Structure and Mechanism of the Human ER Membrane Protein Complex

Thesis by

Giovani Pinton Tomaleri

In Partial Fulfillment of the Requirements for the

degree of

Doctorate in Philosophy

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2023

Defended May 26th, 2023

Giovani Pinton Tomaleri
ORCID: 0000-0001-9661-6480

ACKNOWLEDGEMENTS

"Success is not about competition; it's about contribution." As Adam Grant wisely puts it, the true champions are those who go out of their way to uplift others. When I reflect on this thesis, I am overwhelmed with gratitude for the countless individuals who offered their unwavering support and valuable advice. While my name may be the only one on the title of this document, it does not diminish the profound impact that many people had on its creation. Their contributions are etched deep within my heart.

First and foremost, I want to express my deepest gratitude to my advisor, Rebecca M. Voorhees, for her exceptional mentorship and unwavering guidance throughout my research journey. Rebecca's support has not only allowed me to excel academically but has also encouraged me to embrace my true self and pursue my passion. Her mentorship has honed my research skills and fostered personal growth. I am incredibly fortunate to have had such an exceptional advisor, and I will forever carry the invaluable lessons I have learned from her.

I extend my heartfelt appreciation to my mentor, Maria Cristina Nonato, whose support, and patience during my undergraduate years ignited my passion for scientific exploration. Cristina's introduction to the world of science in my freshman year of college has had a lasting impact on my career. I am forever grateful for her instrumental role in shaping my scientific path. Additionally, I would like to acknowledge Ricardo A. P. Padua, who, as Cristina's Ph.D. student at the time, became my mentor and dear friend. Ricardo's invaluable guidance in the laboratory and his brilliant scientific mind continue to inspire me. Both Cristina and Ricardo have played a significant role in my development as a scientist.

Cristina's guidance also led to a transformative opportunity when she introduced me to Paul D. Adams. Paul provided me with the chance to conduct research in the United States, exposing me to the intricacies of Structural Biology. Under Paul's expert guidance, I developed a deep understanding and appreciation for this captivating field of study. I am profoundly grateful to Cristina and Paul for their support, mentorship, and the immeasurable impact they have had on my academic and professional trajectory. Their mentorship has been a true gift, and I consider myself incredibly fortunate to have had their guidance throughout my journey.

ABSTRACT

The successful synthesis, targeting, insertion, folding, and assembly of membrane proteins into designated membranes is a crucial process in cell biology. Recent research has shed valuable light on this process through the discovery of the endoplasmic reticulum (ER) membrane protein complex (EMC) and its role in membrane protein biogenesis and quality control. As part of my Ph.D. research in the Voorhees lab, I collaborated with esteemed scientists to investigate the EMC in atomic detail. Described in this thesis is the mechanistic basis of EMC function in membrane protein biogenesis, including recent insights into its broader role beyond its well-defined insertase function. First, we determined the structure of the human EMC using single-particle cryo-electron microscopy (cryo-EM). The structure revealed that it utilizes a mechanism similar to other protein-conducting channels, which involves membrane thinning and polar intramembrane residues to transport substrate transmembrane domains from the cytosol into the membrane. The EMC structure provided the foundation for the subsequent rigorous analysis of its role in membrane protein biogenesis. Through this work, we demonstrate the molecular mechanisms involved in the EMC-dependent path a membrane protein takes, from its initial cytosolic capture by methionine-rich loops of the EMC to its eventual membrane insertion via a hydrophilic vestibule. We further show that specific polar intramembrane residues on the EMC serve as an ER "selectivity filter" which uses charge-repulsion properties to reject mis-targeted mitochondrial membrane proteins and maintain organelle integrity. We also demonstrate that the EMC ensures that transmembrane-spanning substrates adopt the correct topology by promoting the "positive-inside" rule, which states that positively and negatively charged amino acids localize to the interior (cytoplasmic) and exterior (non-cytoplasmic) sides of membranes, respectively. Finally, our studies suggest that the EMC has a broader role beyond its well-defined insertase function. Specifically, we found that the EMC physically binds to other factors involved in membrane protein biogenesis, providing a shared interaction surface that acts as a hub to integrate signals from other pathways. Using a combination of structural and functional approaches, we identified the EMC's interaction with Nodal modulator (NOMO) complex, which is part of the multipass translocon complex and facilitate membrane protein biogenesis. Together, these results define and expand the model for membrane protein biogenesis at the ER membrane by the EMC and highlight the complex interplay between different factors in this important process. Together, these results define and expand the model for membrane protein biogenesis at the ER membrane by the EMC and highlight the complex interplay between different factors in this important process.

PUBLISHED CONTENT AND CONTRIBUTIONS

Pleiner, T.*, **Pinton Tomaleri, G.***, Januszyk, K.*, Inglis, A.J., Hazu, M., Voorhees, R.M. (2020) "Structural basis for membrane insertion by the human ER membrane protein complex" Science, 369(6502): 433-436. doi: 10.1126/science.abb5008.

Contributions: G.P.T. participated in the conception of the project, expressed, purified, and reconstitute the human EMC into lipid nanodisc, and built and refined the atomic model.

Pleiner, T., Hazu, M., **Tomaleri, G.P.**, Januszyk, K., Oania, R.S., Sweredoski, M.J., Moradian, A., Guna, A., Voorhees, R.M. (2021) "WNK1 is an assembly factor for the human ER membrane protein complex." Mol Cell, 81(13): 2693-2704.e12. doi: 10.1016/j.molcel.2021.04.013

Contributions: G.P.T. expressed, purified, and assemble the tetrameric EMC cytosolic domain (EMC2, -8, -3, and -5), and also performed and analyzed the biochemical experiments that confirmed the interaction of EMC2 with the WNK1 helix is mutually exclusive to the interaction with EMC8.

Guna, A., Hazu M., **Tomaleri, G.P.**, Voorhees, R.M. (2022) "A TAle of Two Pathways: Tail-Anchored Protein Insertion at the Endoplasmic Reticulum" Cold Spring Harbor Perspectives in Biology doi: 10.1101/cshperspect.a041252

Contributions: G.P.T. prepared figures for the manuscript.

Stevens, T. A., **Tomaleri, G.P.**, Hazu, M, Wei, S., Nguyen, V. N., DeKalb, C., Voorhees, R.M., Pleiner, T. (2023) "A nanobody-based strategy for rapid and scalable purification of native human protein complexes" bioRxiv doi:10.1101/2023.03.09.531980

Contributions: G.P.T. cloned the genes, generated stable mammalian cell lines, and expressed and purified NOMO complex and AE2.

Pleiner, T.*, Hazu, M.*, **Tomaleri, G.P.***, Nguyen, V. N., Januszyk K., Voorhees, R.M. (2023) "A selectivity filter in the EMC limits protein misinsertion at the ER" in press, JCB, bioRxiv doi: 10.1101/2022.11.29.518402

Contributions: G.P.T. participated in the conception of the project, expressed and purified the human EMC in detergent, and collected and processed the single particle cryo-electron microscopy data sets.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND TABLES

C h a p t e r   1

GENERAL INTRODUCTION

1.1 Overview

Membrane-spanning proteins are a crucial component of human cell function, and approximately one fourth of all genes in a human cell encode these proteins (Boyd et al., 1998; Krogh et al., 2001; Wallin & Von Heijne, 1995). They play diverse roles in biological processes such as intracellular trafficking, small molecule transport, signal transduction, organelle biogenesis, and cell adhesion (Almén et al., 2009; Michael Gromiha & Ou, 2014). Due to their essential role in cell function, mutations or aberrant activity in membrane protein biogenesis and quality control can lead to developmental disorders and diseases (Ng et al., 2012; Sanders & Myers, 2004; Schlebach et al., 2015). Accordingly, more than half of current therapeutics for the treatment of human disease target membrane-spanning proteins (Overington et al., 2006; Rask-Andersen et al., 2014).

Integral membrane proteins (IMPs) are a challenging and diverse class of proteins due to the presence of transmembrane domains (TMDs) composed of nonpolar amino acids that span the membrane (White & Von Heijne, 2005). As TMD-containing proteins are prone to aggregation in the cytoplasm, the efficient recognition, targeting, and insertion of TMDs is crucial for their proper folding, assembly, and function (Cymer et al., 2015; Shao & Hegde, 2011b). IMPs are classified into single-pass, multi-pass, and tail-anchored proteins based on the number and orientation of their TMDs. Single-pass and multi-pass IMPs are further classified into Type I, II, or III, based on the presence of a signal peptide and the orientation of their N-terminal flanking domain (Chou & Shen, 2007). The N-terminal flanking domain is classified as N_{cyt} if it faces the cytosol, or N_{exo} if it faces the exoplasmic side of the membrane (Hartmann et al., 1989). Type I IMPs are signal peptide-containing membrane proteins that adopt the 'N_{exo}' topology, while Type II IMPs adopt the 'N_{cyt}' topology. Type III IMPs lack a signal peptide but adopt the 'N_{exo}' topology (Chou & Shen, 2007; Hartmann et al., 1989). On the other hand, tail-anchored (TA) proteins have a single TMD near the C-terminus.

Efficient recognition, targeting, and insertion of TMDs are essential for proper folding and assembly of IMPs and to prevent aggregation or nonproductive interactions of TMDs before they reach their correct cellular compartment for maturation and assembly (Cymer et al., 2015; Shao & Hegde, 2011b). Failure in any step of the biogenesis process or mislocalization of IMPs can lead to cellular stress, requiring degradation of the protein (Hegde & Ploegh, 2010; Meusser et al., 2005). Thus, specialized factors are necessary to identify TMDs upon their synthesis in the cytosol and shield them from the aqueous environment until they engage with the appropriate insertion machinery for membrane insertion (Borgese & Fasana, 2011; Keenan et al., 2003).

The endoplasmic reticulum (ER) membrane is crucial for the quality control of membrane-spanning proteins (Shao & Hegde, 2011a). Consistent with this, it houses numerous molecular chaperones and proteins that aid in IMPs maturation (Shao & Hegde, 2011a). The diverse nature of IMPs requires specialized machinery due to their variable length, hydrophobicity, helical propensity, and protein context (Borgese & Fasana, 2011; Keenan et al., 2003). One mechanism alone cannot meet the requirements for recognition, targeting, translocation, and insertion challenges. Over several decades, scientists have gained knowledge about co-translational and post-translational pathways for protein biogenesis into the ER. These pathways include both the signal recognition particle (SRP)-dependent and SRP-independent pathways. Recent studies expanded this knowledge and revealed three additional pathways addressing targeting, translocation, and insertion challenges: the SRP-independent (SND) (Aviram et al., 2016) , the transmembrane and coiled-coil domains 1 (TMCO1) (Anghel et al., 2017), and the ER membrane protein complex (EMC) (Christianson et al., 2012; Jonikas et al., 2009).

1.2 Integral membrane protein biogenesis at the ER

The co-translational pathway is a crucial mechanism for the biosynthesis of secreted proteins, and it is the primary route relied upon by most ER-targeted IMPs, including Type I and the majority of Type II IMPs (Demangel & High, 2018; M. McKenna et al., 2016, 2017). During IMP synthesis, the conserved ribonucleoprotein SRP tightly associates with the ribosome to both prevent its aggregation and to ultimately steward it to the ER, where it is then transferred to membrane protein insertase for its insertion and assembly. Specifically, The M domain of the SRP54 subunit sits by the ribosomal exit tunnel and engages with the hydrophobic nascent protein chain as it emerges (Halic et al., 2004;

Schaffitzel et al., 2006). This initial engagement allows SRP to shield the TMD from the aqueous environment of the cytosol directly at the point of its synthesis. At the ER membrane, the SRP-ribosome nascent chain (RNC) complex is targeted and delivered to the ER membrane via the SRP receptor (SR), forming a prehandover complex, in a GTP-dependent manner (Akopian et al., 2013; Halic & Beckmann, 2005; Keenan et al., 2003). Finally, the RNC is transferred to the secretory translocation channel (SEC translocon) either for its translocation into the ER lumen or insertion into the ER membrane. This critical interaction between the ribosome and SEC translocon aligns the ribosome exit tunnel with the central pore of the SEC translocon, enabling direct engagement of TMDs and preventing their exposure to the cytosol (Beckmann et al., 1997).

The SEC translocon, also known SEC61 protein-conducting channel is composed of three subunits, α, β, and γ. The Sec61α subunit, which is widely conserved, creates a central channel or conduit for the translocation of polypeptides into the ER lumen. Structural data of mammalian SEC translocon reveal that Sec61α forms an hourglass-shaped channel divided into two halves, amino (N) and carboxyl (C) terminus. The N-terminal half consists of transmembrane segments 1-4, whereas the C-terminal region consists of segments 5-10. The cytosolic side of the channel appears completely exposed, while the lumenal side is occluded by the plug domain (Park & Rapoport, 2012; Van Den Berg et al., 2003).

The mammalian SEC61-mediated protein translocation involves a series of important conformational changes (Voorhees et al., 2014; Voorhees & Hegde, 2016). Notably, the opening of the 'lateral gate' between TMDs 2 and 7 of SEC61 is a key step in translocase function. Binding of the RNC to the SEC6, more specifically the C-terminal half of Sec61a, triggers a series of conformational changes in the SEC translocon (Voorhees & Hegde, 2016). This triggers an "idle" state that leads to the opening of the lateral gate on both the cytosolic and intramembrane sides. This opening is essential for substrate binding, which in turn facilitates the dislodgement of the plug domain and full opening of the central pore (Voorhees et al., 2014; Voorhees & Hegde, 2016). These changes enable the substrate to move either into or across the membrane in an iterative insertion way (Blobel, 1980; Matlack et al., 1998).

While it has been assumed that the lateral gate of SEC61is used to insert all TMDs from multi-pass membrane proteins, there is limited experimental evidence to support this idea (Blobel, 1980; Matlack

et al., 1998). The variability in TMD sequences, biophysical properties, and spacing raises questions about whether late TMDs always insert in the same way as initial TMDs, as postulated by the iterative insertion model.

Recent studies combining biochemistry and structural biology propose the existence of a multipass translocon that responds to a multipass protein substrate (McGilvray et al., 2020). This "supercomplex" consists of several protein complexes, including PAT, NOMO, the heterodimer TMCO1/C20Orf24, and SEC61 (Smalinskaitė et al., 2022; Sundaram et al., 2022). While the proposed framework is speculative due to the unknown functions of some components and the limited amount of substrate tested, preliminary observations suggest that the insertion of TMDs in multi-pass membrane proteins is a complex process that requires coordinated action by several protein complexes. Further studies are necessary to fully comprehend the mechanisms involved in this process and validate the proposed model of multipass translocon assembly.

These findings highlight the intricate and complex nature of protein insertion and translocation across the ER membrane via the co-translational pathway, which is the main route for most ER-inserted IMPs. However, certain proteins, including a subset of IMPs and tail-anchored membrane proteins (TAs), do not use this pathway. For these proteins, alternative mechanisms such as the SRP-independent post-translational pathway are used to target fully synthesized proteins to the ER.

1.3 Tail-anchored membrane protein biogenesis at the ER

TA proteins represent a distinctive membrane protein class that poses a considerable biosynthetic challenge for cells. Unlike most IMPs, TAs have a single TMD located near the carboxyl terminus, which is occluded by the ribosome during translation (Kutay et al., 1993, 1995). Nascent TAs undergo post-translational targeting to either the outer-mitochondrial membrane (OMM) or to the ER membrane. The guided entry of TA proteins (GET) pathway is specifically used for TA protein insertion at the ER (Favaloro et al., 2008; Stefanovic & Hegde, 2007a).

The GET pathway involves a set of unique factors for targeting and insertion from those used in the SRP pathway. The co-chaperone small glutamine-rich tetratricopeptide repeat-containing alpha (SGTA) binds and shields the TMDs of TA proteins upon their synthesis in the cytosol before handing them off to a targeting factor called the transmembrane recognition complex subunit of 40 kDa

(TRC40, now widely called GET3), an adenosine triphosphatase. The pre-targeting complex, Ubl4A-Bag6-GET3, allows substrate triage for either membrane insertion or degradation by the ubiquitin-proteasome system. Once sorted among these factors, the GET3-TA complex is subsequently trafficked to the ER where it binds specifically to the GET3 receptor, WRB/CAML, in a process regulated by ATP hydrolysis. This receptor dislodges the TMD from GET3 allowing WRB/CAML to mediate its insertion into the lipid bilayer (Shao et al., 2017; Stefanovic & Hegde, 2007a; Wang et al., 2010).

It is worth noting that the SRP and GET pathways were initially established as the canonical routes for IMP biogenesis, as they satisfied the general requirements for co- and post-translational synthesis. However, the enormous topological and biophysical diversity of TMDs suggests that some facets of this process are still not fully understood. Recent studies on the SRP-independent (SND) pathway and the ER membrane protein complex (EMC) have shed light on the level of complexity of IMP targeting and insertion (Aviram et al., 2016; Casson et al., 2017; Haßdenteufel et al., 2017).

The SND pathway was identified using high-throughput screening in yeast, and it is composed of three proteins, namely Snd1, Snd2, and Snd3 (Casson et al., 2017). These proteins have been implicated in the trafficking of IMPs with central TMDs to the ER membrane. Of these three, SND2 is the only one with an identified mammalian homolog, known as hSnd2, and it has been postulated that it functions in an alternative ER-delivery route for TA proteins (Haßdenteufel et al., 2017). However, it is unclear how both substrate recognition and specificity are mediated by SND2. Thus, it remains unknown how a large fraction of TA proteins that cannot engage the GET pathway get inserted into the ER membrane.

1.4 The ER membrane protein complex (EMC)

The EMC is a multi-subunit complex that is highly expressed and conserved across eukaryotes. The EMC is involved in the biogenesis of IMPs and the maintenance of ER homeostasis (Jonikas et al., 2009). The complex was first discovered in yeast by Jonikas et al. (2009) who identified six genes, EMC1 through EMC6, that yielded a seemingly stoichiometric complex upon protein co-purification. Christianson et al. (2012) later independently identified the mammalian EMC ortholog, a larger complex, consisting of ten proteins named EMC1-10, using mass spectrometry to map the ER-associated degradation (ERAD) interaction network (Christianson et al., 2012).

Subsequent studies and bioinformatics analyses revealed that the yeast genes Sop4 and YDR056C are homologs of the mammalian EMC7 and EMC10 genes, and that the lower eukaryotes lack the orthologs of EMC8 and 9 (Louie et al., 2012; Wideman, 2015). Moreover, these studies demonstrated that EMC8 and EMC9 are paralogs (Wideman, 2015) that, originated from a gene duplication event, and therefore, EMC in most eukaryotes is composed of nine subunits: EMC1 through EMC7, EMC8 or 9, and EMC10.

Although the functions of the EMC were initially unclear, later studies demonstrated that it influences the biogenesis, quality control, and trafficking of a wide range of membrane proteins (Bircham et al., 2011; Richard et al., 2013; Satoh et al., 2015), including those with low-to-moderate hydrophobicity TMD-containing TA proteins that cannot engage the GET pathway (Guna et al., 2018). Moreover, the EMC can insert the first TMDs of Type III IMPs co-translationally and may potentially function as a chaperone for TMDs during co-translational processes (Chitwood et al., 2018). Thus, the EMC is involved in both co-translational and post-translational pathways for protein biogenesis at the ER membrane.

The EMC-dependent co-translational insertion model suggests that Type III IMPs are co-translationally inserted into the ER membrane after the initial engagement of RNC·SRP with the SRP receptor in what is known as the formation of a prehandover complex (Chitwood et al., 2018). The EMC then samples the prehandover complex and selects only suitable substrates for insertion and rejects others that immediately engage with the SEC61 complex (Chitwood et al., 2018). Once the proper substrates are identified and their TMDs are correctly inserted by EMC, the RNC docks tightly onto the SEC61 complex, allowing the remainder of the protein to resume synthesis at the ER membrane, where the nascent protein region is directly inserted into the ER. The correct topology of insertion, N_{exo}, of the TMDs by EMC is thus the primary EMC-dependent step for Type III IMP biogenesis.

Compared to the GET pathway, the mechanisms underlying TA protein targeting and delivery to the EMC are poorly understood. Whereas GET3 targets high hydrophobicity TMDs of TA proteins during GET-dependent pathway insertion, there is no equivalent factor for the low-to-moderate hydrophobicity TMDs in the EMC (Guna et al., 2018). Therefore, it has been proposed that direct substrate-EMC interaction is the mechanism of EMC targeting. Recent studies have shown that

cytosolic TMD-binding factors like calmodulin (CaM) and SGTA can prevent TA protein aggregation and facilitate EMC-mediated ER insertion (Guna et al., 2018). In fact, substrate-loaded CaM alone can promote EMC-dependent TA protein insertion into the ER membrane, suggesting that the release of CaM at physiological Ca^{2+} concentrations allow EMC to engage with substrates (Guna et al., 2018). However, further research is necessary to fully understand the mechanisms involved in TA protein targeting to the EMC.

The structural architecture of the EMC has been the subject of much research (Guna et al., 2023; Hein et al., 2015; Jonikas et al., 2009; Lahiri et al., 2014; Li et al., 2014; Shurtleff et al., 2018). Biochemical experiments have suggested a stoichiometric assembly of the subunits, such that each subunit is present in a single copy within the complex. Notably, the disruption of the core subunits (EMC1, 2, 3, 5, and 6) leads to the degradation of the entire complex and the loss of its activity (Volkmar et al., 2019). However, the absence of more peripheral EMC subunits, such as EMC4 and EMC7, has less effect on the structural integrity of the complex yet compromises its function as seen in disease-causing mutations within these genes (Bagchi et al., 2020; Shurtleff et al., 2018; Volkmar et al., 2019).

The identification of the EMC as an insertase sheds light on why the EMC is implicated in many cellular functions and indicates its key role in the biogenesis of many crucial IMPs. However, without a structure of the EMC, our understanding of precisely how the EMC inserts substrates into the ER and the overall path of EMC-dependent substrates from their synthesis in the cytoplasm to eventual insertion into the ER membrane was limited. As described in this thesis, I took a structural biology approach to answer questions related to the macromolecular assembly of the EMC in the context of the ER membrane and to define the precise structural features that enable EMC to productively engage with its diverse substrates.

1.5 Overview of the thesis

My thesis research aims to elucidate the non-canonical protein insertion pathway by studying the EMC. Chapter 2 focuses on the structural basis of membrane insertion by the human ER membrane protein complex. Using cryo-electron microscopy, we determined the structure of the human EMC and built a nearly complete atomic model. Furthermore, we employ structure-guided mutagenesis to identify critical EMC features involved in IMP biogenesis. Our study enhances our understanding of

the mechanisms underlying membrane protein insertion and provides crucial insights into the function of the EMC.

In Chapter 3, we show data describing the presence of a selectivity filter in the EMC that functions to limits protein misinsertion. Using structural biology and biochemical approaches, my work sheds light on how the EMC ensures accurate and efficient insertion of its substrates into the ER membrane. Chapter 4 focuses on how the EMC interacts with other factors to facilitate membrane protein biogenesis, specifically the EMC's role in the multipass membrane protein insertion by interaction via the NOMO complex. Taken together, Chapters 1-4 describe the structure-function relationship of the EMC. By understanding how the EMC fits into the larger context of membrane protein biogenesis, we hope to identify new avenues for research.

Finally, Chapter 5 concludes this thesis by reviewing the findings discussed in the previous chapters, giving a high-level overview of the field prior to and after the research presented here. The work presented here sheds new light on the molecular mechanisms of membrane protein biogenesis and provides insights into the molecular basis of protein biogenesis at the ER.

Chapter 2

STRUCTURAL BASIS FOR MEMBRANE INSERTION BY THE HUMAN ER MEMBRANE PROTEIN COMPLEX

The following chapter is adapted from Pleiner, Pinton Tomaleri and Januszyk et al., 2020 and modified according to the Caltech Thesis format.

Pleiner, T., Tomaleri, G. P., Januszyk, K., Inglis, A. J., Hazu, M., & Voorhees, R. M. (2020). Structural basis for membrane insertion by the human ER membrane protein complex. Science, 369(6502), 433–436. https://doi.org/10.1126/SCIENCE.ABB5008

2.1 Abstract

A defining step in the biogenesis of a membrane protein is the insertion of its hydrophobic transmembrane helices into the lipid bilayer. The nine-subunit ER membrane protein complex (EMC) is a conserved co- and post-translational insertase at the endoplasmic reticulum. Here we report the structure of the human EMC in a lipid nanodisc determined to an overall resolution of 3.4 Å by cryo-electron microscopy, permitting building of a nearly complete atomic model. We used structure-guided mutagenesis to demonstrate that substrate insertion requires a methionine-rich cytosolic loop and occurs via an enclosed hydrophilic vestibule formed by the subunits EMC3 and EMC6. We propose that the EMC uses local membrane thinning and a positively charged patch to decrease the energetic barrier for insertion into the bilayer.

2.2 Main

The human genome encodes over 5000 integral membrane proteins, all of which contain hydrophobic transmembrane helices (TMs) that must be inserted into the lipid bilayer (Bateman et al., 2017). At the endoplasmic reticulum (ER), multiple insertion pathways operate in parallel to accommodate the enormous topological and biophysical diversity of these substrates (Aviram et al., 2016; Görlich & Rapoport, 1993; Guna et al., 2018; Mariappan et al., 2011; Schuldiner et al., 2008; Stefanovic & Hegde, 2007a). The ER membrane protein complex (EMC) is a ubiquitously expressed and widely

conserved membrane protein insertase (Chitwood & Hegde, 2019; Guna et al., 2018; Jonikas et al., 2009; Wideman, 2015), which both post-translationally inserts tail-anchored proteins and co-translationally inserts some multipass membrane proteins (Chitwood et al., 2018; Guna et al., 2018). One of the membrane-spanning subunits of the EMC, EMC3, belongs to the Oxa1 superfamily of insertases, which includes the bacterial YidC, the archaeal Ylp1, and the eukaryotic WRB (Anghel et al., 2017; Hennon et al., 2015). An atomic model of the EMC would thus provide insight into substrate insertion at the ER, and further define the general principles of membrane protein biogenesis across all domains of life.

Using cells stably expressing GFP-tagged EMC2, the human EMC was affinity purified using an immobilized, protease-cleavable GFP-nanobody and reconstituted into lipid nanodiscs (Fig. S2.1) (Pleiner et al., 2015). Using cryo-electron microscopy (cryo-EM), we generated a reconstruction of the resulting nine-subunit EMC to an overall resolution of 3.4 Å; correction of inter-domain flexibility resulted in reconstructions of the cytosolic and lumenal regions to 3.6 and 3.2 Å resolution, respectively (Fig. 2.1; Figs. S2.2, S2.3; Table S2.1).
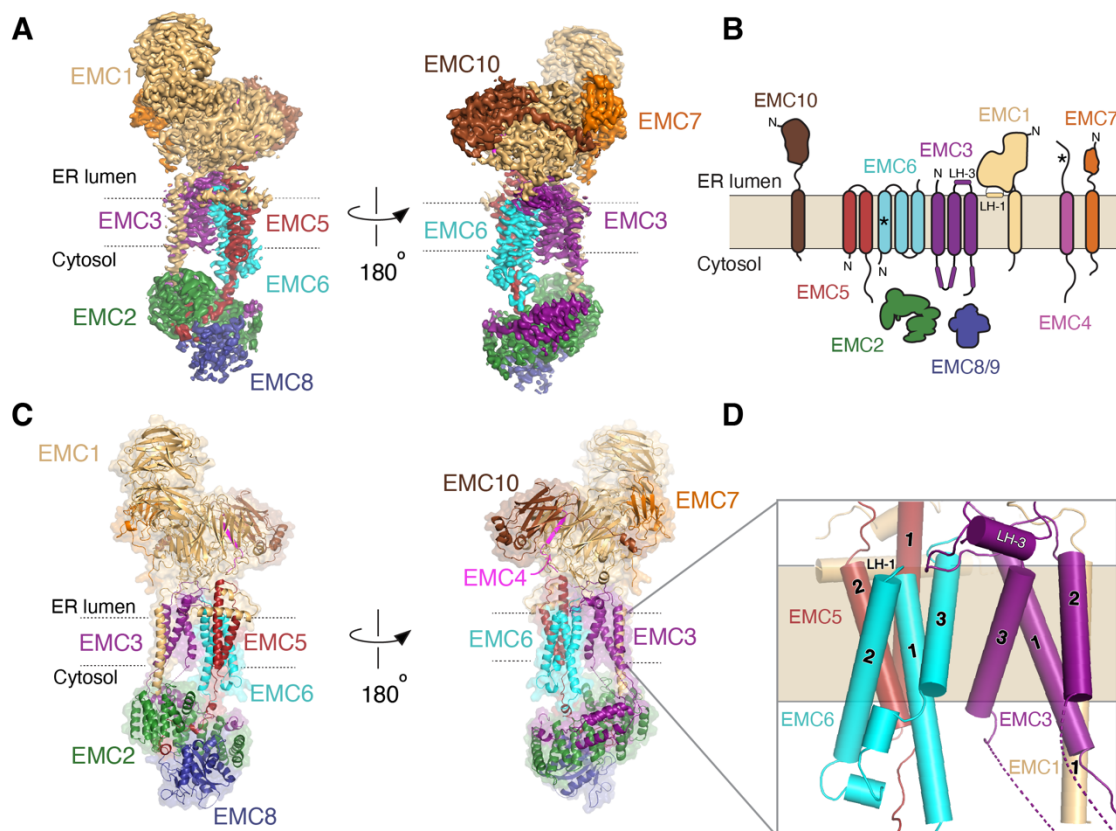
Figure 2.1 **The structure of the human EMC.**

(A) Two views of the sharpened 'overall' density map (Fig. S2.3) from the perspective of the two intramembrane sides of the EMC colored by subunit. (B) Schematic representation of the topology of the nine EMC subunits as determined by the structure. EMC8 and 9 are functional paralogs, and their binding to EMC2 is mutually exclusive. For simplicity, we refer only to EMC8 throughout the text, though most observations will apply to both EMC8 and 9. Helices of EMC1 and 3 that are positioned in the lumenal plane of the membrane are labeled LH-1 and LH-3. Asterisks indicate newly determined topologies based on the structure and experimental data. Note, we cannot unambiguously define the topology of EMC4, but structural data is most consistent with it containing a single TM (Fig. S2.7). (C) Atomic model of the EMC, in the same orientation as the density map in (A). (D) Close-up of the nine core TMs of the EMC and their subunit assignment.

The EMC extends ~200 x 70 x 100 Å and has a tripartite organization: (i) the membrane spanning region is composed of twelve TMs, nine of which form the central ordered core; (ii) a basket-shaped cytosolic region anchored by EMC2 and 8; and (iii) an L-shaped lumenal region comprised of EMC1, 4, 7, and 10. The cryo-EM density maps were sufficient to unambiguously assign and build nearly complete atomic models for EMC1, 2, 3, 5, 6, and 8, and the lumenal domains of EMC4, 7, and 10 (Fig. 2.1B, C; Figs. S2.3-5, Table S2.2). In doing so, we found that EMC6 unexpectedly contained three TMs. We demonstrated that the poorly hydrophobic TM1 of EMC6 (ΔG=3.8; (Hessa et al., 2007)) inserted only upon assembly with EMC5, a conserved strategy for stabilizing poor TMs at subunit interfaces (Fig. S2).(Carvalho et al., 2019; Inglis et al., 2020). We also observed weak density for three putative TMs extending from the lumenal domains of EMC4, 7, and 10 (Fig. S2.7; movie S2.2). The dynamic nature of these TMs is consistent with the limited roles of EMC4, 7, and 10 in EMC stability (Volkmar et al., 2019).

Thus, the membrane-spanning region of the EMC is pseudo-symmetric: three TMs of EMC6 abut the three TMs of EMC3 at the complex's midline. On either side, EMC5 and 1 anchor the cytosolic and lumenal domains, respectively. In addition, the EMC contains at least two helices within the lumenal plane of the bilayer contributed by EMC1 and 3 (Fig. 2.1D). Like the amphipathic EH-1 helix of YidC, these helices may position the complex within the membrane and locally remodel the bilayer (Drin & Antonny, 2010; Kumazaki et al., 2014).

In the cytosol, EMC2 acts as an architectural scaffold for EMC8 and the cytosolic regions of EMC3, 5, and 1, consistent with its essential role in stability of the EMC (Fig. 2.2A) (18). EMC2 forms an α-solenoid that binds the three-helix bundle formed by the coiled coil and C-terminus of EMC3 (Fig. 2.2A; Fig. S2.8). The C-terminus of EMC1 forms π- stacking interactions within a cleft of EMC2, and EMC2 clamps around EMC8 through an extensive hydrophobic surface (Table S2.3). Together, EMC2 and 8 form a composite interface with the C-terminal tail of EMC5, which traverses through the center of EMC2 to the cytosolic face of the complex. Mutations at the interfaces between EMC2, 3, 5, and 8 disrupted subunits binding in vitro, verifying the atomic model (Fig. 2.2B, C; Fig. S2.8).

Finally, the lumenal region is composed of the N-termini of EMC1, 7, 10, and the tail of EMC4 (Fig. 2.2D, E; Fig. S2.8). EMC1 contains two eight-bladed β-propellers, and stabilizes the entire complex,

contacting six of the remaining eight subunits. EMC7 and 10 form β- sandwiches that are anchored to EMC1 via primarily hydrophobic contacts (Table S2.3).
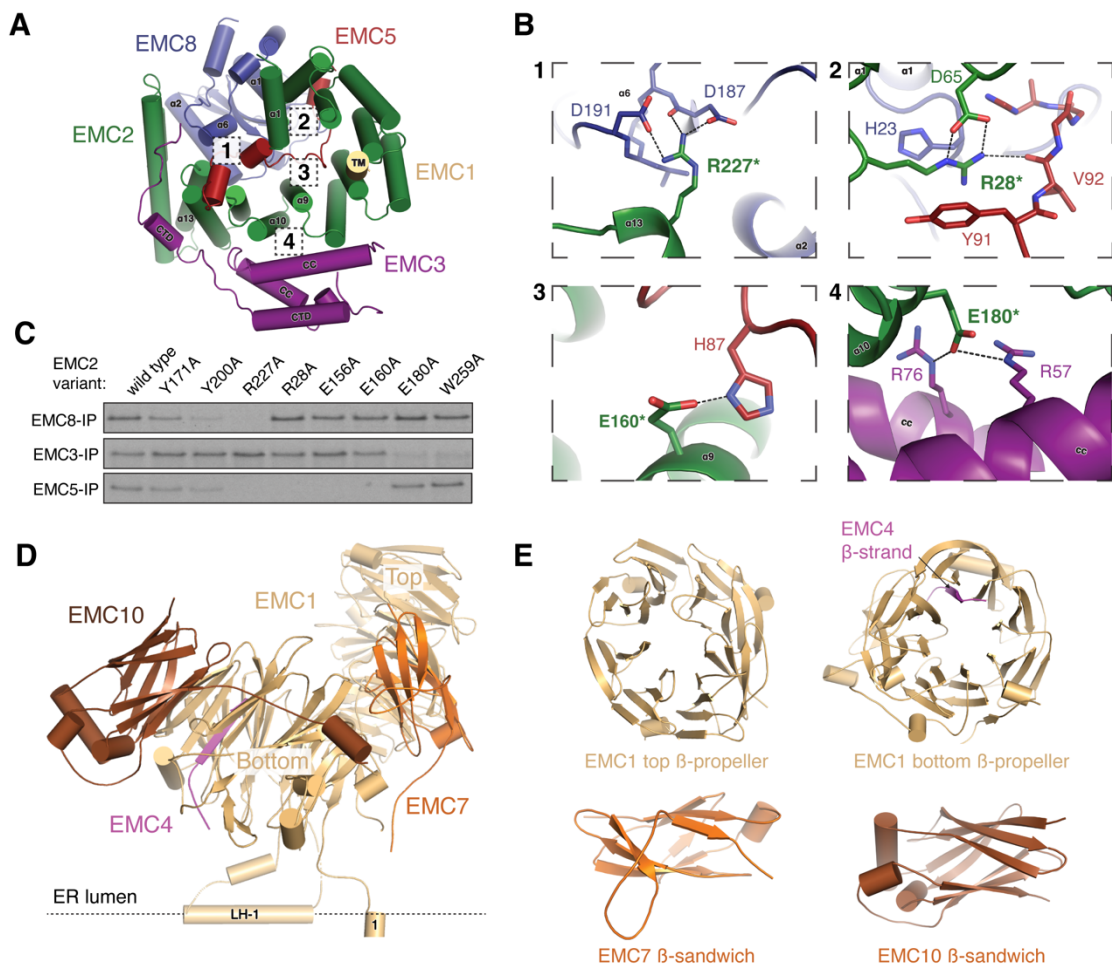
Figure 2.2 **Architecture of the cytosolic and lumenal regions of the EMC.**

(A) View from the membrane of the cytosolic domain of the EMC. (B) Close-up of the primary interfaces between the cytosolic subunits of the EMC indicated in (A). Dashed lines represent polar interactions, and asterisks indicate mutations that disrupt complex assembly (Fig. S2.8). (C) 35S-methionine labeled wild type EMC2 or the indicated point mutants were translated in rabbit reticulocyte lysate (RRL) and tested for binding to FLAG-tagged EMC8, EMC3, or EMC5 by co-immunoprecipitation using anti-FLAG resin. (D) Side view of the EMC lumenal domain. (E) Cartoon

model of the globular N-termini of EMC1, 7, and 10. EMC1 and EMC4 together form one of the four-stranded blades of the bottom β-propeller.

This atomic model permitted detailed interrogation of how EMC facilitates substrate insertion. We first used site-specific crosslinking to establish EMC3 as the primary interaction partner of a tail-anchored substrate with purified EMC (Fig. S2.9). However, the EMC contains two intramembrane surfaces that could be involved in insertion and are consistent with an EMC3-substrate crosslink (Fig. 2.3A). On one side is a hydrophobic crevice that runs perpendicular to the plane of the membrane and could accommodate a TM (Fig. S2.10). On the opposite side is a lipid-exposed cytosolic vestibule composed of EMC3 and 6, which is partially enclosed by the TMs of EMC4, 7, and 10, and sealed by EMC3's lumenal helix (Fig. 2.3B; Fig. S2.7). Within this vestibule, EMC3 contains a positive patch in the bilayer that is surrounded by hydrophobic residues (Fig. 2.3B; Fig. S2.11). We postulated that this hydrophilic vestibule formed the insertase based on analogy to YidC: the bacterial homolog of EMC3 relies on positively charged residues in the membrane for insertion (Fig. S2.12) (Kumazaki et al., 2014). Consistent with this model, the sequence conservation of the hydrophilic vestibule is significantly higher than that of the hydrophobic crevice (Fig. S2.11). In particular, R31 of EMC3 is a positive charge in all eukaryotes.

We introduced mutations to residues of EMC3 and 6 that line the hydrophilic vestibule but do not affect complex assembly and tested their effect on substrate biogenesis using an established assay for EMC insertion (Fig. 2.3C-E; Fig. S2.13) (Chitwood et al., 2018; Guna et al., 2018). The mutations R31A and R180A in EMC3 destabilized representative post- and co-translational EMC dependent substrates (SQS and OPRK1) but had no effect on the matched EMC independent controls (VAMP2 and TRAM2). Furthermore, a positive charge at these positions is required for EMC insertion: R31E and R180E caused an insertion defect that was rescued by R31K/R180K for some substrates (Fig. 2.3E; Fig. S2.13). Mutations to the polar residues of EMC6 had a modest, but detectable, effect on EMC-dependent insertion (Fig. S2.13). We therefore concluded that the hydrophilic vestibule of the EMC is required for insertion of both post- and co-translational substrates.
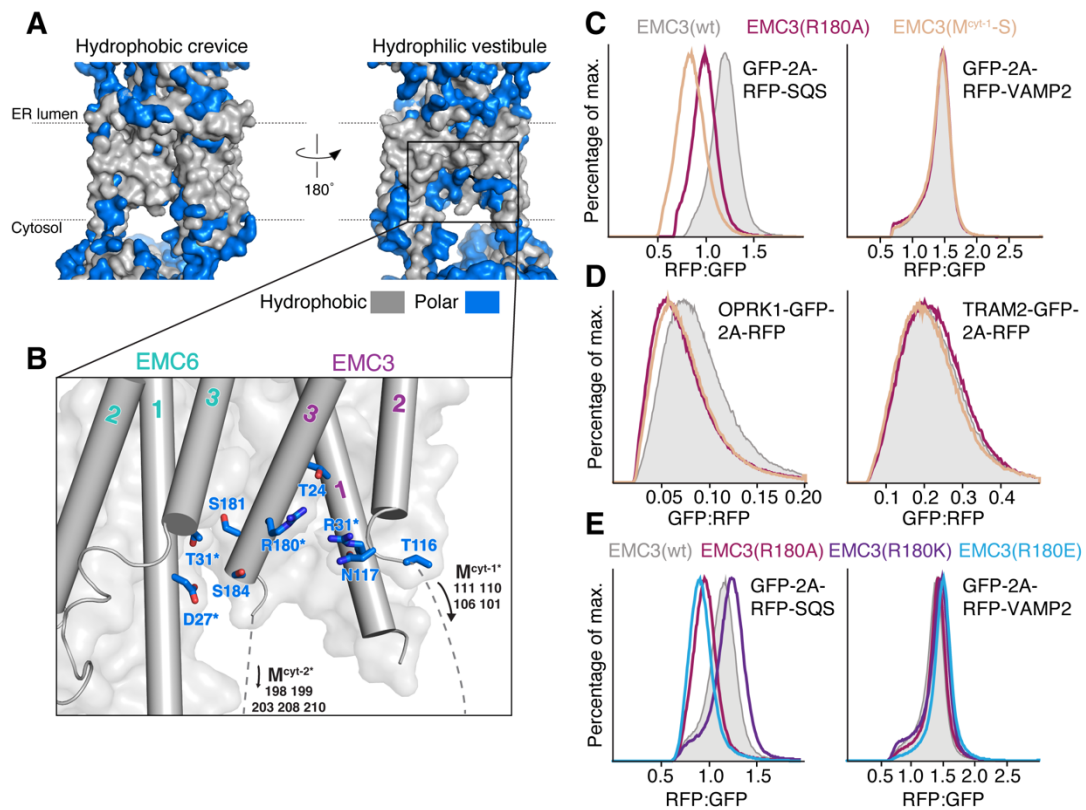
Figure 2.3 **Substrate insertion by the EMC**

(A) Surface filling representation of the membrane-spanning region of the EMC colored with hydrophobic residues in grey and polar residues in blue. Displayed are the two sides of the complex: the 'hydrophobic crevice' (left) and the 'hydrophilic vestibule' (right) as in Fig. 2.1A and C. (B) Close-up view of the hydrophilic vestibule formed by EMC3 and EMC6, with polar residues shown in blue and displayed as sticks. Residues that were mutated in functional assays are highlighted with asterisks (Fig. S2.13). (C) HEK293 cells were generated that stably expressed exogenous wild type or mutant EMC3, as well as the tail-anchored substrates RFP-squalene synthase (SQS; EMC-dependent) or RFP VAMP2 (EMC-independent) (3). The relative RFP fluorescence, normalized to an internal expression control (GFP), is plotted as a histogram. (D) As in (C) but with the co-translational substrates Opioid Receptor Kappa 1 (OPRK1)-GFP (EMC-dependent) and TRAM2-GFP (EMC-independent). (E) As in (C), analysis of the role of positive charge in the hydrophilic vestibule.

It was recently noted that Oxa1 superfamily insertases contain methionine-rich cytosolic loops that were proposed to interact with substrates, in analogy to SRP54 and GET3 (Borowska et al., 2015; Mitra et al., 2004). In the structure of the EMC, these loops were dynamic. Nevertheless, mutation of methionines adjacent to TM2 in EMC3 specifically disrupted biogenesis of co- and posttranslational EMC substrates (Fig. 2.3C, D; Fig. S2.13). These methionines are positioned in the cytosol just below the insertase vestibule and could transiently orient a substrate on its path into the membrane. We do not exclude a direct role for EMC2 in substrate binding; however, we could not identify a suitable hydrophobic surface or groove in the cytosolic domain.

We therefore propose a model for EMC-mediated co- and post-translational substrate insertion (Fig. 2.4). A substrate would first be captured and guided towards the membrane by the flexible methionine-rich cytosolic loop of EMC3, possibly assisted by those of nearby EMC4 and 7. Prior to substrate engagement, the lipid-exposed hydrophilic vestibule of the EMC is axially sealed by the lumenal helix of EMC3 and laterally partially enclosed by the dynamic TMs of EMC4 and 7. The few contacts between EMC3 and 6 suggest this may be a potential site for subunit rearrangement during insertion.

The EMC decreases the energetic cost of insertion in two ways. First, by inducing a local thinning of the membrane by ~10 Å. Similar to other translocases, EMC thus decreases the distance that a substrate's soluble lumenal domain must travel through the hydrophobic bilayer (Fig. 2.4A; Fig. S2.14) (Y. Chen et al., 2017; Mitra et al., 2004; Wu et al., 2020). Second, the EMC positions polar and positively charged residues within the bilayer, which could provide a way station for the substrate's soluble domain, enforce the positive-inside rule, and potentially stabilize the hydrophilic residues that are highly enriched in the TMs of EMC substrates (Fig. 2.4B)(Shurtleff et al., 2018; Tian et al., 2019).

Once within the membrane, the hydrophobic core of the substrate could interact with the hydrophobic surface of EMC3 above and below these polar residues. The dynamic nature of the TMs of EMC4 and 7, which partially enclose the insertase, permit sampling of the lipid bilayer, and may further serve a gating function and/or interact directly with substrate. Finally, the shortened TMs of EMC3 and 6 cannot stably bind a membrane spanning substrate, favoring its partitioning into the bilayer and dissociation from the complex. The substrate's soluble lumenal domain would then encounter the β-

propellers of EMC1, which may serve as a platform for recruitment of co-factors at the site of nascent protein insertion (Shurtleff et al., 2018).
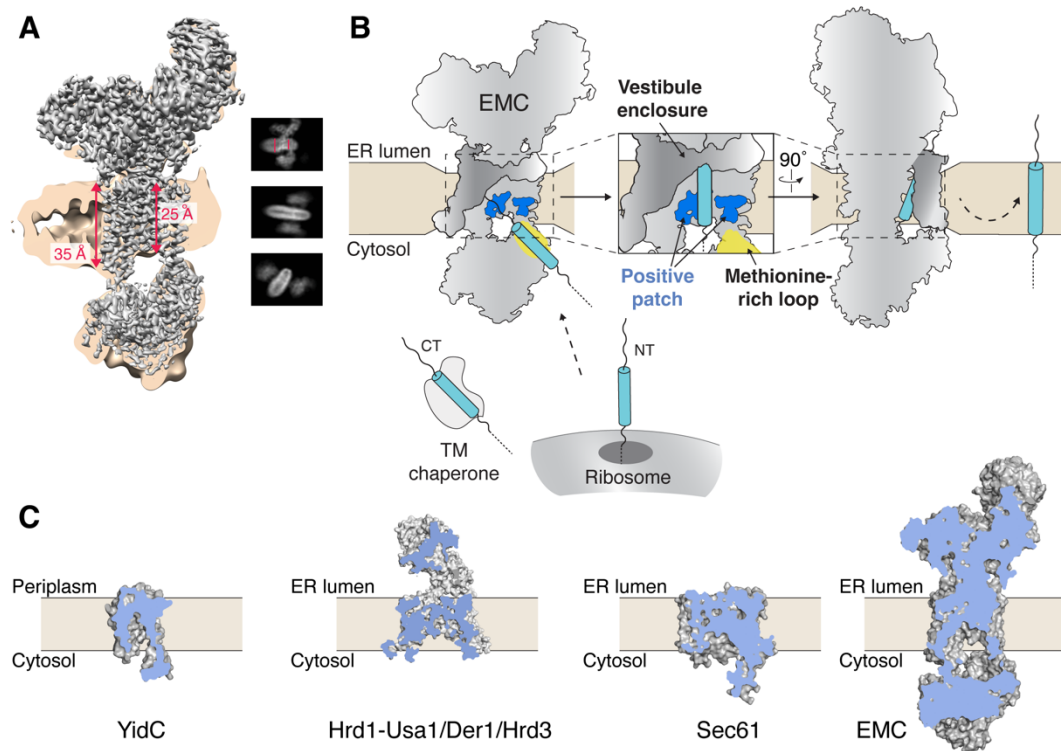
Figure 2.4 **Model for membrane protein insertion by the EMC.**

 (A) Unsharpened EM density maps are shown at low (tan) and high contour (grey) to highlight the thickness of the lipid nanodisc. Distances measured within the density are shown in red (Fig. S2.14). Insets are representative 2D class averages that depict the local thinning of the lipid bilayer by the EMC. (B) Post- and co-translational EMC substrates are released from either a TM chaperone (e.g. calmodulin) or the ribosome, respectively. The flexible methionine-rich loop of EMC3 could capture substrates for insertion through the hydrophilic vestibule along the surface of EMC3. The EMC decreases the energetic barrier for insertion via local thinning of the membrane and a positively charged patch in the bilayer. The TMs of EMC4 and 7 enclose the cytoplasmic vestibule and facilitate insertion. (C) Cut-away view of the space filling-models for the bacterial YidC (PDB 3WO6), the fungal Hrd1-Usa1/Der1/Hrd3 complex (6VJZ), mammalian Sec61 (3J7Q), and the human EMC. A

hydrophilic conduit from the cytosol to the membrane is a general feature of evolutionarily diverse protein conducting channels.

We therefore conclude that the Oxa1 superfamily insertases all rely on qualitatively similar mechanisms for insertion, and there is a marked similarity between EMC3/6 and WRB•CAML (Inglis et al., 2020). More broadly, the presence of a hydrophilic cytosolic funnel is a conserved feature of all protein conducting channels including Sec61, Hrd1, YidC, and now EMC (Fig. 2.4C) (Kumazaki et al., 2014; Voorhees et al., 2014; Wu et al., 2020). However, the significant increase in complexity of the EMC compared to YidC or even Sec61 suggests that its well-defined insertase function represents only a part of its more general role in membrane protein biogenesis and quality control.

2.3 Supplementary Material
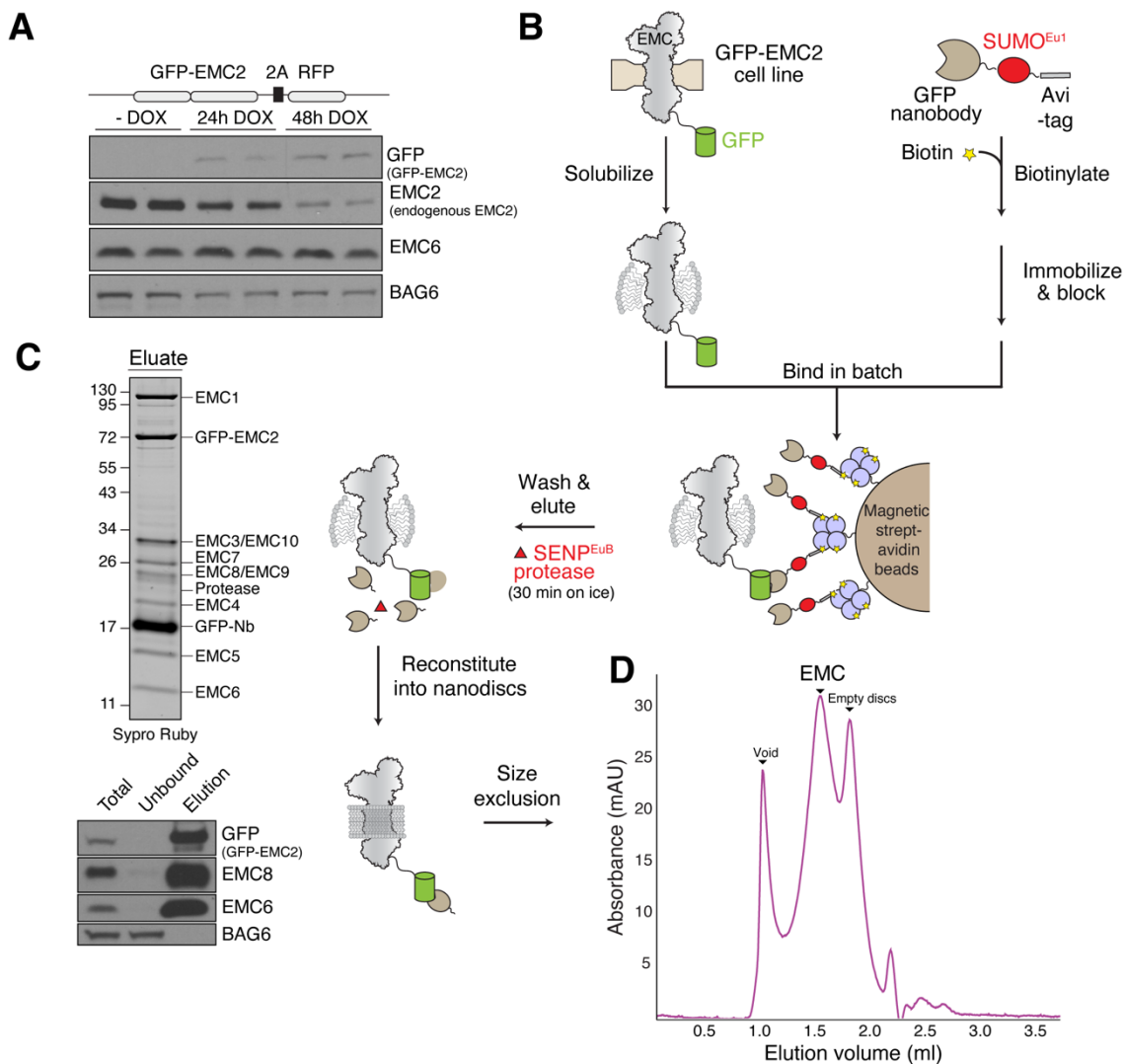
Figure S2.1 **Purification and nanodisc reconstitution of the EMC.**

(A) HEK293 cell lines stably expressing GFP-EMC2-P2A-RFP, where '2A' indicates the viral P2A sequence that causes peptide bond skipping by the ribosome, were harvested either 24 or 48 hours after treatment with carrier or doxycyline (DOX) and analyzed by Western blotting with the indicated antibodies. GFP-tagged EMC2 replaces its endogenous counterpart, which is largely degraded. (B)

Schematic of the GFP-nanobody affinity purification strategy. HEK293 cells stably expressing GFP-EMC2 were adapted to grow in suspension to facilitate large-scale growths. GFP-tagged EMC2 was efficiently incorporated into the EMC, resulting in intact EMC fused to GFP and expressed at endogenous levels. In parallel an Avi tag-SUMOEu1-GFP-nanobody fusion was purified from E. coli, biotinylated using BirA, and immobilized on magnetic Streptavidin beads. Detergent solubilized EMC was affinity purified via the GFP tag, and specifically eluted from the resin under native conditions using SENPEuB protease cleavage. All constructs needed to implement this strategy are available via Addgene (see Materials and Methods). (C) Top: SDS-PAGE gel of the single-step purified EMC, stained with Sypro Ruby. GFP-tagged EMC2 is efficiently incorporated into the intact EMC and thus co-purifies all endogenous EMC subunits, which were identified by mass spectrometry. The native single-step purification described in (B) yielded EMC at sufficient quantities and purity to be utilized directly for reconstitution into nanodiscs. Bottom: A fraction of the anti-GFP nanobody IP eluates was analyzed by Western blotting with the indicated antibodies. (D) Size exclusion chromatography of the DDM solubilized EMC reconstituted into lipid nanodiscs using the MSP2N2 scaffold and POPC lipid. UV absorbance at 280 nm was monitored during the purification to identify fractions containing the reconstituted complex.
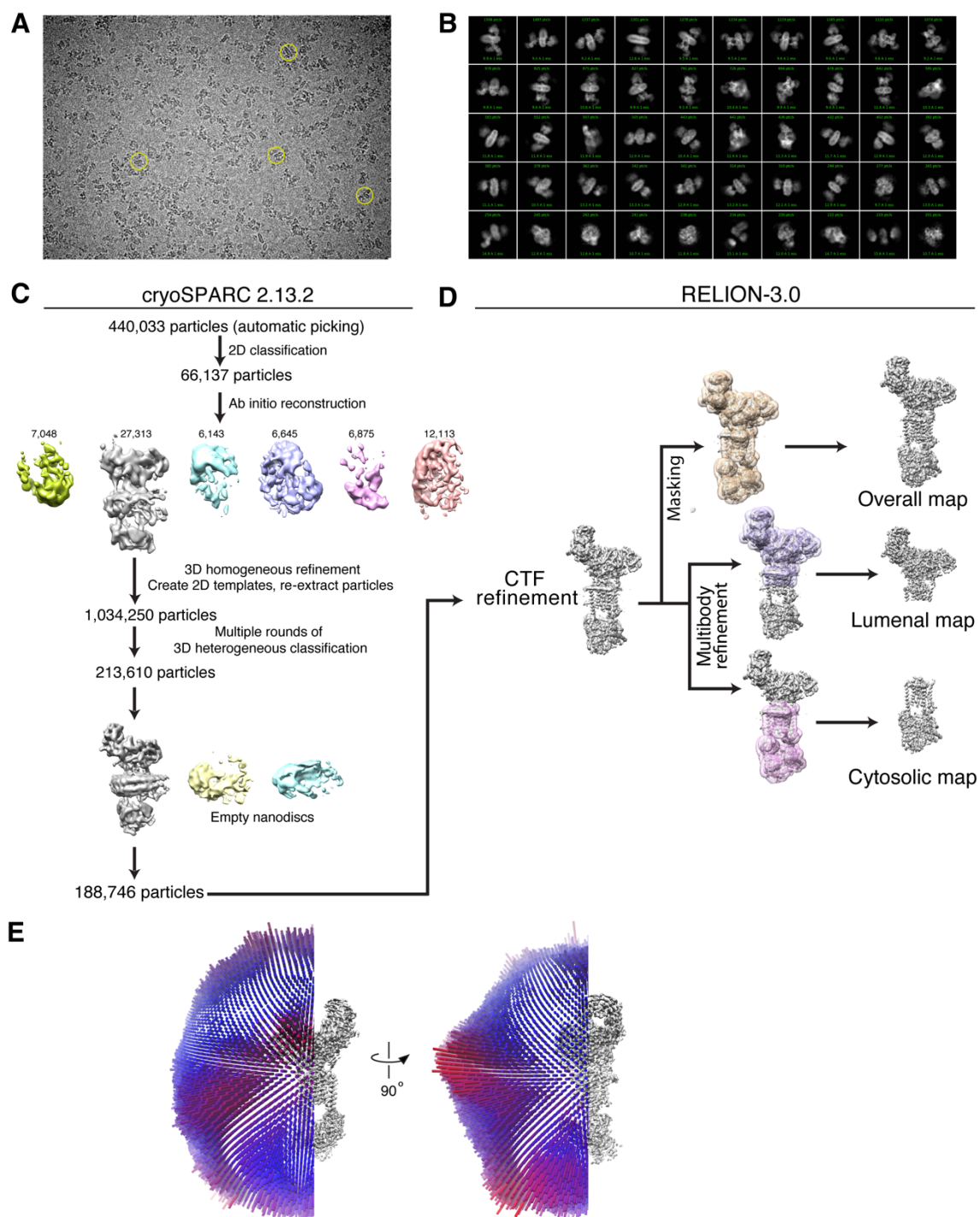
Figure S2.2 **Classification and refinement procedure.**

(A) A representative image with several particles circled in yellow. A total of 6,345 images were collected, and 6,277 were used for further processing based on a high CTF Figure of Merit score and a maximum resolution better than 4.5 Å. (B) Representative 2D class averages after selecting and filtering for intact particles. (C) Classification scheme in cryoSPARC 2.13.2 used to identify a population of particles containing the intact EMC in a lipid nanodisc. Following an initial round of automatic particle picking, iterative rounds of 2D classification and template-based particle picking were used to filter out the majority of empty nanodiscs and broken complexes. After re-extraction of 1,034,250 particles, multiple rounds of 3D heterogeneous classification were employed to characterize a subpopulation of 213,610 particles; 188,746 of these particles appeared to contain intact EMC in a lipid nanodisc. (D) These 188,746 particles were exported from cryoSPARC for further refinement in RELION-3.0. Following CTF refinement, a mask that excluded the nanodisc was applied during 3D refinement resulting in an improvement in the overall resolution of the complex and a final 'overall map' used primarily for assignment and building of the TMs. Multibody refinement was used to generate improved maps for the cytosolic and lumenal domains, which allowed de novo modeling of these regions. This would not have been possible using the 'overall map'. (E) Orthogonal views of the angular distribution of particle views for the final data set (188,746 particles). Each cylinder represents one view. The number of particles with their particular orientations are represented by both length and color (from blue to red) of the cylinders.
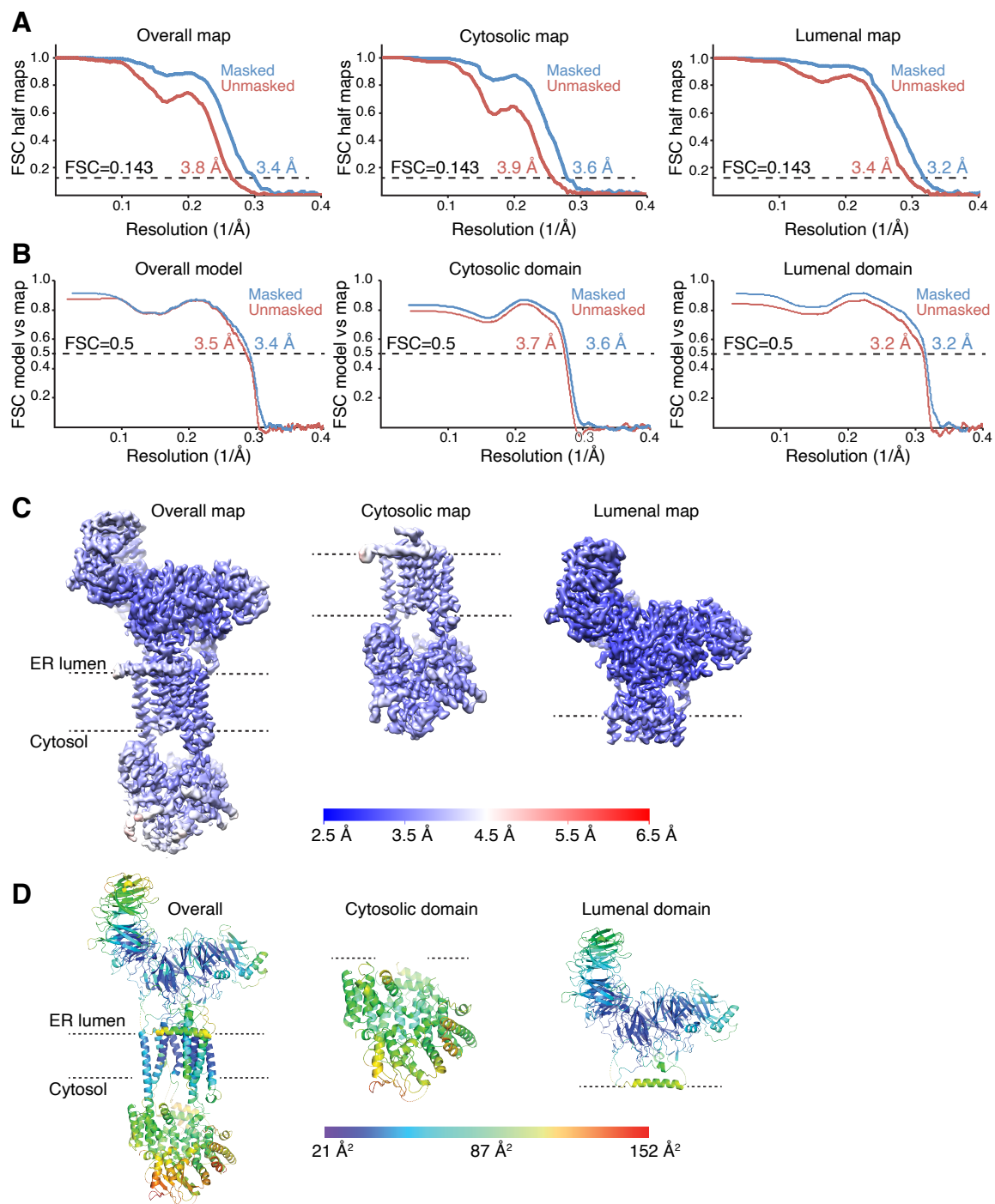
Figure S2.3 **Map and model quality.**

(A) Gold-standard Fourier Shell Correlation (FSC) curves for the masked and unmasked half maps used for building of the EMC calculated using RELION-3.0. While the 'overall map' was sufficient for assigning and building the TMs, Multibody refinement in RELION-3.0 was able to correct for intersubunit flexibility to significantly improve the local resolution and molecular detail visible in the 'cytosolic' and 'lumenal' maps. The nominal resolution of each map is demarcated using the FSC = 0.143 criterion. (B) FSC curves of the model vs. map, masked and unmasked calculated using phenix.real_space_refine. Overall is the complete model vs. the 'overall map', cytosol is the model of the cytosolic domain shown in (D) vs. the 'cytosolic map', and lumen is the lumenal domain vs. the 'lumenal map'. FSC = 0.5 is indicated for the masked and unmasked maps. (C) View of the final unsharpened cryo-EM density maps colored by local resolution in Å as calculated by RELION-3.0. (D) The models used in (B) that were built in each map colored according to local B-factor.
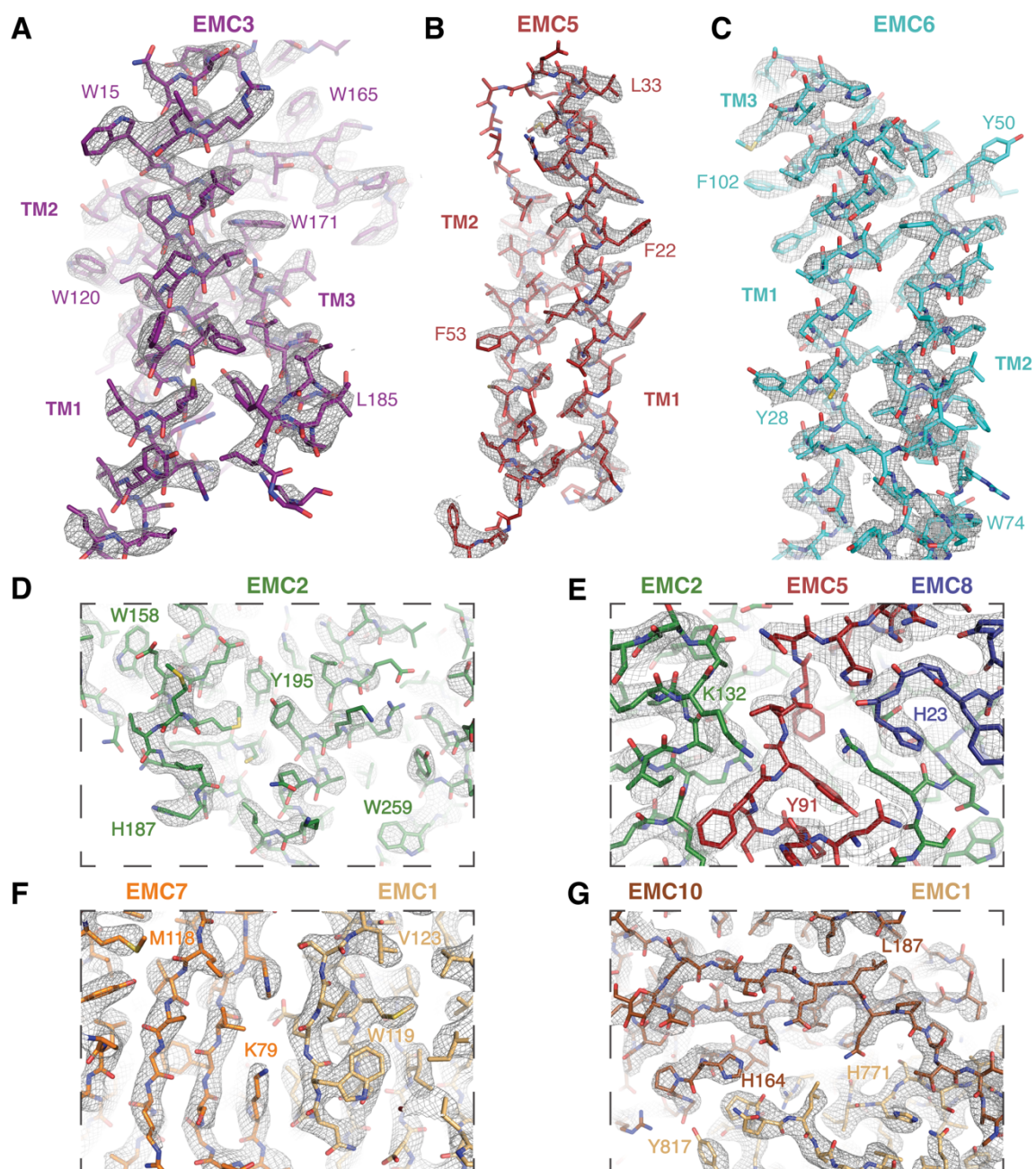
Figure S2.4 **Representative density.**

(A-C) Representative density for the TMs of EMC3 (purple), 5 (red), and 6 (cyan), shown in the sharpened 'overall map' at 11, 9, and 10σ, respectively (Fig. S2). (D+E) Representative density for

the cytosolic region of the EMC, including EMC2 (green), the C-terminus of EMC5 (red), and EMC8 (blue) shown in the sharpened 'cytosolic map' at 16σ. (F+G) Representative density for the lumenal region of the EMC including EMC1 (gold), 7 (orange), and 10 (brown) shown in the sharpened 'lumenal map' at 8σ.
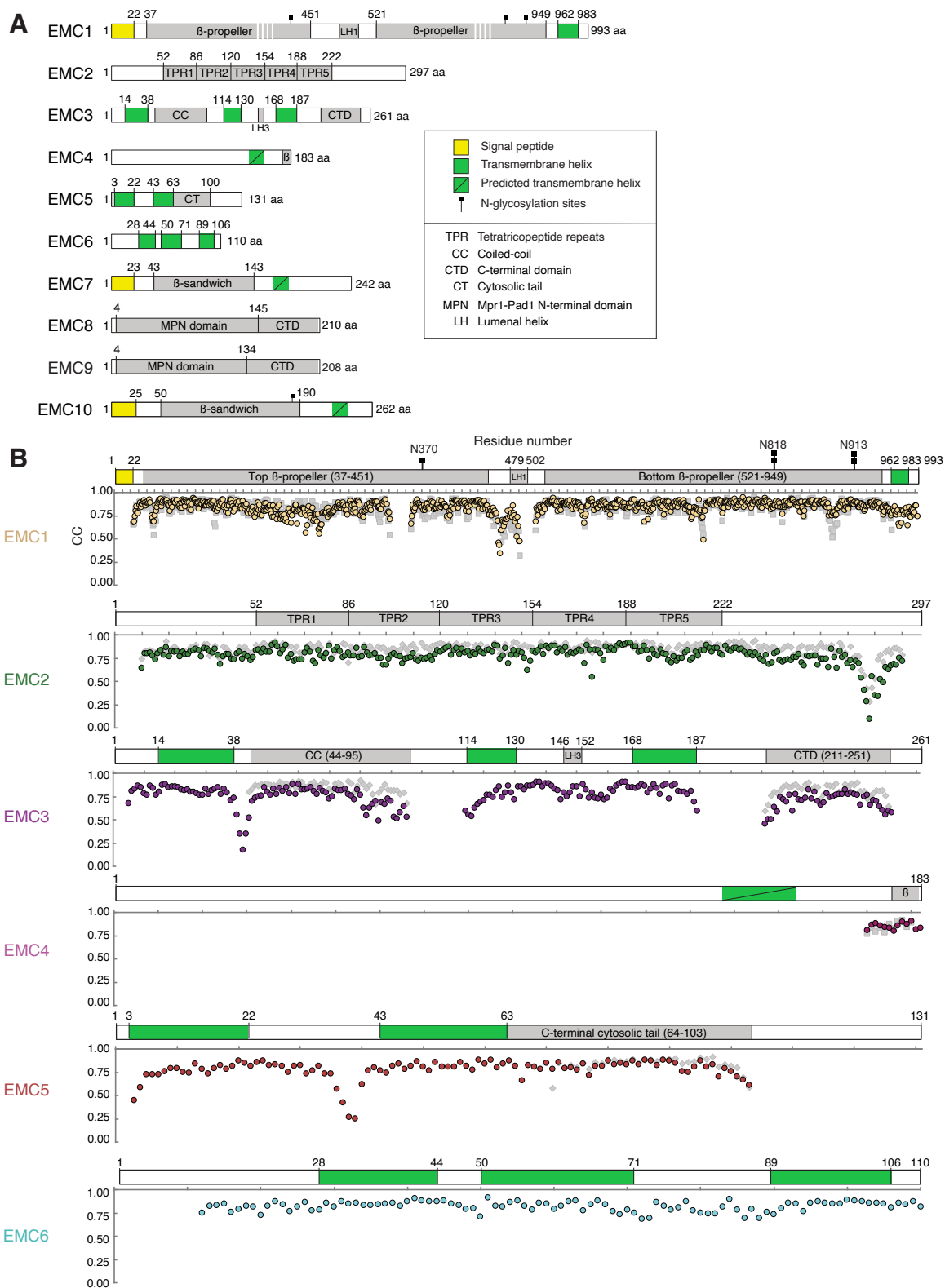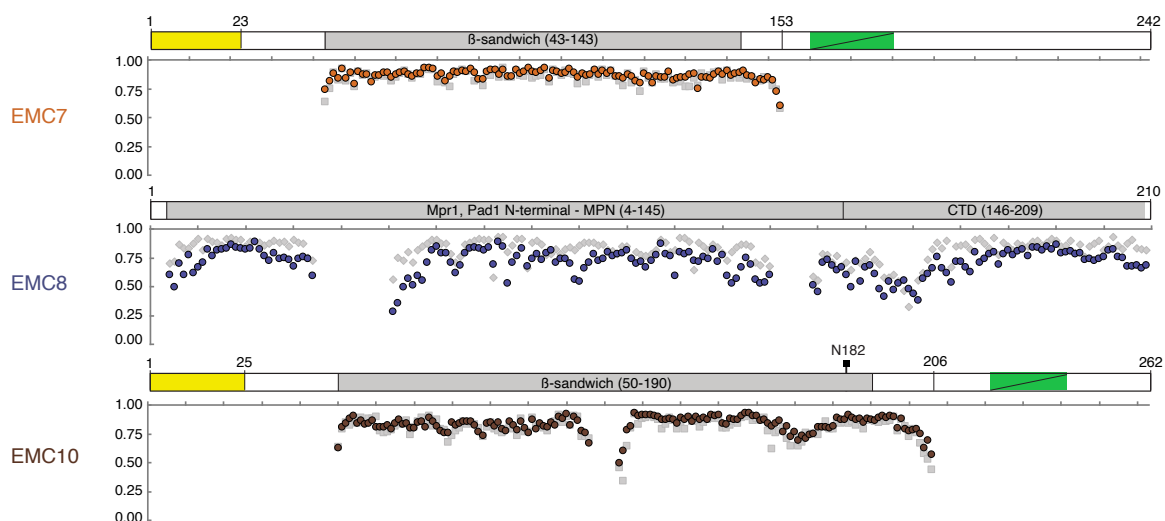
Figure S2.5 **Domain architecture and local properties the EMC model.**

(A) Schematic representation of the individual domain architecture of the EMC subunits. TPR domains in EMC2 were predicted using TPRpred (Karpenahalli et al., 2007) and validated by visual inspection. TPR domains are commonly found to mediate protein-protein interactions and in EMC2 serve to contact EMC8 and the cytosolic regions of EMC3, 5, and 1. EMC8 and EMC9 are vertebrate-specific paralogs resulting from a duplication event of an ancestral protein in the lineage leading to vertebrates (Wideman, 2015). Binding of EMC8 to EMC2 is mutually exclusive to binding of EMC9. This is consistent with previous work showing that knockdown of EMC8 upregulates the cellular levels of EMC9 (Volkmar et al., 2019). EMC8 and 9 contain Mpr1- Pad1 N-terminal (MPN) domains, commonly found in quality control machinery, and could thus provide a potential link between the EMC and the ubiquitin proteasome pathway. The cytosolic domain of the EMC seems well-positioned to mediate interactions with the ribosome or other cytosolic co-factors of the EMC. Though the function of the EMC lumenal domain remains poorly defined, the β-propellers of EMC1 are homologous to other WD40-repeat proteins, including those found in the bacterial Bam insertase (61). These domains are known to function as scaffolds for protein-protein interactions and therefore may bind lumenal co-factors. Recent proteomics experiments suggest that the EMC is a platform for

chaperone recruitment (Shurtleff et al., 2018), and human mutations to surface exposed residues in EMC1 lead to neurocognitive delays (Harel et al., 2016), potentially consistent with defects in protein folding or assembly. (B) Cross-correlation (CC) of the model for each subunit in the 'overall map' (colored circles), as well as the 'cytosolic' and 'lumenal' maps (grey diamonds and square respectively), calculated in phenix.real_space_refine. The variation in CC values illustrates the non-uniform resolution of the map (Fig. S3) and reflects the local quality of the model. It also highlights the improved local resolution of the 'cytosolic' and 'lumenal' maps. The density for the flexible linkers between the top β-propeller and lumenal helix of EMC1 (amino acids 475-478), the α-helix 14 and α-helix 15 of EMC2 (amino acids 276-279), the TM1 and the coiled-coil of EMC3 (amino acids 40-43), and between the TM1 and TM2 of EMC5 (amino acids 37-39) was not sufficient to unambiguously assign the position of the side chains, thus they are included in the model as poly-Ala, yielding a local low cross-correlation value. EMC4 was initially built and refined as P170 to L183 (Figs. S2.7, S2.8), yielding an overall cross-correlation of 0.79. However, fourteen residues are not sufficient to unambiguously assign the registry; therefore, in the model the region is assigned as poly-Ala/Gly.
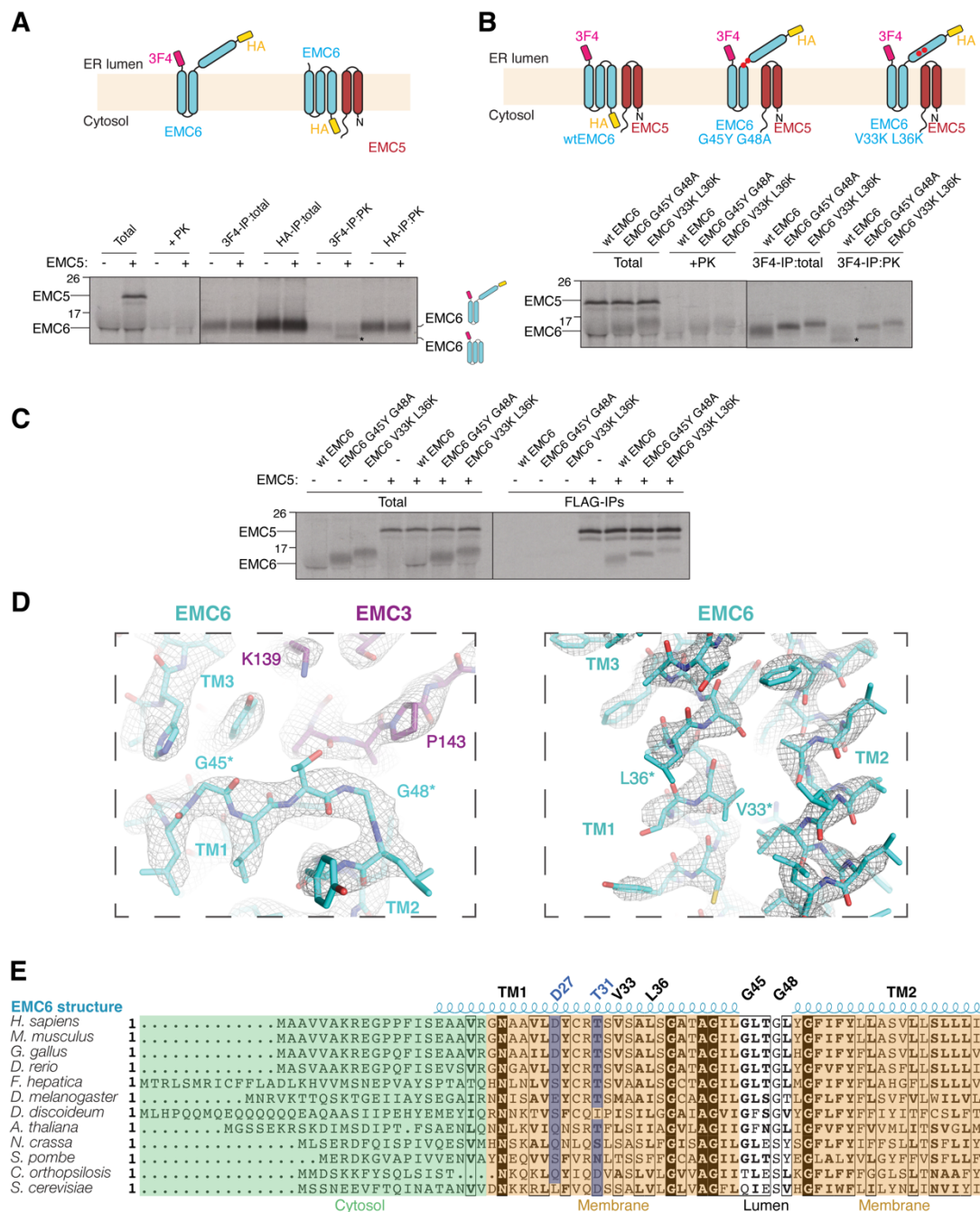
Figure S2.6 **EMC6 N-terminus forms a TM in the presence of EMC5.**

(A) 35S-methionine-labeled HA-EMC6-3F4 was translated in rabbit reticulocyte lysate (RRL) in the presence of canine-derived rough microsomes (cRMs) either alone or with EMC5-FLAG. The translations were treated with proteinase K (PK) and then analyzed directly, or following immunoprecipitation via the 3F4 or HA tag, by SDS-PAGE and autoradiography. Without EMC5, EMC6 inserts with the predicted topology, with two transmembrane helices and both termini in the lumen. The short loop between TMs 2 and 3 is protected from cleavage by the membrane and so the complete protein (with both tags) is recovered after PK treatment. Upon co-translation with EMC5, a second band appears (*) that is 3F4-tagged, but not HA-tagged. This represents an EMC6 species in which TM1 is now inserted into the bilayer resulting in the loss of the N-terminal HA-tag upon PK treatment. A schematic is shown above. (B) To validate this model, two conserved glycine residues in the TM1-2 loop (G45Y and G48A) or two hydrophobic residues in TM1 (V33K and L36K) were mutated and tested in the protease protection assay described in (A), in the presence of EMC5. In both cases we see a reduction in TM1 insertion compared to wild type (wt:*), suggesting both the flexibility of the TM1-2 loop and the hydrophobicity of TM1 are required for EMC5-dependent insertion. A schematic is shown above; the positions of the mutations are indicated with red circles. (C) To ascertain whether the inability of the EMC6 mutants to insert their first TM was due to inhibition of the interaction with EMC5, 35S-methionine-labeled EMC6 variants (wild type, G45Y G48A, or V33K L36K) were translated in the presence or absence of EMC5-FLAG, and then FLAG-tagged complexes were immunoprecipitated. Both EMC6 mutants were recovered with EMC5-FLAG at similar levels to wild type. (D) Density for EMC6 TM1 and TM2 in the sharpened 'overall map' at 11σ. Asterisks indicate residues that were mutated. (E) Sequence alignment of the N-terminus of EMC6 from higher to lower eukaryotes. The determined secondary structure of EMC6 for TM1 and TM2 is depicted above the alignment. Residues mutated within TM1 and the lumenal turn described in panel (B) are highlighted. Polar residues located in the hydrophilic vestibule and mutated in Fig. S2.13D are highlighted in blue. Note that the highly conserved N22 in TM1 of EMC6 is within hydrogen-bonding distance of the main chain of EMC5. This is one of the many interactions of the new TM1 of EMC6 with EMC5 that may explain how this poor TM is stabilized in the bilayer (Table S2.3).

A

EMC10

ER lumen

EMC4

EMC7

Cytosol

120°

EMC10

ER lumen

EMC4

EMC7

Cytosol

120°

B

EMC10

ER lumen

EMC4

EMC7

Cytosol

120°

EMC10

ER lumen

EMC4

Cytosol

120°

C

EMC5

EMC6

EMC7

EMC4

EMC1

EMC10

Views from ER lumen

120°

EMC1

EMC3

EMC5

EMC10

EMC6

EMC4

EMC7

180°

180°

D

Views from cytosol

EMC10

EMC7

EMC3

EMC1

EMC6

EMC4

EMC5

Hydrophilic
vestibule

180°

EMC5

EMC6

Hydrophilic
vestibule

EMC3

EMC4

EMC1

EMC7

EMC10

Figure S2.7 **Density for the dynamic TMs of EMC4, 7 and 10.**

(A) Orthogonal views of the overall map (grey) at 6σ and a low pass filtered map at 4σ, corresponding to EMC4 (magenta), EMC7 (orange), and EMC10 (brown) EM density regions. (B) Orthogonal views of the atomic model displayed as a cartoon superimposed with the low pass filtered map for EMC4, EMC7, and EMC10. EM density consistent with the C-terminus of EMC4 was tentatively built as a poly-Ala/Gly model. Continuous density could be traced from this putative EMC4 β-strand, past the lumen-facing surface of EMC3, and through the lipid region of the nanodisc. We also detect strong EM density directly below the putative TM region of EMC4 that protrudes from the nanodisc/lipid and forms a surface with the cytosolic region of EMC2. It is likely that this additional density represents the N-terminal cytosolic region of EMC4. Therefore, though we cannot definitively assign the topology of EMC4, the structure is most consistent with a model in which it contains a single TM. Weak EM density for a TM is also detected C-terminal to the residue R152 of EMC7 (orange) and this density continues into the nanodisc lipid region. EM density that is C-terminal to L205 of EMC10 (brown) can be detected that continues toward the nanodisc/lipid at very low contour (and is depicted as a dashed line). Together the TMs of EMC4,7, and 10 partially enclose the hydrophilic vestibule formed by EMC3 and 6. (C) Orthogonal views of the cross-section through the transmembrane region with the same coloring as (B), with a view from the lumen region facing toward the cytosol. (D) Orthogonal views of the cross-section through the transmembrane region with a view from the cytosol facing up toward the membrane. When viewed from the cytosol, it is clear that substrate would encounter a hydrophilic vestibule that is capped by the lumenal helix of EMC3 (LH-3) and enveloped by the TMs of EMC4,7, and 10.
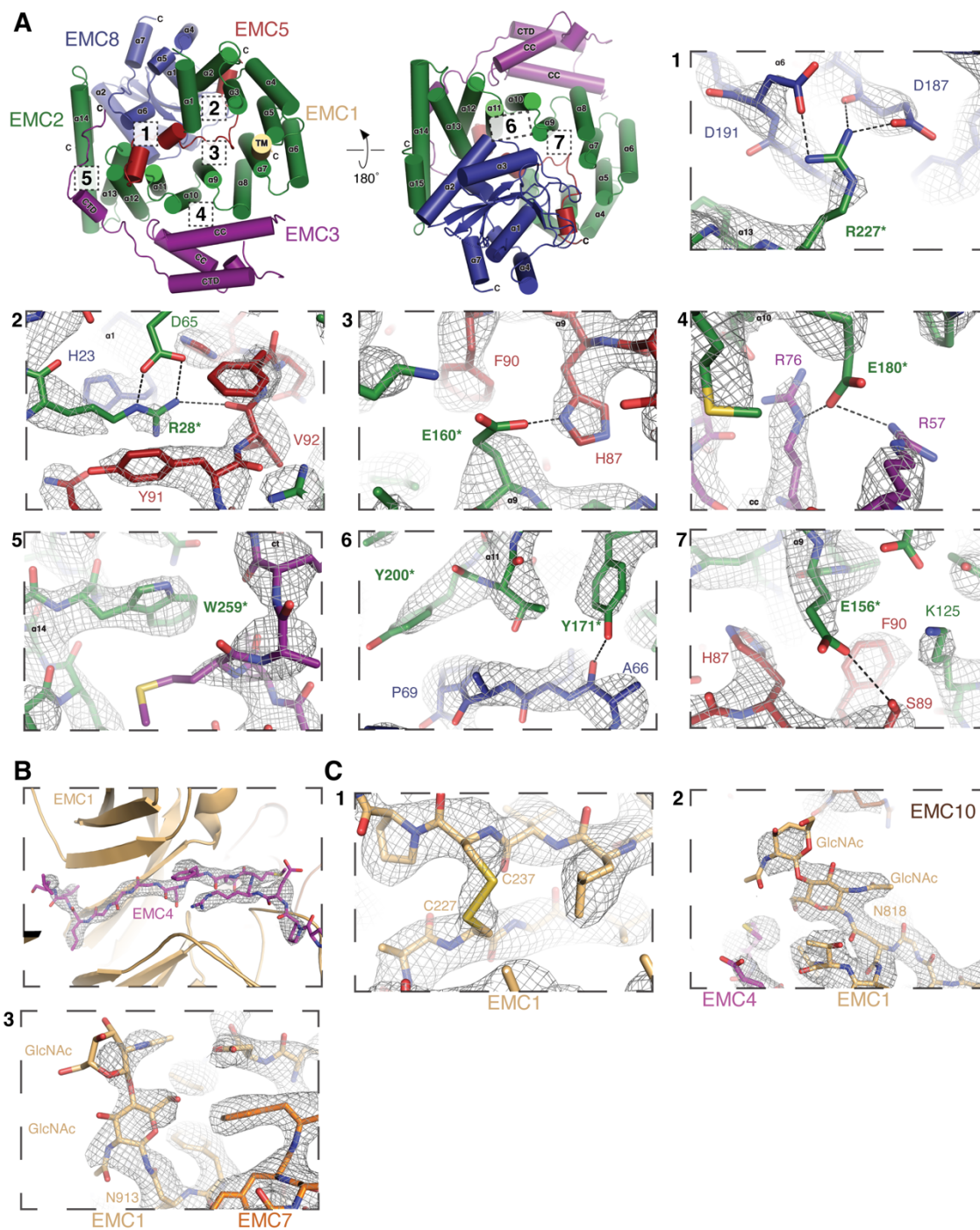
Figure S2.8 **Density for the post-translational modifications.**

(A) Top and bottom view of the cytosolic domain of the EMC. C = Carboxyl-terminus, CTD = EMC3 C-terminal domain, CC = EMC3 coiled coil, α = alpha helix, TM = transmembrane helix. EMC2 serves as the scaffold for the cytosolic region of the EMC with five tetratricopeptide repeats (α3 - α12; Fig. S5), which partly mediate protein-protein interactions with EMC8 and the cytosolic regions of EMC3, 5, and 1. Panels 1-7 display the density for the interaction surfaces between EMC2, 3, 5, and 8 in the cytosol. The views 1-4 correspond to those in Fig. 2B shown in the sharpened 'cytosolic map' at 17σ (15σ in panel 5). Asterisks indicate residues in EMC2 that when mutated disrupt binding (Fig. 2C). (B) Close-up view of the density for the lumenal region of EMC4. EMC1 contains two eight-bladed β-propellers, one of which incorporates the single β-strand of EMC4 into a four-stranded blade (Fig. 2D, E). (C) 'Lumenal map' details. Panel 1 displays a disulfide bond in EMC1 (between residues 227 and 237), shown at 8σ. Panels 2-3 show N-linked GlcNAc residues attached to N818 and N913 of EMC1 at the interface of EMC4, 7, and 10, shown at 7σ.
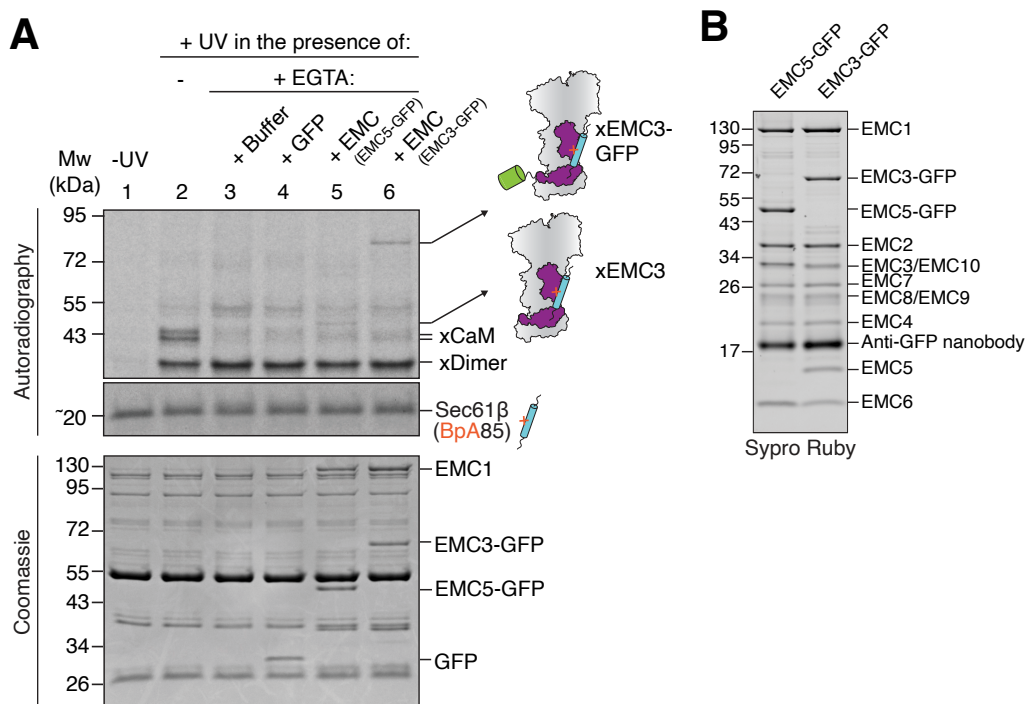
Figure S2.9 **EMC3 crosslinks to tail-anchored substrates.**

(A) SEC61β, a known EMC tail-anchored substrate, was expressed with an amber stop codon at position 85 in its transmembrane helix in the presence of radioactive 35Smethionine, 100 nM Ca^{2+} and 10 μM purified calmodulin (CaM) in the *E. coli*-derived PURE system in the absence of release factor RF1. The photoreactive unnatural amino acid 4-Benzoylphenylalanine (BpA) was incorporated at the amber stop codon using BpA synthetase and its supressor tRNA. After translation for 2 hours at 32°C, 1 mM puromycin was added, the reaction was layered on top of a 20% (w/v) sucrose cushion prepared in physiologic salt buffer with 100 nM Ca^{2+}, and spun to remove aggregates. SEC61β-CaM complexes were retrieved from the cushion and incubated in the presence or absence of 1 mM EGTA, buffer (lane 3), GFP (lane 4), or detergent solubilized EMC affinity purified from either EMC5-GFP or EMC3-GFP cell lines (lanes 5 and 6 respectively). Except for the -UV control (lane 1), all reactions were irradiated with a UV source for 15 min on ice before quenching with SDS-PAGE sample buffer. Samples were analyzed by SDS-PAGE and autoradiography (top) or Coomassie staining (bottom). The ~20 kDa substrate produces a ~50 kDa crosslink in lane 5, consistent with crosslinking to ~30 kDa untagged EMC3. In lane 6, the EMC3-substrate crosslink shifts to ~80 kDa, consistent with addition of ~30 kDa from a linker-GFP fusion. Note that the absence of crosslinks to other EMC subunits does not exclude that these are also involved in directly contacting substrate transmembrane helices. (B) EMC complexes purified via EMC5- and EMC3-GFP as detailed in Fig. S2.1 were analyzed by SDS-PAGE and Sypro Ruby staining.
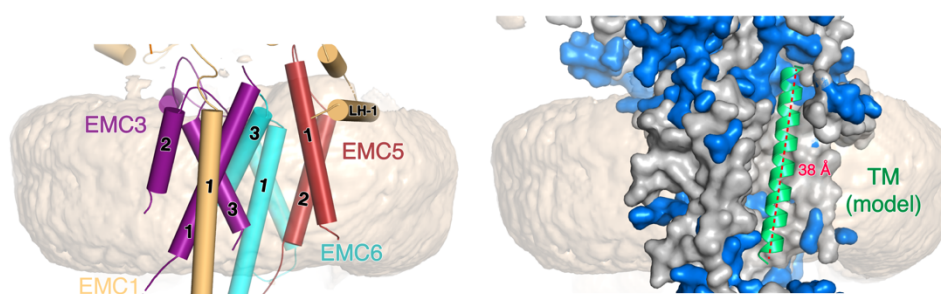
Figure S2.10 **The hydrophobic crevice can accommodate a full TM.**

The 'hydrophobic crevice' shown in cartoon (left) and surface representation (right). Both views are depicted within a subtracted map of the nanodisc/lipid region to highlight the presence of membrane. Hydrophobic amino acids (in grey) are contributed by TMs 1 and 3 of EMC3, TMs 1 and 3 of EMC6,

and TMs 1 and 2 of EMC5. Polar residues are shown in blue. The crevice is partially enclosed on the lumenal face by the lumenal helix of EMC1 (LH-1). A helix (lime) was modeled in the space-filling model to demonstrate that a 30 amino acid TM can be fully accommodated within the membrane spanning region of the hydrophobic crevice.
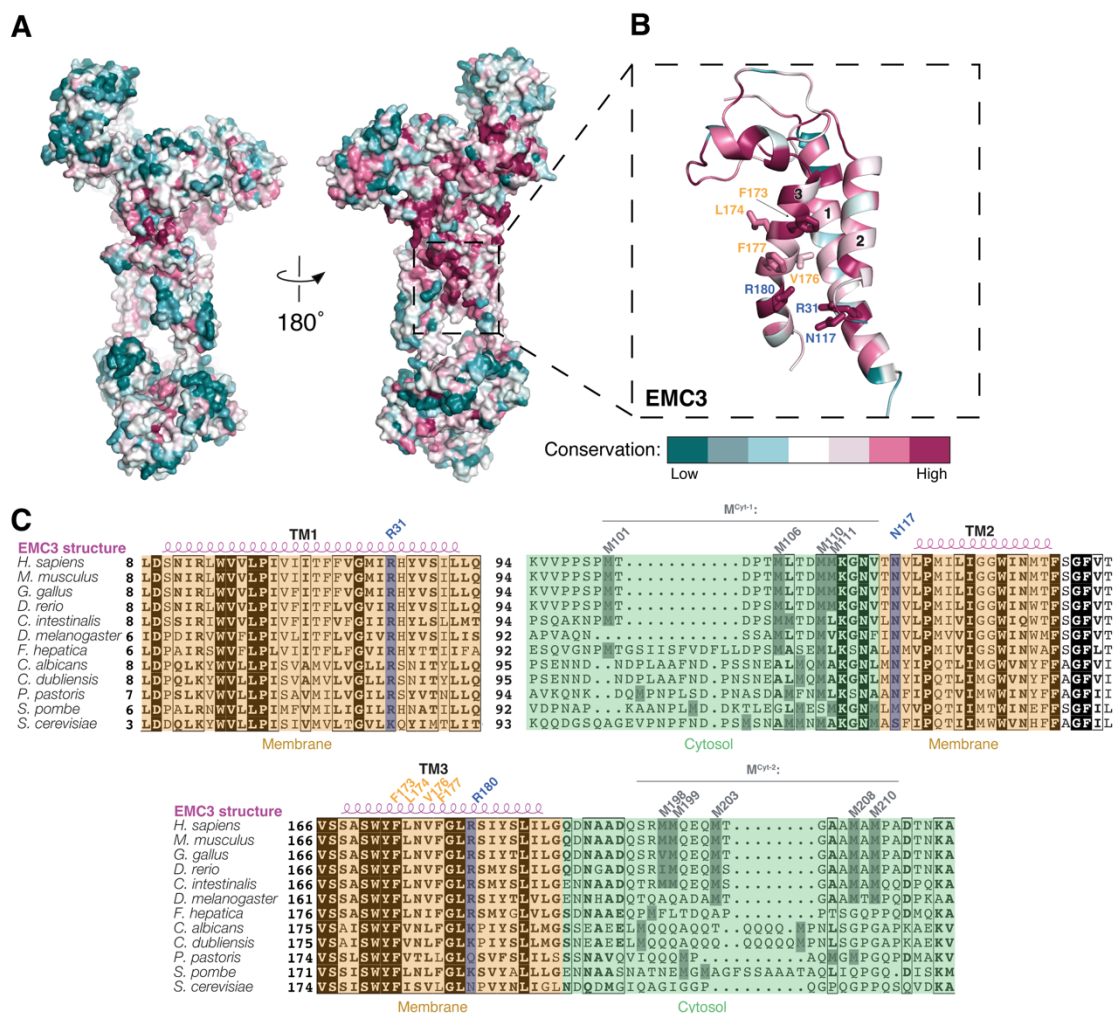
Figure S2.11 **Conservation of the hydrophilic vestibule and EMC3.**

(A) A conservation plot mapped onto the structure of EMC to highlight the high conservation of the hydrophilic vestibule (right) as compared to the less conserved hydrophobic crevice (left). Many highly conserved residues on this side are located in EMC3 (boxed region), a member of the Oxa1 superfamily of insertases. (B) Conservation is mapped on to the cartoon representation of EMC3. Highly conserved residues are displayed as sticks and include the polar residues (labeled in blue) highlighted in Fig. 2.3B and mutated in Fig. 2.3C-E and Fig. S2.13. (C) Sequence alignment of EMC3 that spans sequences derived from higher to lower eukaryotes. The panel is a sub-selection of species that were used for the conservation plot in panels A-B. The determined secondary structure of EMC3

is depicted above the alignment. The depicted regions are located in the membrane (TMs in tan) and cytosol (light green). Polar residues that were mutated are highlighted in blue. Conserved hydrophobic residues located above the polar patch are highlighted in gold. Methionine residues that comprise the 'Methionine-rich cytosolic loops' (MCyt-1: methionines 101, 106, 110, 111; MCyt-2: methionines 198, 199, 203, 208, 210) on the cytosolic side of TM2 and TM3 are highlighted in grey.
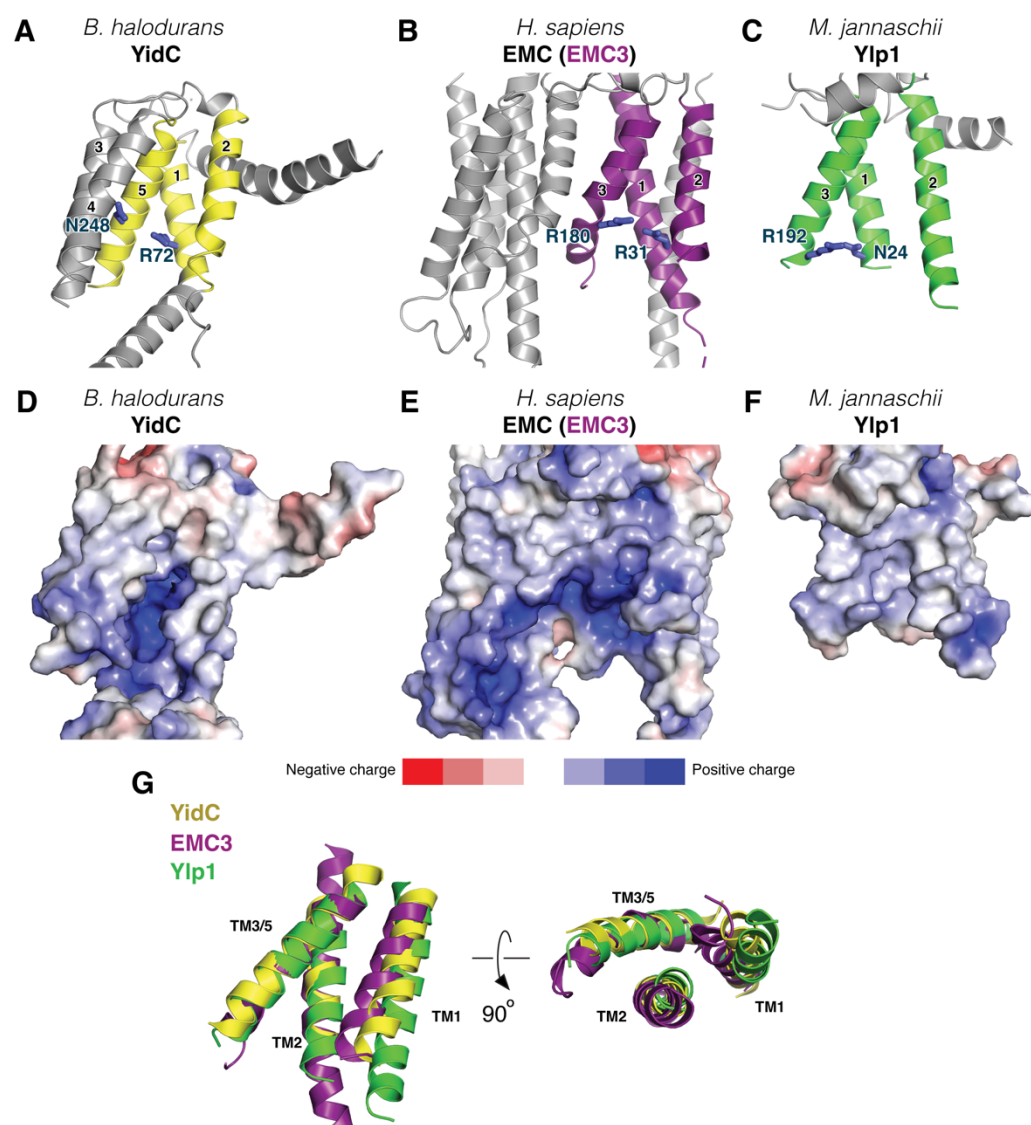
Figure S2.12 **Comparison of EMC3 to YidC and Ylp1.**

(A-C) Secondary structure alignment of the putative insertase regions of YidC (TM1, TM2, TM5 in yellow, panel A), EMC3 (TM1, TM2, TM3 in deep purple, panel B), and Ylp1 (TM1, TM2, TM3 in green, panel C) in the context of their entire scaffold (grey). Polar residues implicated in the insertase function are depicted as blue sticks. (D-F) Electrostatic plots (generated using PDB2PQR and APBS with range -5 to +5 kT/e) mapped onto the surface of structures of *H. sapiens* EMC3, B. *halodurans* YidC (3WO6) (Kumazaki et al., 2014), and *M. jannaschii* Ylp1 (5C8J) (Borowska et al., 2015). Surface representations are shown in a similar orientation to panels A-C to highlight the positively charged vestibule. Seven polar residues can be found in EMC3 on the insertase side. (G) Orthogonal views of the superposition of all three insertases using the 'super' algorithm in PyMol.
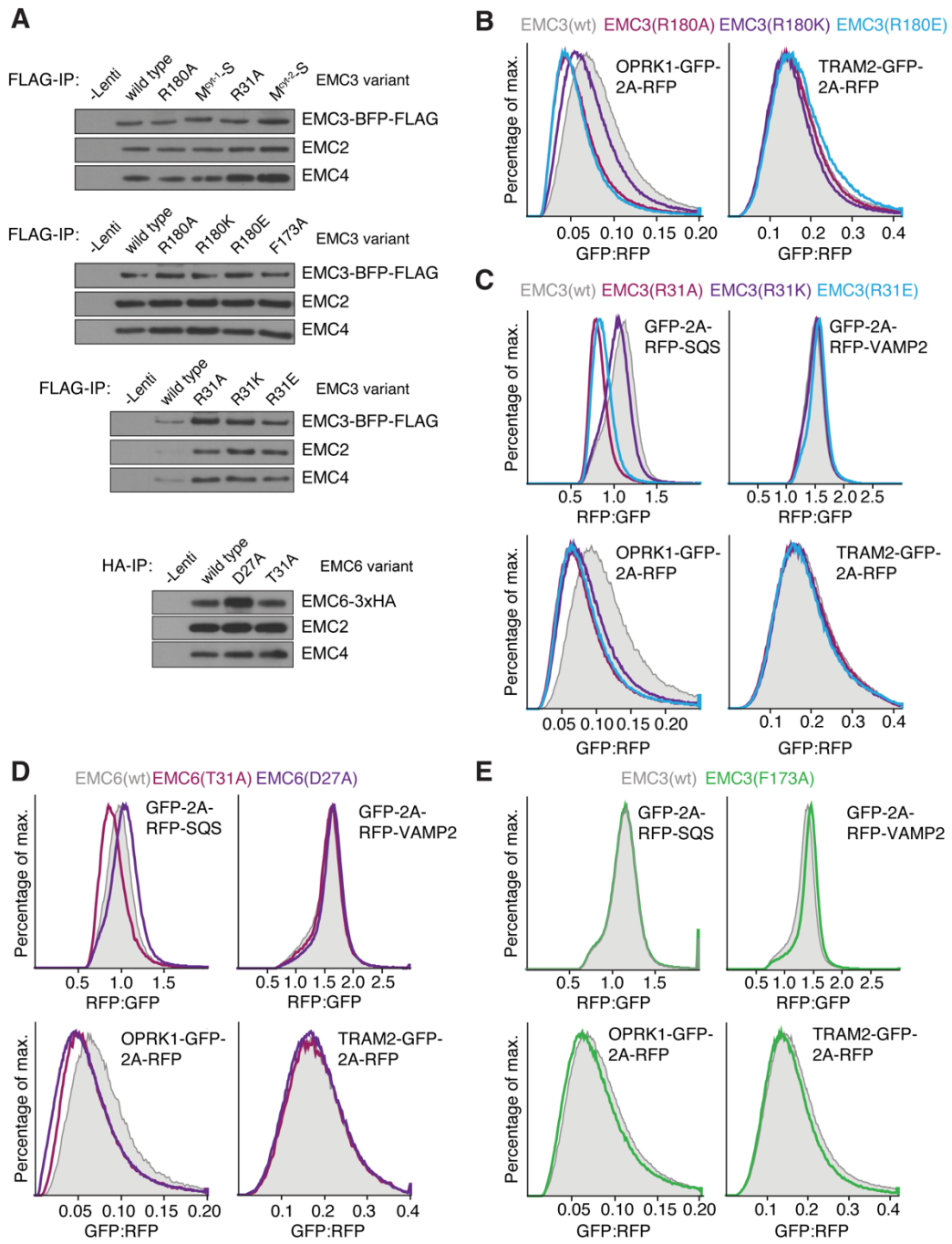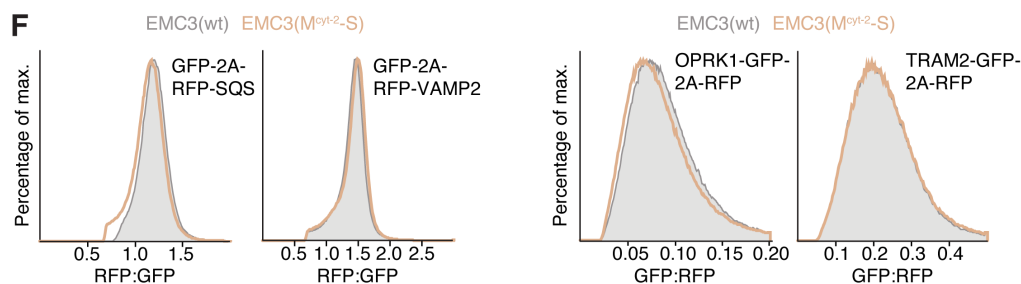
Figure S2.13 **Functional analysis of EMC3 and 6 mutations.**

(A) EMC3 and EMC6 mutants incorporate efficiently into the EMC. HEK293 cells were transduced with lenti-viral vectors encoding either wild type or the indicated EMC3 mutants with a C-terminal TagBFP-3xFLAG tag. Wild type or mutant EMC6 was expressed with a C-terminal 3xHA-tag, followed by a viral P2A sequence and TagBFP. After 72 hours, cells were solubilized in detergent and subjected to FLAG or HA-IP. Bound proteins were eluted with 3xFLAG or 3x HA peptide and analyzed by SDS-PAGE and Western blotting with the indicated antibodies. Note that all shown EMC3 and EMC6 mutants co-purify soluble (EMC2) and other membrane-bound subunits (EMC4) to a similar extent as wild type. MCyt-1-S: mutation of methionines 101, 106, 110, 111. Though EMC2 could potentially contain a hydrophobic groove, in analogy to GET3, no Met-rich or otherwise hydrophobic structured binding site could be identified in the structure. In line with a previous observation (Borowska et al., 2015), flexible Met-rich loops can be found adjacent to the TMs of EMC3. We found that Met-rich loop MCyt-1 was functionally important in cells (Fig. 2.3C), whereas MCyt- 2 seemed to be dispensable (see F). (B) HEK293 cells stably expressing GFP-tagged OPRK1 (co-translational EMC substrate) or TRAM2 (EMC-independent multipass membrane protein) were transduced with lenti-viral vectors to stably express BFP-tagged wild type or mutant EMC3 to analyze the role of the positive charge on R180 in insertion. The GFP:RFP ratio of BFP-positive cells was analyzed by flow cytometry and then plotted as a histogram. (C) As in (B), testing the role of the positive charge on R31 in insertion. Additionally, cell lines expressing the post-translational EMC substrate SQS or the EMC-independent control protein VAMP2 were analyzed. (D) Assay as in (C), testing the effect of the two polar residues of EMC6 in the hydrophilic vestibule. (E) As in (C). No effect of the EMC3 F173A mutation on insertion of co- or post-translational substrates was observed.

(F) As in (C). No effect of the methionine-rich loop MCyt-2-S mutant EMC3 on insertion of co- or post-translational substrates was observed.
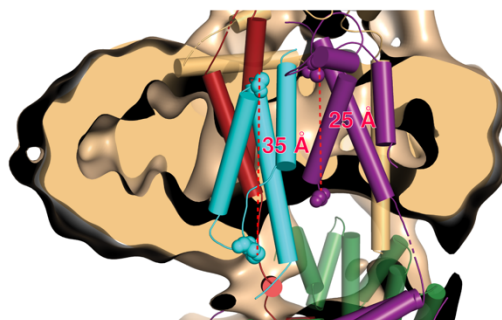
Figure S2.14 **Distance measurements of the lipid/nanodisc thickness.**

Cartoon representation of the TM region of EMC with a cut-away view of a low-pass filtered, unsharpened overall EM density map (contoured at 3σ). Distances (in red) were measured from residues that are coincident with the boundaries of the lipid for the largest lipid width (EMC6 amino acids G48 to K79) and shortest width (EMC3 amino acids T138 to G188).

Table S2.1 Cryo-EM data collection, refinement, and validation statistics

	Overall (EMDB-21929) (PDB 6WW7)	Lumenal (EMDB-21930)	Cytosolic (EMDB-21931)
Data collection and processing			
Microscope	FEI Titan Krios		
Voltage (kV)	300		
Camera	Gatan K3		
Energy filter	BioQuantum		
Energy filter slit width (eV)	20		
Magnification (nominal)	59,808 (130,000)		
Defocus range (μm)	-0.5 to -2.0		
Calibrated pixel size (Å/pix)	0.836		
Electron exposure (e⁻/Å²)	59.2		
Exposure rate (e⁻/Å²/frame)	1.48		
Number of frames per movie	40		
Automation software	SerialEM		
Number of micrographs	6345		
Initial particle images (no.)	1,034,250		
Final particle images (no.)	188,746	188,746	188,746
Estimated accuracy of translations (pix) (RELION)	0.87		
Estimated accuracy of rotations (°) (RELION)	1.571		
Local resolution range	2.5 – 6.5	3.0 – 6.0	2.5 – 4.0
Map resolution (Å, FSC=0.143)	3.4	3.2	3.6
Refinement			
Software (*phenix.real space refine*)	PHENIX 1.17.1-3660		
Initial model used (PDB code)	de novo	de novo	de novo
Resolution of unmasked reconstructions (Å, FSC= 0.5)	3.5	3.2	3.7
Resolution of masked reconstructions (Å, FSC= 0.5)	3.4	3.2	3.6
Correlation coefficient (CC$_{mask}$)	0.83	0.87	0.84
Map sharpening B factor (Å²)	-72	-97	-91
Model composition			
Non-hydrogen atoms	16576	9300	4872
Protein residues	2060	1167	592
Ligands	*UNK:14 NAG:6	NAG: 6	---
B factors (Å²)	min/max/mean		
Protein	21/152/65	22/101/45	51/134/81
Ligand	23/89/54	47/66/56	---
R.M.S. deviations			
Bond lengths (Å) (# > 4σ)	0.003 (0)	0.004 (0)	0.004 (0)
Bond angles (°) (# > 4σ)	0.544(6)	0.622(2)	0.538 (0)
Validation			
MolProbity score	1.89	1.93	1.90
Clashscore	6.96	5.49	7.39
Poor rotamers (%)	2.92	3.62	2.7
Cß deviations (%)	0	0	0
CaBLAM outliers (%)	1.25	1.15	0.35
EMRinger score	2.50	3.65	2.91
Ramachandran plot			
Favored (%)	97.1	96.7	97.0
Allowed (%)	2.9	3.3	3.0

Disallowed (%)	0	0	0

*UNK is the code for the unknown amino acids of EMC4 assigned as poly-Ala/poly-Gly.

Table S2.2 Model building of the EMC

EMC	Chain	Region (residue range)	Region Name	Template for model building
1	A	24-36	NTD	de novo
		37-451	Top β-propeller	trRosetta
		452-474	Coil/Helical	de novo
		479-502	LH1	de novo
		521-949	Bottom β-propeller	trRosetta
		950-961	Coil	de novo
		962-983	TM	de novo
		984-993	CTD	de novo
2	B	10-51	NTD	de novo
		52-222	Core (TPR)	trRosetta
		223-290	CTD	de novo
3	C	5-13	NTD	
		14-38	TM1	
		39-43	Coil	
		44-95	Coiled coil	de novo
		114-130	TM2	
		131-167	Lumen region	
		168-187	TM3	
		211-252	CTD	
4	D	170-183	bottom β-propeller	de novo (poly-Ala/poly-Gly)
		1-169	EMC4	sequence not modeled
5	E	3-22	TM1	
		23-42	Lumen region - Helical	de novo
		43-63	TM2	
		64-103	CT	
6	F	12-27	NTD - Helical	trRosetta / de novo
		28-44	TM1	trRosetta / de novo
		45-49	Lumen - Loop	de novo
		50-71	TM2	de novo
		72-88	Cytosol - Loop	de novo
		89-106	TM3	de novo
		107-110	CTD	de novo
7	G	43-143	β-sandwich	trRosetta
		144-153	Coil	de novo
		154-242	TM / CTD	sequence not modeled
8	H	4-162	MPN	trRosetta
		163-209	CTD	de novo
10	I	50-190	β-sandwich	trRosetta
		191-206	Coil	de novo
		207-262	TM / CTD	sequence not modeled

TM – transmembrane helix
LH-1 – lumenal helix of EMC1
NTD – amino-terminus domain
CTD – carboxyl-terminal domain
TPR – tetratricopeptide repeats
CT – cytosolic tail
MPN – Mpr1, Pad1 N-terminal

Table S2.3 Interface area between adjacent EMC subunits

Adjacent subunits of EMC	Buried area* (Å^2)
EMC 1 and EMC 2	386
EMC 1 and EMC 3	1595
EMC 1 and EMC 4	1139
EMC 1 and EMC 5	665
EMC 1 and EMC 7	1700
EMC 1 and EMC 10	1672
EMC 2 and EMC 3	1721
EMC 2 and EMC 5	1582
EMC 2 and EMC 8	1393
EMC 3 and EMC 5	76
EMC 3 and EMC 6	1053
EMC 3 and EMC 8	94
EMC 5 and EMC 6	1477
EMC 5 and EMC 8	634
EMC 7 and EMC 10	100

*Buried area indicates, in Å^2, the interface area calculated as difference in total accessible surface areas of isolated and interfacing structures divided by two (Krissinel & Henrick, 2007; Laskowski et al., 2018).

2.4 Materials and Methods

Plasmids and antibodies

The plasmid for generating the nanodisc-forming protein MSP2N2 was acquired from Addgene (ID 29520). Expression constructs for the bdSENP1 and bdNEDP1 proteases (Addgene IDs 104962 and 104129) (Frey & Görlich, 2014; Pleiner et al., 2018), as well as the SUMOEu1 module and its cognate SENPEuB protease (Addgene ID 149333) (Rodriguez et al., 2019) were kind gifts from Dirk Görlich. 2nd generation lenti-viral packaging and envelope plasmids were kind gifts from Carlos Lois.

The following antibodies were used in this study: EMC2 (25443-1-AP, Proteintech, USA); EMC4 (27708-1-AP, Proteintech, USA); EMC5 (A305-833, Bethyl Laboratories, USA); EMC6 (ab84902, Abcam, UK); EMC8 (STJ117038, St. John's Laboratory, UK); Anti- FLAG-HRP (Millipore-Sigma, USA); Anti-HA (Sigma); Anti-3F4 and Anti-BAG6 were kind gifts from Ramanujan Hegde and have been described previously (Mariappan et al., 2010).

E. coli protein expression and purification

Untagged GFP used as a negative control in Fig. S9 was expressed with an N-terminal His14-bdNEDD8 tag (Pleiner et al., 2018) in *E. coli* NEB express Iq (New England Biolabs, USA). Expression was carried out at a 1 L scale in Super Broth medium for 6 hours at 18°C using 0.2 mM IPTG for induction. Harvested cells were resuspended in lysis buffer (50 mM Tris/HCl pH 7.5, 300 mM NaCl, 20 mM imidazole, 1 mM DTT, 1 mM PMSF) and lysed by sonication (4x 1 min pulsing in thin-walled metal containers in an ice-water bath, Macro-Tip at 100% amplitude, Branson Sonifier). GFP was purified from lysate by Ni^{2+}- chelate affinity chromatography and eluted with protease elution buffer (300 nM bdNEDP1 in 50 mM Tris/HCl pH 7.5, 300 mM NaCl, 20 mM imidazole, 250 mM sucrose) to remove the N-terminal tag.

The expression and purification of GST-3xFLAG-Calmodulin and BpA-RS used for in vitro crosslinking (Fig. S2.9) was carried out as described previously (Shao et al., 2017). Expression and purification of the MSP2N2 protein was achieved as described by established protocols (Ritchie et al., 2009).

Expression and purification of biotinylated anti-GFP nanobody

GFP-tagged proteins were purified using a previously described native single-step strategy (Pleiner et al., 2015) (see Fig. S2.1B, C). It relies on a multifunctional tag that allows biotinylation and immobilization of nanobodies onto magnetic Streptavidin beads. A specific protease cleavage module in the fusion tag further enables gentle release of the nanobody along with its bound target complex on ice in physiological buffer, while avoiding elution of unspecific background binders. See Figs. S2.1C and S2.9B for representative examples. Magnetic beads are preferred over porous beads, as they show higher capacity for larger protein complexes and also enable elution in a very small volume.

An *E. coli* expression vector (pTP396, Addgene ID 149336) encoding the anti-GFP nanobody "GBP1/Enhancer" (Kirchhofer et al., 2010) fused to an N-terminal His14-Biotin acceptor peptide (Avi)-SUMOEu1-tag was generated (Pleiner et al., 2015; Rodriguez et al., 2019). The Avi-tag is a 15 amino acid peptide with a single lysine that can be specifically biotinylated using purified *E. coli* biotin ligase BirA (Beckett et al., 2008; Fairhead & Howarth, 2015). The engineered SUMOEu1 module is resistant to cleavage by endogenous deSUMOylases in eukaryotic cell lysates, enabling stable isolation of nanobody:target complexes from eukaryotic extracts (Pleiner et al., 2015; Rodriguez et al., 2019). The corresponding engineered SENPEuB protease cleaves SUMOEu1 very efficiently at low concentration on ice. Alternatively, a commercial system can be used, based on the SUMOStar tag and SUMOStar protease (LifeSensors, USA)(Liu et al., 2008).

The nanobody fusion protein was expressed in *E. coli* NEB Express Iq cells. A 100 ml SuperBroth pre-culture was grown overnight at 28°C and diluted to 1 L after ~18 hours. After dilution, the culture was incubated at 18°C until an OD600 of ~2.0 (2-4 hours). Expression was then induced by addition of 0.2 mM IPTG for 18-20 hours at 18°C. Cells were harvested and the pellet was resuspended in 120 ml lysis buffer (50 mM Tris/HCl pH 7.5, 300 mM NaCl, 20 mM imidazole, 1 mM DTT, 1 mM PMSF) and lysed by freeze thaw followed by sonication (4x 1 min pulsing in thin-walled metal containers in an ice water bath, Macro-Tip at 100% amplitude, Branson Sonifier). The lysate was cleared by centrifugation for 45 min at 17,000 rpm and 4°C in an SS-34 rotor. Half of the lysate was frozen after addition of 250 mM sucrose for later purifications. The other half (60 ml) was incubated in batch with ~4 ml bed volume Ni^{2+}-NTA agarose (Thermo Fisher Scientific, USA) for 1 hour at

4°C and then transferred to a gravity flow column. The resin was washed with three column volumes of lysis buffer before stepwise elution with imidazole elution buffer (50 mM Tris/HCl pH 7.5, 300 mM NaCl, 500 mM imidazole, 1 mM DTT, 250 mM sucrose). The buffer of the eluate was then exchanged to storage buffer (50 mM Tris/HCl pH 7.5, 200 mM NaCl, 1 mM DTT and 250 mM sucrose) using a PD-10 desalting column (GE Healthcare, USA). The purified protein was aliquoted, snap-frozen in liquid nitrogen, and stored at -80°C until further use.

Purification of E. coli biotin ligase BirA and nanobody biotinylation

The expression of biotin ligase BirA from *E. coli* and its use for biotinylation of Avi-tagged proteins has been described extensively before (Beckett et al., 2008; Fairhead & Howarth, 2015) (commercial products by Avidity, USA). His14-bdNEDD8-tagged BirA Biotin ligase was expressed from pTP264 (Addgene ID 149334) in *E. coli* NEB express Iq (New England Biolabs, USA) for 18-20 hours at 18°C in 1 L SuperBroth using 0.2 mM IPTG for induction. After cell harvest, lysis, and binding to Ni^{2+}-resin as described above, tagged BirA was eluted with imidazole elution buffer. The buffer was exchanged to storage buffer as described above. The purified protein was aliquoted, snap-frozen in liquid nitrogen, and stored at -80°C until further use.

An in-solution biotinylation reaction contained 300 μl 5x biotinylation buffer (250 mM Tris/HCl, 500 mM NaCl, 62.5 mM MgCl2, 50 mM ATP, 50 mM biotin), at least 50 μM purified Avi-tagged anti-GFP nanobody (from pTP396), a 1:50 molar ratio of purified His tagged BirA, and water to add up to 1.5 ml final volume. The reaction was incubated for 3 hours at 25°C and then applied to a PD-10 desalting column (GE Healthcare, USA) equilibrated in storage buffer to remove excess biotin. Fractions with normal 260/280 ratio (comparable to starting ratio of the prep, usually ~0.6-0.7) were pooled and quantitative biotinylation was assessed by test binding to Streptavidin beads. If fully biotinylated and added below bead capacity, all of the nanobody should be in the bound fraction and almost none should remain in the unbound fraction.

Expression of bdSENPEuB protease

His14-Tev-tagged bdSENPEuB protease (Addgene ID 149333) (Rodriguez et al., 2019) was expressed in *E. coli* NEB express Iq as described in the '*E. coli* protein expression and purification'

section and purified by Ni^{2+}-chelate affinity chromatography using imidazole elution. The buffer of the eluate was then exchanged to storage buffer as described above.

Mammalian in vitro translation

Cell free in vitro translations were carried out in rabbit reticulocyte lysate (RRL) as previously described (Sharma et al., 2010). Constructs for expression in RRL were based on the SP64 vector (Promega, USA). Briefly, templates for in vitro transcription were prepared by PCR and contained a 5' SP6 promoter, an in-frame stop codon, and a short 3' UTR. PCR products were purified and used for in vitro transcription (1.5 hours at 37°C) and translation (20-30 min at 32°C) in the presence of radioactive 35S-methionine as described (Sharma et al., 2010). Translation of membrane subunits was performed in RRL supplemented with nucleased canine pancreatic microsomes (cRM), prepared as previously described (Walter & Blobel, 1983), at a ratio of 0.5 μl of cRM per 10 μl translation reaction. Native immunoprecipitations were carried out for soluble complexes in physiologic salt buffer (PSB) (50 mM HEPES pH 7.5, 130 mM KAc, 2 mM MgAc$_2$, 1 mM DTT) using Anti-FLAG M2 affinity resin or Anti-HA Agarose (both Millipore-Sigma, USA).

Immunoprecipitation of membrane subunits was carried out as follows. Following the translation reaction, membranes were isolated by centrifugation (20 min at 55,000 rpm in a Beckman TLA55 rotor) through a 25% (w/v) sucrose cushion in PSB. Following resuspension, complexes were solubilized in 1.75% (w/v) digitonin in PSB on ice for 10 min. Insoluble material was removed by centrifugation for 15 min at 20,000 x g, and the supernatant was incubated with the appropriate resin for 1.5 hours at 4°C. The resin was washed with PSB supplemented with 0.25% (w/v) digitonin, and eluted with either 3xFLAG or 3xHA peptide. Complexes were analyzed by SDS-PAGE and autoradiography.

Protease digestions were performed for one hour on ice by the addition of 0.5 mg/ml proteinase K to translation reactions. The digestion was quenched by the addition of 5 mM PMSF in DMSO, followed by transfer to boiling 1% (w/v) SDS in 0.1 M Tris/HCl pH 8.0 (pH determined at room temperature). Immunoprecipitation of protected fragments was performed in IP buffer (50 mM HEPES/KOH pH 7.5, 100 mM KAc, 2 mM MgAc$_2$, and 1% (v/v) Triton X-100) using the 3F4 or HA antibody and protein G resin. Samples were subsequently analyzed by SDS-PAGE and autoradiography.

In vitro translation in the PURE System and photocrosslinking

Site-specific UV crosslinking was carried out as previously described (Shao et al., 2017). Constructs for translation in the PURE system were derived from the T7 promotor based PURExpress plasmid provided by New England Biolabs (USA). Human Sec61β with residue Phe85 in its transmembrane helix mutated to an amber stop codon, was translated in the presence of 35S-methionine, 100 nM $CaCl_2$, and 10 μM purified Calmodulin (CaM) in the coupled transcription/translation PURE system (New England Biolabs, USA). This translation extract contains a defined set of purified *E. coli* ribosomes and translation factors and lacks all release factors (ΔRF). The release factors RF2 and RF3, but not RF1 (which recognizes the UGA [amber] stop codon) were added back to the reaction. The unnatural amino acid BpA (100 μM) was incorporated at the amber stop codon using recombinant BpA synthetase (100 μg/ml) and suppressor tRNA purified as described (Shao et al., 2017). After translation for 2 hours at 32°C and addition of 1 mM puromycin, the reaction was layered on top of a 20% (w/v) sucrose cushion prepared in PSB with 100 nM $CaCl_2$ and spun for 1 hour at 55,000 rpm in a TLS-55 rotor (Beckman Coulter, USA) to remove aggregates. Sec61β- CaM complexes were retrieved from the cushion and incubated in the presence of 1 mM EGTA with either buffer, purified GFP, or EMC purified via EMC5-GFP or EMC3-GFP from the respective stable HEK293 cell lines (described below). Except for the -UV control lane, all reactions were irradiated at a distance of ~7-10 cm with a UVP B-100 series lamp (Analytik Jena, Germany) for 15 min on ice, before quenching with SDS-PAGE sample buffer. Samples were analyzed by SDS-PAGE and autoradiography. Comparable crosslinking efficiency was observed in at least ten independent replicates. A representative gel is shown in Fig. S2.9.

Generation of stable HEK293 cell lines

Flp-In 293 T-Rex cells were purchased from Thermo Fisher Scientific (USA) (RRID: CVCL U421). Cell lines were grown in DMEM supplemented with 2 mM glutamine, 10% (w/v) FBS, 15 μg/ml Blasticidine S, and 100 μg/ml Zeocin. The open-reading frame to be integrated into the genomic FRT site was cloned into the pcDNA5/FRT/TO vector backbone and cell lines were generated according to the manufacturer's protocol. Briefly, the open-reading frame was cloned into the pcDNA5 backbone and transfected together with pOG44 Flp-In recombinase in a 9:1 ratio using Trans-IT 293 transfection reagent (Mirus, USA) according to the manufacturer's instructions. 48 hours after

transfection, 100 μg/ml Hygromycin B was added to select for cells that had undergone successful integration. This procedure was used to generate cell lines stably expressing GFP-EMC2- 2A-RFP, EMC5-GFP-2A-RFP, EMC3-GFP-2A-RFP, OPRK1-GFP-2A-RFP, and TRAM2- GFP-2A-RFP. Constructs for generating cell lines expressing the RFP-tagged transmembrane helix and flanking regions of human squalene synthase (SQS/FDFT1) or vesicle-associated membrane protein 2 (VAMP2) were prepared as previously described (Guna et al., 2018).

To facilitate large-scale growths, the adherent cell line expressing GFP-EMC2 was adapted to grow in suspension. This was achieved over the course of ~10 days, during which the FBS-supplemented DMEM was serially diluted with FreeStyle 293 Expression Medium (Thermo Fisher Scientific, USA). Once growing in 100% FreeStyle Medium, the cells were transferred to 1-2 L roller bottles (Celltreat, USA) and grown in a shaking incubator operating at 8% CO_2 and rotating at 125 rpm.

Flow cytometry analysis of reporter cell lines

To measure protein stability in cells, stable cell lines were generated as described above containing a fusion of a protein-of-interest with either GFP or RFP in a GFP-2A-RFP backbone (Itakura et al., 2016). Note, the mCherry sequence is used throughout, but is referred to as RFP for simplicity. The viral 2A sequence induces skipping of a peptide bond by the ribosome (De Felipe et al., 2006), resulting in two separate polypeptide chains being produced from a single transcript in a 1:1 ratio. The GFP and RFP fluorescence intensity can be measured by flow cytometry to derive a GFP:RFP or RFP:GFP ratio. Any changes in the GFP:RFP ratio (for OPRK1 and TRAM2) or RFP:GFP ratio (for SQS and VAMP2) reflect changes in the posttranslational stability of the fused substrate or control protein.

To test the effect of EMC subunit mutations on EMC function in cells, lenti-viral vectors in a pHAGE2 backbone were generated that stably integrate and express either wild type or mutant subunit from a CMV promotor-driven open-reading frame. EMC3 variants were fused to a C-terminal TagBFP-3xFLAG tag, while EMC6 variants carried a C-terminal 3xHA tag followed by a 2A sequence and TagBFP. A 2nd generation packaging system was used to generate lenti-viral particles. Reporter cell lines expressing GFP or RFP-fused EMC substrates or control proteins were then transduced with wild type or mutant EMC subunit lenti-viral vectors. BFP-positive cells were selected

for analysis of the effect of the mutants on the GFP:RFP or RFP:GFP ratio of the underlying reporter. The effect of mutation of the following residues was tested in at least three independent replicates in all cell lines: EMC3 MCyt-1-S, R31, and R180. Representative data is shown.

In order to verify incorporation of mutant EMC subunits into the intact EMC, an aliquot of the analyzed cells was resuspended in solubilization buffer (50 mM HEPES/KOH pH 7.5, 200 mM NaCl, 2 mM MgAc$_2$, 1% (w/v) DDM [Anatrace, USA], 1 mM DTT, 1x complete EDTA-free protease inhibitor cocktail [Roche, Germany]). After incubation on ice for 30 min, the lysate was centrifuged for 10 min at 18,000 x g at 4°C in a table-top centrifuge. The cleared detergent lysate was then subjected to FLAG-IP and elution with 3xFLAG peptide for EMC3 or HA-IP and elution with 3xHA peptide for EMC6. Eluates were analyzed by SDS-PAGE and Western blotting.

Purification of EMC for structure determination

2 L of HEK293 cells stably expressing GFP-EMC2-2A-RFP were grown in suspension and induced for 48 hours with 1 μg/ml doxycycline (DOX). Cells were harvested, washed with 1x PBS, weighed and resuspended with 6.8 ml solubilization buffer per 1 g cell pellet (50 mM HEPES/KOH pH 7.5, 200 mM NaCl, 2 mM MgAc$_2$, 1% (w/v) DDM [Anatrace, USA], 1 mM DTT, 1x complete EDTA-free protease inhibitor cocktail [Roche, Germany]). After 1 hour of head-over-tail incubation with solubilization buffer, the lysate was cleared by centrifugation for 40 min at 4°C and 30,000 x g in a Sorvall RC6+ centrifuge. In parallel, beads with immobilized nanobody were prepared similarly as described before (14). Briefly, 60 μl resuspended Pierce Streptavidin magnetic beads (Thermo Fisher Scientific, USA) per 1 g cell pellet were pre-equilibrated in wash buffer (solubilization buffer with 0.015% (w/v) DDM). Then 20 μg biotinylated His14-Avi-SUMOEu1-tagged anti-GFP nanobody (see above for expression protocol) per 60 μl beads were immobilized for 20 min on ice with occasional mixing. After this, all remaining biotin binding sites were blocked by incubation with 100 μM biotin in 50 mM HEPES/KOH pH 7.5 for 5 min on ice with occasional mixing. The beads are then washed with wash buffer and incubated with the cleared detergent cell lysate for 1 hour, binding head-overtail at 4°C. The beads were washed four times with 1 ml wash buffer and finally, the anti-GFP nanobody along with all bound proteins was released under native conditions and in minimal volume (~20-30 μl for 60 μl beads) by cleavage with 250 nM SENPEuB protease (Rodriguez et al., 2019) in wash buffer for 30 min at 4°C with occasional mixing.

The eluted EMC was reconstituted into lipid nanodiscs as follows. His6-tagged MSP2N2 (Addgene #29520) was expressed in *E. coli* BL21(DE3) cells and purified as previously described (Ritchie et al., 2009). DDM solubilized EMC at ~5 μM concentration was incubated on ice with POPC 16:0-18:1 lipids (Avanti, USA) for 30 min, after which MSP2N2 was added to a final ratio of EMC:MSP2N2:POPC of 1:4:300. The reaction was incubated for 2 hours on ice. Bio-Beads SM-2 (Bio-Rad Laboratories, USA) were washed sequentially in methanol, water, and finally equilibrated in wash buffer without detergent. The prepared Bio-Beads were added to the EMC:MSP2N2:POPC mixture and incubated for 18 hours at 4°C with continuous shaking. Bio-Beads were removed by centrifugation, and the reconstituted complex was purified via size exclusion chromatography on a 3.5 ml Superose 6 column (GE Lifesciences, USA) (Fig. S2.1). The purified reconstituted complex was concentrated to 0.2 mg/ml using an Amicon Ultra 0.5 ml 100K MWCO concentrator (Millipore-Sigma, USA), and 0.05% (w/v) CHAPSO (Millipore-Sigma, USA) was added immediately prior to vitrification.

Grid preparation and data collection

3 μl of sample was applied to Holey carbon grids (Quantifoil R1.2/1.3) that were glow discharged using a Pelco easiGlow, Emeritech K100X at a plasma current of 20 mA for 60 seconds in air. The grid was blotted at 6°C, 95% humidity with filter paper (Whatman No.1) for 3-5 seconds at a -4 blot force before plunging into liquid ethane using the FEI Vitrobot Mark v4 x2. Data were collected on an FEI Titan Krios equipped with an energy filter (20 eV slit width) operated at 300 keV and a K3 (Gatan) direct detector. Images were recorded using an automated acquisition pipeline in SerialEM (Mastronarde, 2005). Illumination conditions were adjusted in nanoprobe mode to a fluence of 13 e-/pixel/second. Images were recorded between -2.0 and -0.5 μm defocus in super resolution mode at a calibrated magnification of 0.418 Å/pixel. Two-second images with a frame width of 50 ms (1.48 e-/Å2/frame) were collected in counting mode.

Image processing

6,345 movies were processed using cryoSPARC v2.13.2 to generate dose-weighted aligned and averaged micrographs alongside aligned movies (Punjani et al., 2017). Of the 6,345 movies acquired, a sub-set of 6,277 movies were selected for further processing based on a high CTF Figure of Merit

score and an estimated resolution better than 4.5 Å. Using the automated Blob Picker function, particles were picked and then manually filtered to remove obvious debris. The remaining particle stack of 440,033 particles was further filtered by an initial round of 2D classification to remove junk particles (Fig. S2 C). The resulting 66,137 particles were then used in cryoSPARC to generate six ab initio 3D reconstructions. Only one class was consistent in size with the EMC in a nanodisc (27,313 particles), while the other classes had very small sizes more consistent with broken complexes or empty nanodisc. The putative EMC-nanodisc class was then refined, new templates were generated from the initial 3D model, and particles were re-picked to generate a new more homogeneous particle stack of 1,034,250 particles. Multiple rounds of 3D heterogeneous classification were employed to remove junk particles. A final round of classification was employed on a population of 213,610 particles that produced the 188,746 particles that were used for structural determination. Upon analysis of the EM density maps, it became clear that there was weak density that flanked one side of the putative TM region. Therefore, an additional round of 3D heterogeneous classification/refinement was performed that produced a population of 42,506 particles that had stronger EM density within this region with a worse apparent overall resolution (Fig. S2.7).

The 188,746 pool of particles were exported from cryoSPARC using the csparc2star.py, (UCSF pyem v0.5. Zenodo) suite of conversion scripts for additional processing in RELION-3.0 (Fig. S2.2D) (Zivanov et al., 2018). Following an initial 3D refinement, CTF refinement was performed, resulting in an overall EM density map to 3.4 Å resolution (Fig. S2.3). To address the apparent flexibility between the lumenal, membrane, and cytoplasmic domains, Multibody refinement in RELION-3.0 was performed using two different masking strategies. First, to better elucidate the lumenal domain, Multibody was performed using two bodies: (i) the lumenal domain and a portion of the transmembrane region, and (ii) the cytoplasmic domain and the remaining portion of the membrane region. Second, to better elucidate the cytoplasmic domain, Multibody was performed again with two bodies: (i) the membrane and cytoplasmic region, and (ii) the lumenal domain alone. Together these resulted in EM density maps for the lumenal region composed primarily of EMC1 to an overall resolution of 3.2 Å, and the cytoplasmic region composed of EMC2, 8 and the cytoplasmic portions of EMC3 and 5 to an overall resolution of 3.6 Å (Fig. S2.3). Postprocessing was performed with a soft-mask of 6 pixels, extended by 10 pixels, and the Bfactor was estimated automatically in RELION-3.0 following standard procedures. Local resolution was calculated using RELION-3.0.

TM assignment

Two orthogonal strategies were used to assign the observed TMs. First, segmented maps were generated for each TM in Chimera (Pettersen et al., 2004). Based on the observed side chain density, an initial PDB was built using phenix.map_to_model, phenix.map_to_sequence (Terwilliger et al., 2018), or manual building based on an idealized helix in COOT (Casañal et al., 2020; Emsley et al., 2010). This PDB was then input into CryoID (Ho et al., 2020) to search against the predicted sequences of all twelve EMC TMs, with an additional five flanking residues included on either side. The top-scoring hits were manually inspected in the map for assignment.

In the second strategy, TMs with clear connectivity in either the cytosol or the lumen were manually identified. The 'overall map' was also segmented in Chimera to generate individual maps for these TMs and their connecting loops. A putative assignment was made based on the predicted topology of each EMC subunit, and the sequence and segmented maps were used in phenix.map_to_model (Terwilliger et al., 2018). The initial fitted TMs and loops were visually inspected and further refined by rounds of manual building in COOT (Casañal et al., 2020; Emsley et al., 2010) as described in Table S2.2, to generate a starting PDB.

These two orthogonal strategies converged on a single TM assignment, allowing for the confident assignment of eight of the well-ordered TMs (Fig. 2.1C-D and Fig. S2.4A-C). However, the final ninth TM could not be assigned to any of the predicted subunit TMs even after extensive phenix.map_to_model (Terwilliger et al., 2018) was carried out using the segmented unassigned TM map in combination to all unassigned TM sequence. Careful inspection of the EM density map revealed continuous density from residue 49 of EMC6 to an additional membrane-spanning region (Fig. S2.6D). After careful manual building in COOT followed by phenix.real_space_refine (Afonine et al., 2018) residues 12 to 48 of EMC6 could be modeled in this density. Proteinase K digestion analysis confirmed that indeed, EMC6 contains three TMs when translated in the presence of EMC5 (Fig. S2.6A-C).

The density for the three TMs of EMC4, 7, and 10 was poorly ordered, and putatively assigned based on connections to their structured lumenal domains (Fig. S2.7). Using the subpopulation of 42,506 particles (described above), stronger EM density was detected within this region, and could be used

to successfully model a backbone through the contiguous density that protruded from the termini of EMC4, 7, and 10. However, this region could not be unambiguously assigned and therefore these TMs were not included in the final deposited model.

Model building and refinement

Initial models for all the EMC soluble domains were generated using the transform restrained Rosetta (trRosetta) algorithm, a deep learning-based modeling method based on the Rosetta energy minimization pipeline with additional distance and interaction restraints generated from co-evolution (Yang et al., 2020).

Initially, the individual maps were inspected (cytosolic and lumenal), and each domain was individually fit each using the phenix.dock_in_map function of the PHENIX Software suite (Liebschner et al., 2019). The initial fitted models were visually inspected and further refined by iterative rounds of de novo model building in COOT as described in Table S2.2.

Initial attempts to build regions that differed from the trRosetta model, including connecting loops, and the N- and C-termini of most subunits, were made by first segmenting the map in Chimera (Pettersen et al., 2004) for use in phenix.map_to_model. Subsequent manual building was performed in COOT, using secondary structure prediction generated by trRosetta and PSIPRED (Jones, 1999; Yang et al., 2020).

The resolution of maps for the lumenal region was in the range of 2.5 - 4.0 Å (Fig. S2.3), allowing unambiguous model building. The presence of glycosylation sites at Asn residues (370, 818 and 913 of EMC1 and 182 of EMC10) and disulfide bond pairs (227-237 and 338-368 of EMC1) (Fig. S2.8), validated the register of the polypeptide chain. For the cytosolic domain, the register was confirmed by the presence of large aromatic side chains and using structure-based mutagenesis (Fig. 2.2 and Fig. S2.8).

The density for the flexible linkers between the top β-propeller and lumenal helix of EMC1 (amino acids 475-478), the a-helix 14 and a-helix 15 of EMC2 (amino acids 276-279), the TM1 and the coiled-coil of EMC3 (amino acids 40-43), and between the TM1 and TM2 of EMC5 (amino acids 37-39) was not sufficient to unambiguously assign the position of the side chains, and thus they are

included in the model as poly-Ala. Similarly, the density for the lumenal region of EMC4 that forms a β-strand and completes one blade of the bottom β-propeller of EMC1, was tentatively assigned as P170 to L183 of EMC4 (Fig. S2.8B). However, it was not sufficient to unambiguously assign the registry; therefore, in the model, the region is assigned as unknown amino acid (UNK) poly-Ala/Gly.

Initially, models for each region of the complex (TMs, cytosol and lumen) were refined against the respective EM density maps (overall, cytosolic and lumenal) using phenix.real_space_refinement and ISOLDE (Croll, 2018). Secondary structure restraints were generated in PHENIX (Liebschner et al., 2019) and were maintained during the refinement. For the lumenal region, a set of restraints for N-acetyl-D-glucosamine (NAG) were generated using phenix.elbow (Moriarty et al., 2009). Finally, all models were assembled and refined against the 'overall map'. The FSC curves between individual models and full maps (overall, cytosolic and lumenal) as well as per-residue B factors and cross-correlation were calculated after the refinement and plotted in Fig. S2.3 and S2.5. Statistics of the map reconstruction and model refinement are reported in Table S2.1. The final models were evaluated using MolProbity (V. B. Chen et al., 2010). All figures and movies were generated using Pymol (www.pymol.org) and Chimera. Subunit interface analyses were performed using the PDBsum (R. A. Laskowski et al., 2018) and PDBePISA server (R. A. Laskowski et al., 2018).

2.5 Data availability

Data and materials availability: atomic coordinates and cryo-EM maps have been deposited in the Protein Data Bank under accession code 6WW7 and in the Electron Microscopy Data Bank under accession code EMD-21929, 21930, and 21931. Plasmids for implementing the GFP-nanobody-based purification strategy are available via Addgene (IDs 149333, 149334, and 149336). Requests for materials should be addressed to R.M.V.

2.6 Acknowledgements

We thank Israel Sanchez-Fernandez, Chi-min Ho, Naima Sharaf, Chengcheng Fan, Gabe Lander, Andrey Malyutin, and Songye Chen for technical assistance, as well as Bil Clemons and the entire Voorhees lab for thoughtful discussion. Cryo-electron microscopy was performed in the Beckman

C h a p t e r   3

A SELECTIVITY FILTER IN THE ER MEMBRANE PROTEIN COMPLEX LIMITS PROTEIN MISINSERTION AT THE ER

The following chapter is adapted from Pleiner, Hazu and Pinton Tomaleri et al., 2023 and modified according to the Caltech Thesis format.

3.1 Abstract

Tail-anchored (TA) proteins play essential roles in mammalian cells, and their accurate localization is critical for proteostasis. Biophysical similarities lead to mistargeting of mitochondrial TA proteins to the ER, where they are delivered to the insertase, the ER membrane protein complex (EMC). Leveraging an improved structural model of the human EMC, we used mutagenesis and site-specific crosslinking to map the path of a TA protein from its cytosolic capture by methionine-rich loops to its membrane insertion through a hydrophilic vestibule. Positively charged residues at the entrance to the vestibule function as a selectivity filter that uses charge-repulsion to reject mitochondrial TA proteins. Similarly, this selectivity filter retains the positively charged soluble domains of multipass substrates in the cytosol, thereby ensuring they adopt the correct topology and enforcing the 'positive-inside' rule. Substrate discrimination by the EMC provides a biochemical explanation for one role of charge in TA protein sorting and protects compartment integrity by limiting protein misinsertion.

3.2 Introduction

A hallmark of eukaryotic cells is their organization into subcellular compartments that spatially separate otherwise incompatible biochemical reactions. The evolution of compartmentalization enabled the increasingly complex cellular processes required for emergence of multicellular life. To carry out distinct functions, each compartment must contain a unique and precisely defined set of proteins and metabolites.

Membrane proteins comprise ~20% of the human proteome (Krogh et al., 2001), and their localization is a primary determinant of organellar identity, underscoring the importance of their accurate sorting. Due to the presence of one or more hydrophobic transmembrane domains (TMDs), targeting and insertion of membrane proteins must be tightly regulated to prevent their aggregation in the aqueous cytosol. Canonical localization of many membrane proteins to mitochondria and the endoplasmic reticulum (ER) relies on cleavable targeting sequences that direct proteins to the correct organelle. Both the mitochondrial targeting sequence (MTS) and the ER-specific signal sequence are proteolytically removed upon arrival at their respective compartment, and thus have evolved principally to ensure accurate sorting without the need to serve a functional role in the mature protein.

However, given the functional and topological diversity of the membrane proteome, many nascent proteins cannot utilize these stereotypical biogenesis pathways. In these cases, membrane proteins instead rely on recognition of a TMD and its surrounding residues for accurate sorting (Guna & Hegde, 2018; Rapoport et al., 2017). These sequences must therefore play dual roles, experiencing evolutionary pressure to both function in the mature protein (i.e. insertion, folding, and assembly) and ensure accurate localization.

One important family of membrane proteins that rely on their TMD and its flanking residues for recognition, targeting, and insertion are tail-anchored (TA) proteins (Chio et al., 2017; Guna et al., 2022; Hegde & Keenan, 2011; Kutay et al., 1993). TA proteins are characterized by a single C-terminal TMD followed by a short soluble domain of up to 30-40 amino acids. Their globular N-termini are localized to the cytosol and are responsible for carrying out their diverse functions. Because of their topology, the TMD of a TA protein emerges from the exit tunnel of the ribosome only after translation termination, and they must be post-translationally targeted to the correct organelle. TA proteins are found on all cellular membranes and regulate essential processes such as

neurotransmitter release via exocytosis (SNARE proteins), cholesterol synthesis at the ER (squalene synthase [SQS]), and the onset of apoptosis at mitochondria (BCL-2, Bak). Given their biophysical diversity and the limited information for targeting, how TA proteins are accurately sorted between compartments has been a long-standing open question in the field.

TA protein localization is thought to be primarily dictated by two features: (i) properties of the TMD including its hydrophobicity and helical propensity, and (ii) properties of the C-terminal soluble domain that must be translocated across the bilayer during insertion (Costello et al., 2017; Kalbfleisch et al., 2007; Lin et al., 2021). TAs with highly hydrophobic TMDs are preferentially targeted to the ER membrane for insertion via the guided entry of tail-anchored protein (GET) pathway (Schuldiner et al., 2005, 2008). Its central targeting factor in human cells, GET3 (Favaloro et al., 2008; Stefanovic & Hegde, 2007b), binds TMDs using an ordered methionine-rich substrate binding groove and delivers its substrate TA proteins to the GET1/2 insertase for membrane integration (Mariappan et al., 2011). TAs with lower hydrophobicity TMDs, however, do not efficiently bind GET3 and thus cannot access the GET pathway (Guna & Hegde, 2018). The largest classes of such low hydrophobicity TAs are those targeted to the ER, where they are inserted by the ER membrane protein complex (EMC) (Christianson et al., 2012; Guna et al., 2018; Jonikas et al., 2009), and those targeted to the outer mitochondrial membrane, where they are inserted by MTCH1 and 2 (Guna et al., 2022). Because of their biophysical similarity, there is thought to be some constitutive levels of mistargeting between these compartments, necessitating dedicated quality control machinery at the ER and mitochondria to extract mislocalized TA proteins (Y. Chen et al., 2014; M. J. McKenna et al., 2020; Okreglak & Walter, 2014).

Because functional constraints limit the potential diversity of the TMD alone, a second sequence element, the short polar C-terminal domain, is known to contribute to TA protein sorting (Borgese et al., 2007; Isenmann et al., 1998; Kuroda et al., 1998). Though biophysically diverse, mitochondrial TA proteins are enriched for positive charges in their C-terminal tails, while the C-termini of ER targeted TA proteins are more likely to be net neutral or negatively charged. Manipulation of C-terminal charge is known to be sufficient to shift the localization of TA proteins between the ER and mitochondria (Costello et al., 2017; Horie et al., 2002; Rao et al., 2016). However, the biochemical basis for how changes in charge can alter TA protein sorting is fundamentally not clear. Considering recent advances in our mechanistic understanding of TA protein insertion into the ER (Bai et al.,

2020; Miller-Vedam et al., 2020; O'donnell et al., 2020; Pleiner et al., 2020), we sought to re-examine the molecular basis for sorting specificity between mitochondrial and ER TA proteins at this cellular compartment.

3.3 Results

Selectivity at the ER membrane

Previous studies of the canonical co-translational insertion pathway suggest that sorting fidelity is the combined result of contributions from cytosolic targeting steps and selectivity at the membrane (Akopian et al., 2013; Jomaa et al., 2022; Trueman et al., 2012). In the case of TA proteins, the source of this specificity at either step has remained elusive. While specificity during cytosolic targeting must undoubtedly contribute to TA protein localization, we found that even when loaded onto the identical chaperone *in vitro*, some mitochondrial TA proteins cannot be efficiently inserted into the ER membrane (Figure 3.1A). This selectivity appeared to correlate with C-terminal charge, because when positively charged amino acids were introduced within the C-terminus of the canonical ER TA protein squalene synthase (SQS), its insertion efficiency was dramatically diminished. Based on these observations, we concluded that there must be a source of substrate discrimination directly at the ER membrane, with selectivity occurring at the insertion step.

The EMC is the major insertase for ER-destined TA proteins with lower hydrophobicity TMDs, which are similar to those of mitochondrial TA proteins. Consistent with this biophysical similarity, we and others have demonstrated that the EMC is responsible for misinsertion of mitochondrial TA proteins into the ER (Figures 3.1B-D; (Guna et al., 2023; M. J. McKenna et al., 2022). Using an established split GFP system to specifically query TA integration into the ER (Figure 3.1B; (Inglis et al., 2020)), we found that multiple mitochondrial TA proteins were misinserted in an EMC, but not GET1/2, dependent manner (Figures 3.1C-D). We therefore reasoned that one source of discrimination against TAs with positively charged C-termini at the ER, either mitochondrial or the SQS mutants, must originate from properties of the EMC.
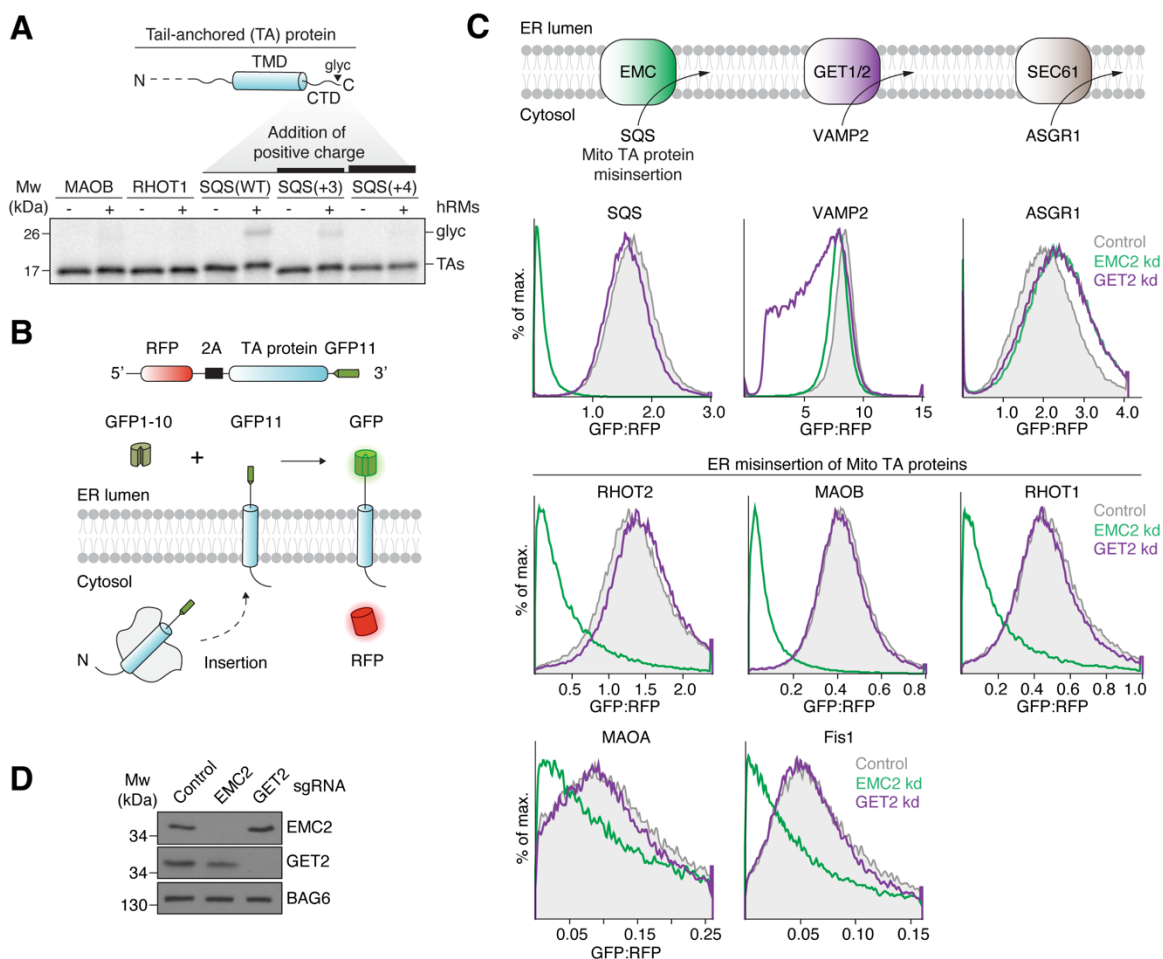
Figure 3.1 **EMC limits misinsertion of mitochondrial TA at the ER**

(**A**) (Top) Topology of a TA protein. (Bottom) ^{35}S-methionine-labeled TA protein with the indicated TMDs and C-terminal domains (CTDs), were expressed in the PURE system and purified as complexes with the cytosolic chaperone calmodulin. Glycosylation (glyc) of the CTD upon incubation with human ER microsomes (hRMs) indicates successful insertion. Samples were analyzed by SDS-PAGE followed by autoradiography. (**B**) Schematic of the split GFP reporter system used to selectively monitor TA protein insertion into the ER. TA proteins fused to GFP11 are expressed in K562 cells constitutively expressing GFP1-10 in the ER lumen, along with a translation normalization marker (RFP). Successful integration into the ER results in GFP complementation and fluorescence. (**C**) (Top) ER insertion pathways. (Bottom) ER insertion of the indicated ER (SQS, VAMP2, ASGR1) and mitochondrial (RHOT2, RHOT1, MAOA, MAOB, Fis1) TA proteins, using

the split GFP system as described in (B), was assessed in cells transduced with either a non-targeting (control), EMC2, or GET2 knockdown (kd) single guide RNA. GFP fluorescence relative to the normalization marker RFP was determined by flow cytometry and displayed as a histogram. (**D**) Cells from (C) were harvested and samples of total cell lysates were analyzed by SDS-PAGE and western blotting with antibodies against EMC2, GET2, and BAG6, a non-targeted control protein.

Substrate TMDs physically associate with the EMC's hydrophilic vestibule

With the goal of determining the biochemical basis of EMC's substrate specificity, we sought to map the path of a TMD from the cytosol into the bilayer through the EMC. Structures of the yeast and mammalian EMC identified two intramembrane surfaces that could potentially catalyze TMD insertion: a hydrophilic vestibule that positions several conserved positively charged residues within the cytosolic leaflet of the bilayer, and a hydrophobic crevice that contains a large lipid-filled wedge within the membrane (Bai et al., 2020; Miller-Vedam et al., 2020; O'donnell et al., 2020; Pleiner et al., 2020). Site-specific crosslinking experiments previously identified EMC3 as the major substrate interaction partner within the purified EMC (Pleiner et al., 2020), consistent with EMC3's homology with other members of the Oxa1 superfamily of insertases (Anghel et al., 2017). However, the path of a substrate TMD has never been directly determined, and potential contributions to insertion from both intramembrane surfaces of the EMC have been proposed.

To map direct physical association of substrates with the EMC, we exploited several independent zero-length crosslinking approaches to chart substrate interaction at single-residue resolution. First, we introduced the site-specific crosslinker BpA into the TMD of a canonical EMC TA substrate and identified UV-dependent crosslinks to both EMC3 and EMC4 by immunoprecipitation (Figure S3.1A). Unlike EMC3, which is present on both sides of the complex, the cytosolic and intramembrane surfaces of EMC4 partially enclose only the hydrophilic vestibule, suggesting substrates must at least transiently localize with this side of the EMC. Second, we exploited the fact that endogenous EMC3 does not contain any naturally occurring cysteine residues to perform disulfide crosslinking between a TA protein and the EMC. Because disulfide-bond formation can only occur between residues within 3-5 Å of each other, productive crosslinking necessarily indicates a direct physical association. Zero-length disulfide formation between single cysteines introduced at defined positions in EMC3, and a unique cysteine at two different positions within a substrate TMD,

identified a strong preference for substrate binding to the hydrophilic vestibule of detergent-solubilized EMC (Figures 3.2A-B and S3.1B). A similar preference was observed when comparing matched positions on either side of EMC3 at the base of the membrane. This preferential crosslinking was independent of cysteine position within the substrate TMD (Figure S3.1C) and was also observed upon incorporation of orientation-independent photo-crosslinkers in EMC3 (Figure S3.1D).
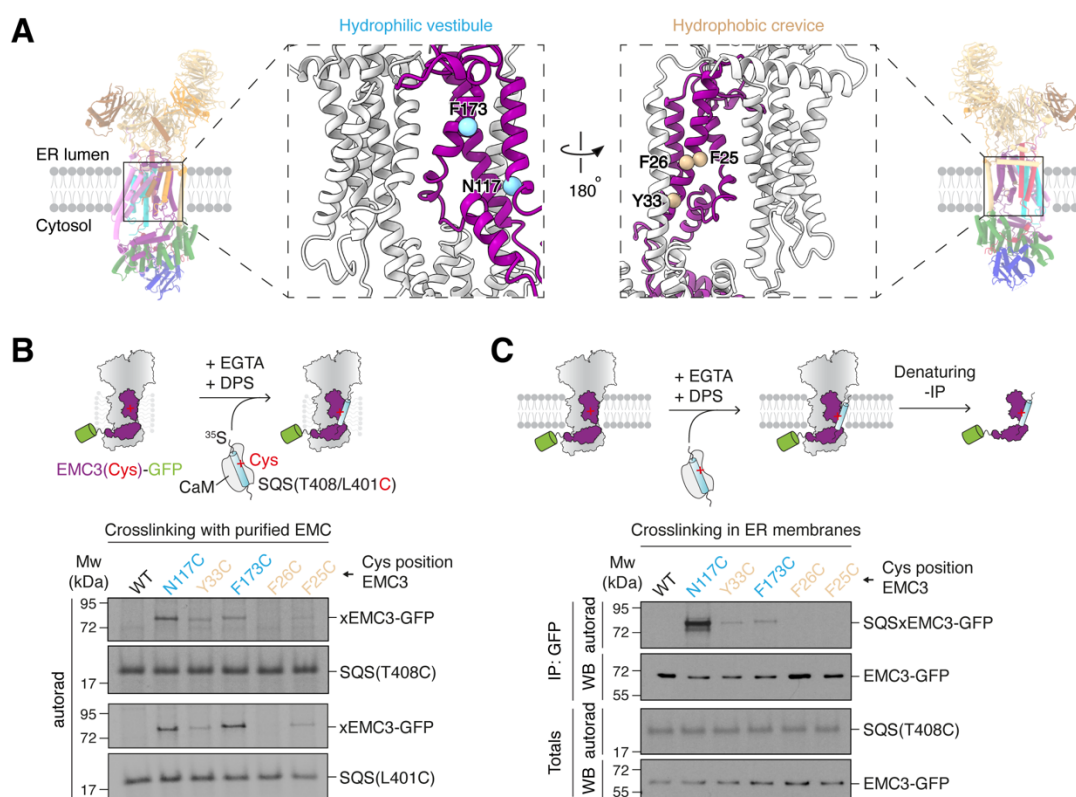
Figure 3.2 **The EMC uses a hydrophilic vestibule for TA protein insertion**

(**A**) Views of the two intramembrane surfaces of the EMC. Residues in EMC3 (purple) lining either the hydrophilic vestibule or hydrophobic crevice were mutated to cysteines for disulfide crosslinking and are highlighted in blue or tan, respectively. EMC4, 7, and 10 are omitted in the inset for clarity. (**B**) Purified wildtype (WT) or EMC3 cysteine (Cys) mutant EMC was incubated with CaM-SQS containing a cysteine in the TMD at either position T408 (CaM-SQS[T408C]) or L401 (CaM-SQS[L401C]). After substrate release from CaM with EGTA, cysteines in close proximity were

crosslinked with the zero-length disulfide crosslinker DPS. Quenched reactions were analyzed by SDS-PAGE and autoradiography. (**C**) Human ER-derived microsomes (hRMs) prepared from EMC3 WT or Cys mutant cell lines were mixed with CaM-SQS (T408C) for crosslinking as described in (B). Substrate crosslinks were enriched by denaturing purification of EMC3-GFP. Samples were analyzed by SDS-PAGE followed by autoradiography or western blotting.

Finally, and most definitively, we developed a strategy to capture the transient interaction between a substrate TMD and the EMC by disulfide crosslinking in native, insertion competent, ER membranes (Figures 3.2C and S3.1E). Using this approach, we again observed a marked preference for interaction of TAs with the hydrophilic vestibule of EMC3 compared to the hydrophobic crevice. In native membranes and with purified EMC, substrates preferentially crosslinked to a cytosol-facing position on EMC3 at the entrance to the lipid bilayer, suggesting a potential increase in dwell time at this location.

To further exclude that the opposite hydrophobic crevice is involved in TA protein insertion, we introduced multiple mutations to polar and hydrophobic residues in this region and found that they are all dispensable for TA protein biogenesis in human cells (Figures S3.1F-H). These data, in combination with sequence conservation, homology to Oxa1 superfamily insertases, and mutational analysis, definitively identify the hydrophilic vestibule as the insertase competent module of the EMC.

An improved model of the EMC defines intramembrane surfaces required for insertion

Having identified the hydrophilic vestibule as the major site of substrate binding to the EMC, we sought to better define its architecture and thereby identify potential sources of substrate specificity. The insertase core of the EMC (composed of EMC3 and 6) is partially enclosed by the dynamic subunits EMC4, 7, and 10. However, whether EMC7 and 10 contain TMDs, how these may be positioned, as well as the specific contributions of all three auxiliary subunits was incompletely defined.

To characterize the biophysical properties of the hydrophilic vestibule we obtained an improved cryo-electron microscopy (cryo-EM) reconstruction of the human EMC that allowed us to unambiguously assign and position the three TMDs of EMC4 and the single TMDs of EMC7 and 10

(Figures 3.3A, S3.2 and S3.3A-C; Table 3.1). In support of this model, we biochemically confirmed that human EMC7 and 10 both contain single C-terminal TMDs that span the lipid bilayer (Figure S3.3D). Examination of the roles of these subunits suggested that, consistent with previous studies, EMC4 and 7, but not 10 are required for TA protein biogenesis (Figure S3.3E; (Lakshminarayan et al., 2020; Louie et al., 2012; Volkmar et al., 2019). These auxiliary subunits do not play an architectural role in complex stability, as their depletion did not affect assembly of the core EMC subunits (EMC2,3,5,6,8) (Figure S3.3F). However, we additionally found that complete loss of EMC4 impaired the assembly of EMC7 and 10 into the EMC. Because EMC4's C-terminal β-strand completes the membrane-proximal β-propeller of EMC1, it is possible that loss of EMC4 disrupts the lumenal binding sites of EMC7 and 10. We concluded that the hydrophilic vestibule formed by the TMDs and cytosolic loops of EMC3 and 6 is partially enclosed by the five dynamic TMDs of EMC4, 7, and 10.
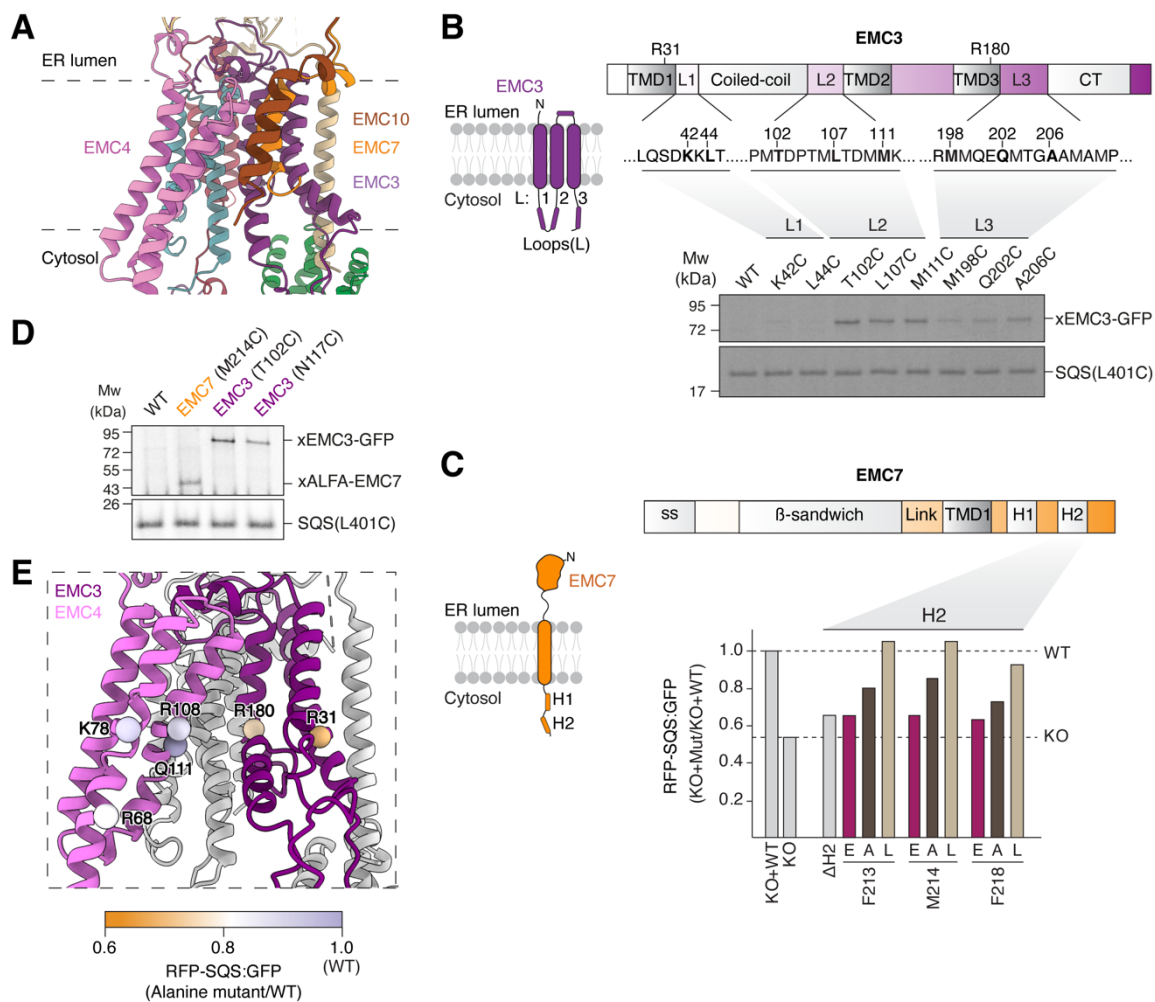
Figure 3.3 **Characterization of the residues required for insertion**

(**A**) Displayed is an improved model of the human EMC determined using cryoelectron microscopy (cryo-EM). View of the insertase core composed of EMC3/6, enclosed by the three TMDs of EMC4, and the single TMDs of EMC7 and 10. (**B**) (Top) Schematic of the topology and domain organization of EMC3, highlighting three flexible cytosolic loops (L1-3) located beneath the hydrophilic vestibule of the EMC. (Bottom) Purified wildtype (WT) or EMC3 Cys mutant EMC were incubated with purified CaM-SQS (L401C) complexes for disulfide crosslinking and analysis as in Figure 2B. (**C**) (Top) Schematic of the topology and domain organization of EMC7. ss = signal sequence; Link = linker; H1 = helix 1; H2 = helix 2. (Bottom) HEK293 EMC7 knockout (KO) cells were transduced with lentivirus to express WT EMC7, or the indicated mutants of EMC7 helix 2 (H2). The effects of

each mutant on biogenesis of SQS was determined using the ratiometric fluorescent reporter assay, normalized to WT and plotted as a bar chart. (**D**) Disulfide crosslinking, as described in Figure 3.2B, of SQS (L401C) with purified EMC complexes, containing cysteines either in H2 of EMC7 (M214S), loop 2 of EMC3 (T102C) or within the membrane (EMC3 N117C). (**E**) View of the hydrophilic vestibule with EMC7 and 10 omitted for clarity. Residues indicated with spheres are colored according to the effects of individual alanine mutations at these positions in EMC3 and 4 on expression of SQS in HEK293T cells. The effect of each mutant was determined by flow cytometry using the ratiometric fluorescent reporter assay as in (C), normalized to wildtype, and is displayed according to the indicated legend.

Capture of substrate TAs in the cytosol by the EMC

Based on this improved model of the EMC, we determined that the cytosolic loops of EMC3 and 7 are positioned immediately below the hydrophilic vestibule, making them prime candidates for cytosolic capture of substrates. We had previously shown that the flexible loops of EMC3 contain conserved methionine residues, commonly found in the TMD binding domains of cytosolic chaperones, that were important for EMC function (Pleiner et al., 2020). We therefore hypothesized that the loops of EMC3 and 7 could be involved in physically interacting with substrate TMDs in the cytosol. We set out to test key facets of this working model, with the goal of understanding whether the molecular details of substrate capture could contribute to discrimination between ER and mitochondrial TA proteins.

Consistent with earlier data, we found that methionine residues within the cytosolic loop of EMC3 were essential for TA protein biogenesis in cells (Figures 3.3B and S3.4A-B). Similarly, we found that the flexible C-terminus of EMC7 was required for EMC function (Figures 3.3C and S3.4C-F). Deletion of twelve residues to disrupt a predicted amphipathic α-helix, but not deletion of a matched upstream α-helix, strongly impaired SQS biogenesis, nearly phenocopying EMC7 knockout. We further demonstrated that the hydrophobicity of conserved residues within both this amphipathic helix of EMC7 and the methionine-rich loops of EMC3 is important, because their mutation to leucine, but not alanine or glutamate supported wild type levels of EMC function in cells (Figures 3.3C and S3.4A-F). However, for these loops to be directly involved in TA protein capture, they must be capable of physically interacting with substrate TMDs. Indeed, using zero-length disulfide

crosslinking, we found that the cytosolic loops of EMC3 and 7 specifically interact with substrates in a TMD-dependent manner (Figures 3.3B, D and S3.4G-H).

We concluded that the primary role of these flexible loops is to position hydrophobic residues within the cytosol, which physically capture substrate TMDs for subsequent insertion into the membrane. To test whether TA capture in the cytosol could contribute to substrate selectivity by the EMC, we used site specific crosslinking to compare the interaction of the TMD of wild type and mutant SQS, containing a positively charged C-terminus, with the loops of EMC3. We observed only a modest decrease in cytosolic capture of the positively charged SQS mutant (Figure S3.4I), suggesting that capture by EMC3 and 7 did not substantially contribute to substrate discrimination based on C-terminal charge. We therefore turned to consideration of the intramembrane surfaces of the hydrophilic vestibule.
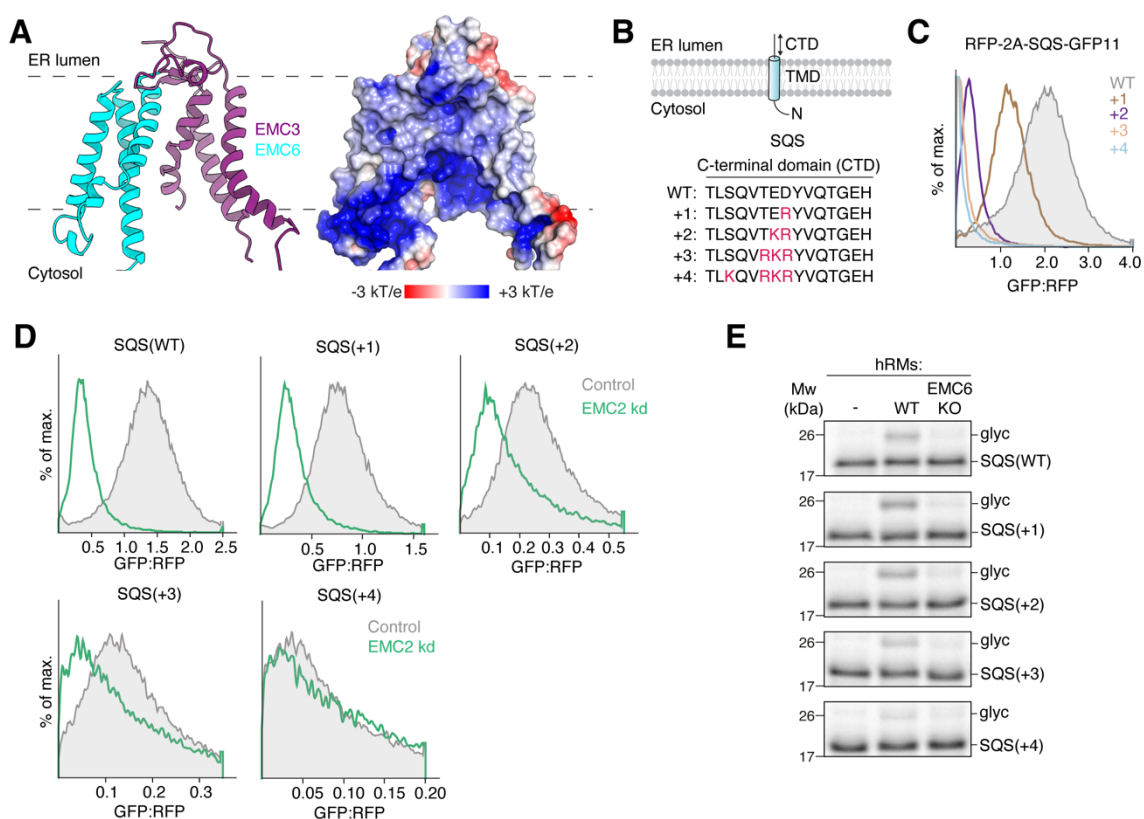
Figure 3.4 **EMC insertion is impaired by C-terminal positively charged residues.**

(**A**) (Left) Model of the TMDs of EMC3 and 6 that constitute the central insertase of the EMC. (Right) Surface representation of the electrostatic potential of the insertase core ranging from -3 to +3 kT/e. EMC4, 7, and 10 were omitted for clarity. (**B**) Schematic of the SQS C-terminal domain (CTD) charge series. The C-terminus of SQS was mutated to introduce positively charged residues at the indicated positions. (**C**) Integration of the indicated SQS mutants into the ER was determined using the split GFP reporter system described in Figure 1B. (**D**) Same assay as in (C), but with cells expressing either a non-targeting (control) or EMC2 knockdown (kd) single guide RNA. (**E**) The indicated ^{35}S-methionine labeled SQS charge mutants were expressed in rabbit reticulocyte lysate and incubated with human ER-derived microsomes (hRMs) prepared from HEK293 wildtype (WT) or EMC6 knockout (KO) cells. ER insertion is monitored by glycosylation (glyc) of an acceptor motif fused to the C-terminus of the TA protein substrates.

Substrates must passage through a positively charged hydrophilic vestibule for insertion

The improved atomic model of the EMC enabled detailed structure-function analysis of the biophysical requirements of the hydrophilic vestibule for TA protein insertion. The defining characteristic of the hydrophilic vestibule is a network of conserved polar and positively charged residues within the cytosolic leaflet of the lipid bilayer. Previous analysis suggests that charged and polar residues required for EMC function are positioned within the TMDs of the core insertase subunits EMC3 and 6 (Pleiner et al., 2020). Mutations to the positively charged residues in EMC3 strongly impaired insertion in cells, whereas mutations to EMC6 had only mild effects.

A more complete understanding of the localization of EMC4, 7 and 10 allowed us to systematically introduce mutations to all the polar residues that face the EMC3/6 insertase core (Figure 3.3E). However, we found that mutations to polar, charged, and methionine residues within EMC4's TMDs had little to no effect on TA protein biogenesis (Figures 3.3E and S3.5A-C). Only mutations of residues that likely affect TMD packing (N140) or lipid headgroup interaction (K67) showed significant phenotypes. If EMC4 does not directly contribute to function, it may instead be playing a role in regulating access to the hydrophilic vestibule, as deletion of its cytosolic EMC2-binding site strongly impaired SQS biogenesis (Figures S3.5D-E). Of all the polar intramembrane residues tested within the hydrophilic vestibule, the highly conserved R31 and R180 of EMC3 are the most crucial

for TA protein insertion, and their combined mutation displayed an additive effect on substrate biogenesis (Figures 3.3E and S3.5F-G).
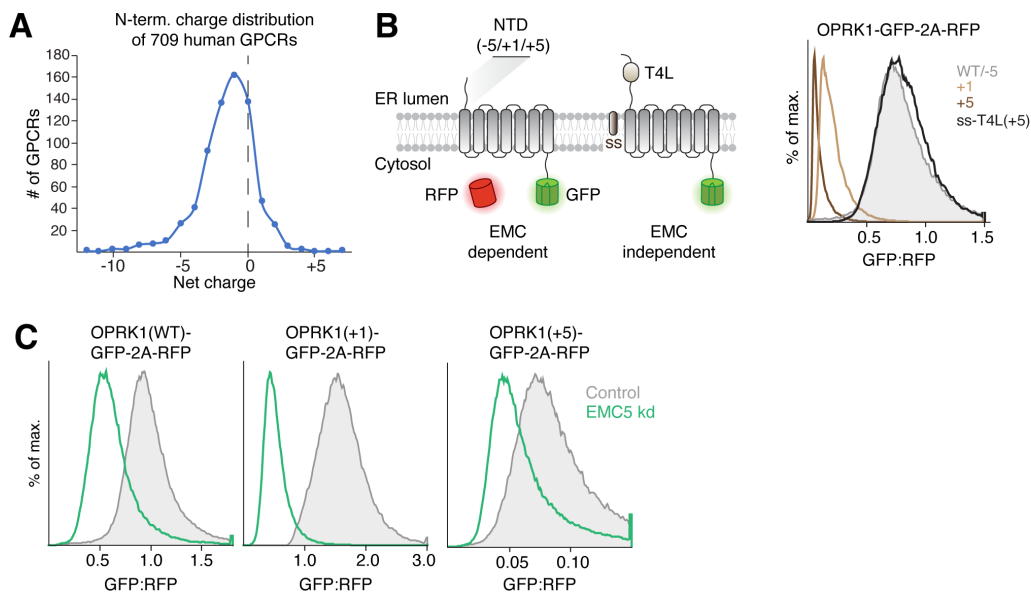
Figure 3.5 **EMC insertion is impaired by N-terminal positively charged residues**

(**A**) Distribution of charge within the soluble N-terminal domain of the 709 human GPCR sequences annotated in the Uniprot database. Only those GPCRs lacking a signal sequence (i.e. signal anchored) were included, because these represent substrates that could potentially rely on the EMC for insertion of their first TMD in an N^{exo} topology (N-terminus facing the ER lumen) (Chitwood et al., 2018). (**B**) WT (total N-terminal charge of -5) or the indicated N-terminal domain (NTD) charge mutants of the GPCR opioid receptor kappa 1 (OPRK1) GFP-fusions were expressed along with an RFP normalization marker in RPE1 cells. Cells were analyzed by flow cytometry and the GFP:RFP ratio is displayed as a histogram. Bypassing insertion by the EMC by fusion to a cleavable signal sequence (ss) enhances ER integration of the OPRK1(+5) charge mutant. (**C**) As in (B), but cells were treated with scrambled (control) or EMC5 knockdown (kd) siRNAs and analyzed by flow cytometry. Note that though the stability of the positively charged NTD variants is reduced, they remain EMC-dependent for their insertion.

Positively charged soluble domains impede insertion by the EMC

Both these mutational data and our crosslinking results together suggest that substrates must passage into the membrane directly along a positively charged surface of EMC3. Mislocalization of a mitochondrial TA protein into the ER requires both insertion of its TMD and translocation of its associated positively charged C-terminal domain. Thus, we reasoned that the positively charged hydrophilic vestibule is ideally positioned to discriminate mitochondrial and ER TA proteins through charge repulsion (Figure 3.4A).

To test the fundamental premise of this hypothesis, we first characterized the impact of charge on insertion by the EMC. In order to directly query the role of C-terminal charge, without confounding effects from comparing different substrates or TMDs, we generated a series of mutants of the canonical ER TA protein, SQS, containing increasing amounts of positive charge within its soluble C-terminal domain (Figure 3.4B). Using the split GFP reporter system, we found that while all SQS mutants inserted into the ER in an EMC-dependent manner, insertion efficiency was inversely correlated with positive charge (Figures 3.4C-D). Even addition of a single positive charge to the C-terminus of SQS resulted in a dramatic decrease in integration into the ER. Validating that this effect is specifically occurring at the insertion step and cannot be explained by other effects in cells (e.g. substrate stability), we observed a similar trend between charge and insertion into ER microsomes *in vitro* (Figure 3.4E).

In addition to its role in TA protein insertion, the EMC co-translationally inserts the first N_{exo} TMD (N-terminus facing the ER lumen) of many GPCRs that do not contain signal sequences (Chitwood et al., 2018). Like the C-termini of ER TA proteins, these GPCRs contain N-termini that are typically short, unstructured, and net negatively charged (Figure 3.5A; (Wallin & Von Heijne, 1995). Using the EMC-dependent GPCR OPRK1, we found that introduction of positive charge is again inversely correlated with insertion propensity by the EMC (Figures 3.5B-C). We therefore propose that inefficient translocation of positively charged extracellular domains is an inherent property of the EMC shared by both its co- and post-translational insertase function.

The EMC selectivity filter enforces TA protein sorting fidelity and the positive inside rule

The EMC's strong bias against translocation of positively charged domains provides a biochemical explanation for discrimination of mitochondrial TA proteins at the ER. To determine if this selectivity is due at least in part to charge repulsion between the hydrophilic vestibule of the EMC and the soluble

C-terminal domain of a substrate TA protein, we tested whether manipulation of the electrostatic potential of the EMC could alter substrate selectivity.

Due to the prominent location of R31 and R180 of EMC3 at the cytosolic entrance to the hydrophilic vestibule, these residues are ideally positioned to form a charge barrier that selectively prevents translocation across the lipid bilayer. If true, mutations that alter the electrostatic potential of these residues could alleviate repulsion between the EMC and positively charged soluble domains, allowing increased misinsertion of mitochondrial TA proteins. Mutation of both EMC3 R31 and R180 to alanine or glutamate did not affect EMC assembly, and as expected markedly impaired insertion of SQS in cells using our ratiometric fluorescent reporter system (Figures 3.6A-B and S5H). However, SQS variants containing increasingly positively charged C-termini showed increased insertion by the glutamate, but not the alanine mutant EMC. A similar trend was observed for insertion of SQS variants *in vitro* into wild type, alanine, or glutamate mutant ER microsomes, validating that charge specifically affects insertion propensity (Figure 3.6C). Similarly, these EMC3 mutations differentially affected the insertion of the co-translational substrate OPRK1 and its positively charged N-terminal domain mutants in cells (Figure 3.6D).
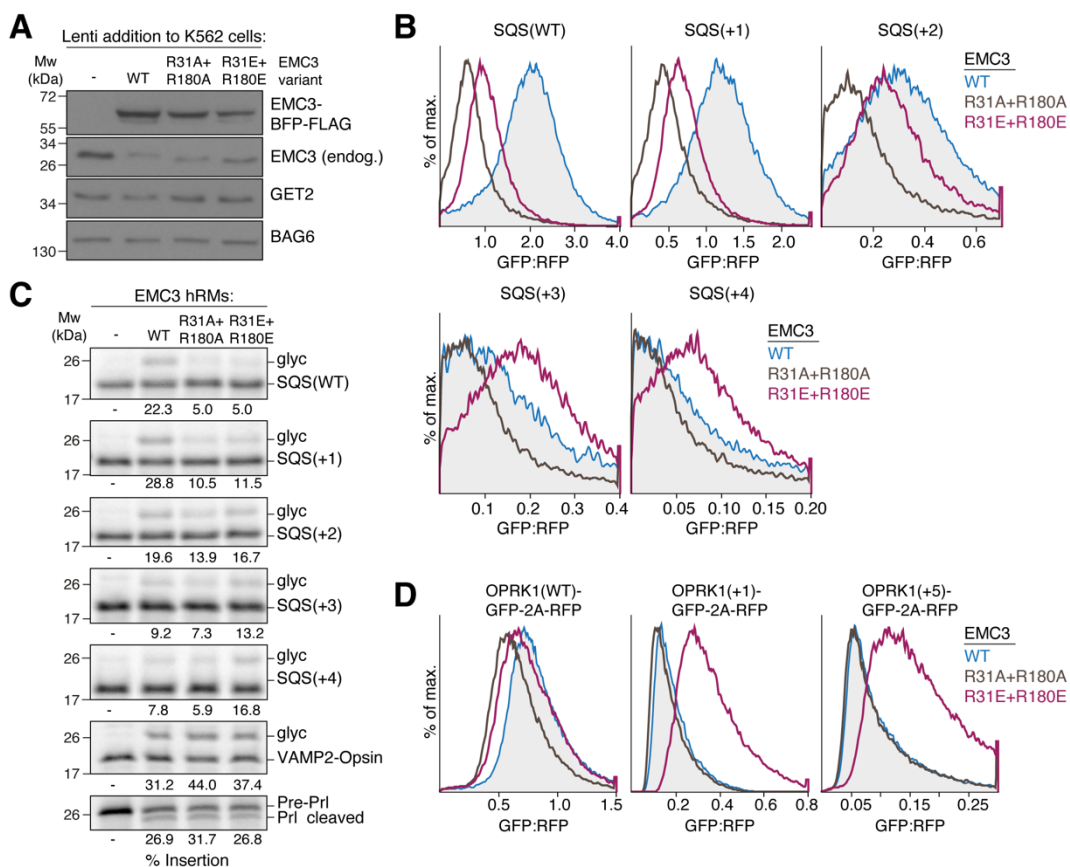
Figure 3.6 **Charge reversal in the hydrophilic vestibule alleviates charge repulsion**

(**A**) K562 ER GFP1-10 cells were transduced with lentivirus to express either WT, R31A+R180A or R31E+R180E EMC3. Cells were harvested, solubilized and samples of the total lysates were analyzed by SDS-PAGE and Western blotting with the indicated antibodies. (**B**) ER insertion of the indicated SQS charge mutants was measured in cells expressing either wild type (WT), R31A+R180A, or R31E+R180E EMC3 using the split GFP reporter system described in Figure 1B. (**C**) The indicated SQS mutants were prepared as in Figure 4E and incubated with hRMs from WT, R31A+R180A or R31E+R180E EMC3 expressing cell lines. Successful ER insertion is monitored with a glycosylation (glyc) acceptor motif fused to the C-terminus of each substrate. The % glycosylated is indicated below the gel. Expression of both EMC3 mutants does not impair the biogenesis of Get1/2-dependent VAMP2 or the secreted protein prolactin (Prl) that depends on the Sec61 complex (translocon). (**D**) WT (-5) or the indicated charge mutants of OPRK1 were fused to GFP and expressed with RFP as a translation normalization marker in RPE1 cells. Cells additionally

expressed either BFP-tagged EMC3 WT, R31A+R180A or R31E+R180E. Cells were analyzed by flow cytometry to derive the GFP:RFP ratio of BFP positive cells.

Because these SQS variants serve only as a proxy for the effects of charge on insertion, we tested whether manipulation of the EMC selectivity filter could also affect mislocalization of bona fide mitochondrial TA proteins into the ER. Indeed, we found that multiple mitochondrial TA proteins, most notably RHOT1, showed increased ER insertion upon expression of the glutamate, but not the alanine mutant of EMC3 in cells and *in vitro* (Figures 3.7A-B). Fis1, MAOA and MAOB similarly showed increased ER insertion. Even with increased mistargeting of TA proteins to the ER, induced by depletion of the outer mitochondrial membrane insertase MTCH2 (Guna et al., 2022) the selectivity filter at the EMC limited mitochondrial TA protein mislocalization to the ER (Figure S3.5I).

Based on this strong preference by the EMC against translocation of positively charged domains, we next tested whether charge repulsion could be used by the EMC to enforce the correct topology of multipass membrane proteins more broadly. Earlier work suggests the EMC assesses the topology-defining signal-anchor of nascent membrane proteins after ER targeting and hand-over from SRP (Chitwood et al., 2018). The N-terminal domains of type II (N_{cyt}) multipass proteins face the cytosol when inserted in the correct topology and are enriched for positive charge.
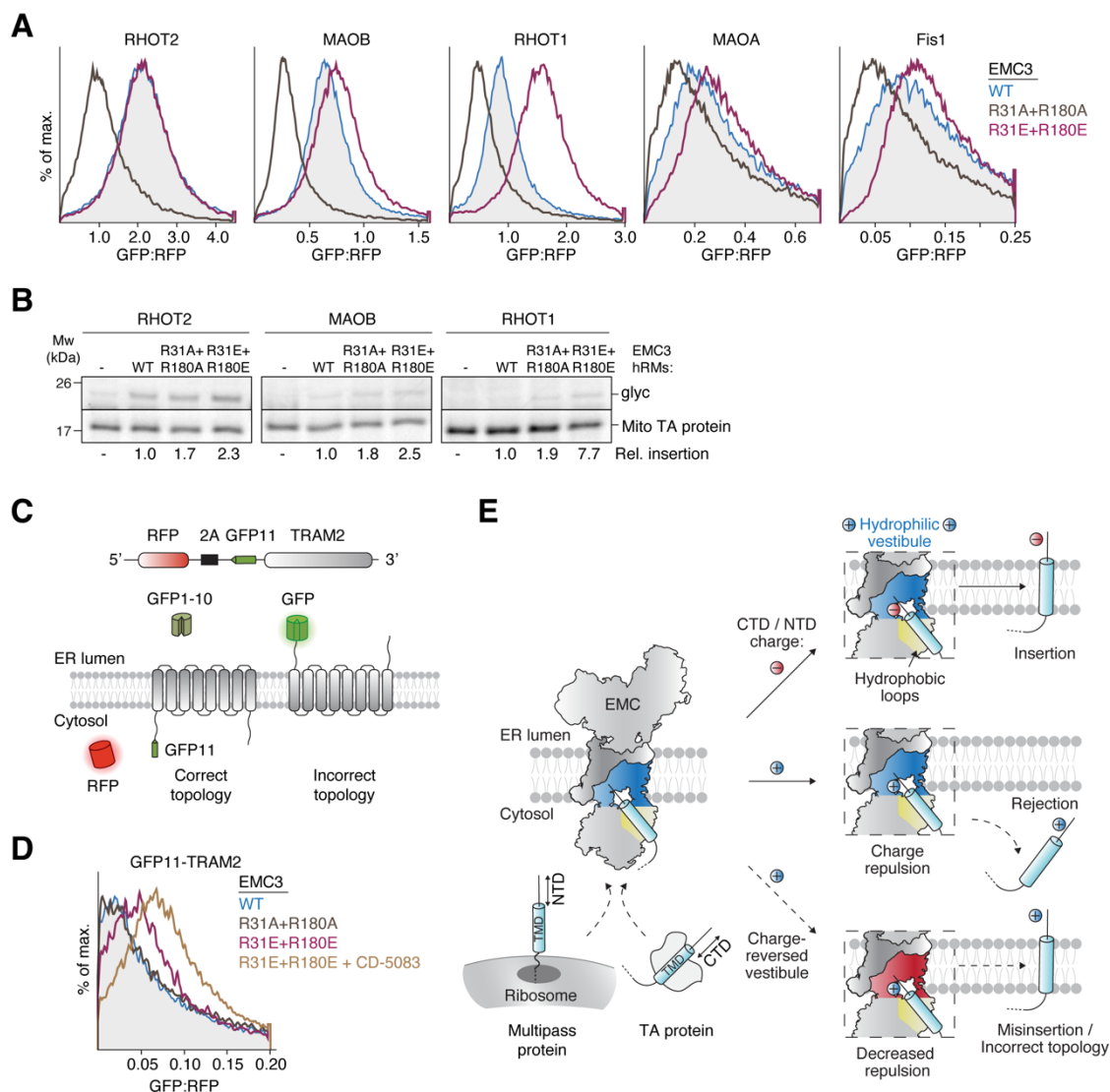
Figure 3.7 **EMC limits mitochondrial TA protein misinsertion at the ER.**

(**A**) As in Figure 6B, but with the indicated mitochondrial TA proteins. Note the strong increase in ER mis-localization of RHOT1 in EMC3 R31E+R180E expressing cells. (**B**) As in Figure 6C but expressing the TMD and C-terminus of the indicated mitochondrial TA proteins in non-nucleased rabbit reticulocyte lysate. (**C**) Schematic of the split GFP reporter system used to selectively monitor TRAM2 insertion in the incorrect topology into the ER. GFP11-tagged TRAM2 is expressed in K562 cells constitutively expressing GFP1-10 in the ER lumen, along with a translation normalization marker (RFP). Successful integration of TRAM2 in the correct topology will result in no

fluorescence. Insertion in the incorrect topology results in GFP complementation and fluorescence. (**D**) ER insertion of GFP11-TRAM2 was measured in cells expressing either wild type (WT), R31A+R180A, or R31E+R180E EMC3 with or without the p97 inhibitor CD-5083 using the split GFP reporter system described above. (**E**) Model for how the EMC distinguishes clients by polar domain charge. A TA protein TMD or the first TMD of a multipass membrane protein is initially captured by flexible hydrophobic loops in the cytosol, allowing their C- or N-terminal domain (CTD/NTD) to probe the net positively charged hydrophilic vestibule. In the absence of positive charge, the polar domain is translocated rapidly, enabling TMD insertion. Insertion of TA proteins with positively charged C-termini or multipass TMDs with positively charged N-termini is slowed by charge repulsion, which facilitates TMD dissociation (rejection). Charge repulsion can be alleviated by introducing negative charge into the hydrophilic vestibule, resulting in increased misinsertion of mitochondrial TA proteins into the ER membrane, as well as increased insertion of multipass proteins in the incorrect topology.

We postulated that the positively charged selectivity filter of the EMC would therefore reject such TMDs. To test this directly, we analyzed the extent of N_{exo} misinsertion of the GFP11-tagged N_{cyt} model protein TRAM2 in the presence of the EMC3 selectivity filter mutations. Indeed, the negatively charged glutamate, but not alanine mutant increased insertion of TRAM2 in the incorrect N_{exo} topology (Figures 3.7C and S5J). This misinserted population is subject to ER-associated degradation because it can be stabilized by the p97 inhibitor CD-5083. We therefore concluded that the EMC selectivity filter additionally limits misinsertion of multipass proteins in the incorrect topology and thus contributes to enforcing the 'positive-inside' rule (von Heijne, 1986).

3.4 Discussion

These results suggest that charge repulsion at the EMC provides a selectivity filter to control the subcellular localization of TA proteins (Figure 3.7D), enforcing their accurate sorting between the ER and mitochondrial outer membrane. The enrichment of positive charge in the C-termini of mitochondrial (and likely peroxisomal) TA proteins, serves as a flag for discrimination at the ER by the EMC. Unlike their TMDs, which must mediate function and targeting, the C-terminal domains of most TA proteins are functionally dispensable and may have evolved primarily to facilitate sorting specificity. The combined evolution of mitochondrial TA protein's positively charged C-termini and

the positively charged hydrophilic vestibule of the EMC thereby limits misinsertion of TA proteins at the ER membrane.

The molecular basis for TA protein discrimination was revealed by a systematic analysis of substrate insertion *in vitro* and in cells that defines the path through the hydrophilic vestibule of the EMC into the membrane. After delivery to the ER by a cytosolic chaperone, the first step in substrate insertion is handover and capture by the EMC. We found that substrate TMDs physically interact with the conserved, hydrophobic loops of EMC3 and EMC7 located immediately beneath the vestibule in the cytosol. Mutational analysis suggests that only the hydrophobicity of these loops, but not their specific amino acid sequence, is important for TA protein insertion. Indeed, comparison of EMC3 with its bacterial and archaeal homologs suggests that methionine-rich cytosolic loops are a conserved feature of Oxa1 superfamily insertases (Borowska et al., 2015), but the specific positioning of these hydrophobic residues is not strictly critical. We propose that these hydrophobic loops represent the first transient, flexible interaction site for substrate TMDs by the EMC.

We observed that substrates crosslink more efficiently to both these loops and the cytosol-exposed residues of the hydrophilic vestibule than to residues within the lipid bilayer. This difference was especially pronounced in native insertion-competent membranes, more likely to represent on-pathway intermediates that are not artefacts of detergent solubilization. These data would be consistent with a longer dwell time of substrates in this cytosolic intermediate followed by faster partitioning into the lipid bilayer. Similarly, a recent kinetic analysis of the bacterial insertase YidC suggests rapid substrate capture via its cytosolic loops and substantially slower translocation of the polar domain and membrane insertion (P. R. Laskowski et al., 2021). A plausible explanation for this observation might be that translocation of a polar domain across the hydrophobic lipid bilayer has a high energetic barrier and thus is a rate-limiting step to insertion.

This would be consistent with molecular dynamics simulations that suggest that TMD partitioning into the membrane is an energetically favorable process and membrane protein insertases are primarily required to decrease the energetic barrier for translocation of a soluble domain across the bilayer (Nicolaus et al., 2021; White & Wimley, 1999). Therefore, interaction of a substrate TMD with EMC's cytosolic hydrophobic loops could prevent aggregation, while its C-terminus probes the hydrophilic vestibule. For correctly targeted TA proteins, the EMC's hydrophilic vestibule serves as

a funnel that catalyzes translocation of their C-termini into the ER lumen by providing a hydrophobicity gradient between the aqueous cytosol and the core of the bilayer. Positioning of similar hydrophilic grooves or vestibules within a locally thinned membrane is a common feature of evolutionary distinct protein translocases (Kumazaki et al., 2014; McDowell et al., 2020; Voorhees et al., 2014; Wu et al., 2020), and represents a striking example of convergent evolution. In the case of the EMC, the dynamic TMDs of EMC4, 7, and 10 provide a protected environment, devoid of any potential off-pathway interaction partners, for the nascent protein to sample the bilayer.

However, for mistargeted mitochondrial or peroxisomal TA proteins, the positive net charge of the hydrophilic vestibule would impose a kinetic barrier to translocation of their positively charged C-terminal domains. In these TA proteins, positive charges are frequently found clustered near their TMD, suggesting that simple net charge alone, may not determine the extent of charge repulsion at the EMC. Repulsion likely delays translocation and thus increases the chance of TA protein dissociation from the hydrophobic loops. Using purified components, we previously showed that the cytosolic domain of the EMC does not contain an ordered high-affinity TMD binding site (Pleiner et al., 2021), as can be found in GET3 or SRP (Guna & Hegde, 2018). A composite transient TMD capture surface formed by flexible hydrophobic loops might allow for faster dissociation of TA protein clients and thus enable quicker accept/reject decisions. Rejected TA proteins in the cytosol could then be either recaptured for targeting to the correct organelle or triaged for degradation by quality control machinery. In this way, the EMC provides an additional layer of specificity to the accurate sorting of the ~600 TA proteins that must be expressed and localized in human cells.

The degree to which mitochondrial TA protein misinsertion into the ER is affected by the EMC selectivity filter is variable and likely influenced by multiple factors. For example, the inherent propensity for mistargeting to the ER differs amongst mitochondrial TA proteins (Guna et al., 2022). Additionally, detailed sequence features of a TA protein's C-terminal domain (i.e. total charge, charge density/positioning, secondary structure propensity) or TMD itself (i.e. helical propensity, length, hydrophobicity) might alter the effect of the EMC selectivity filter. The rules that determine the dependency of an individual TA protein on the selectivity filter represent an important question for future work.

The two positively charged residues in EMC3, which provide the charge barrier for entrance to the hydrophilic vestibule, are universally conserved in all Oxa1 superfamily insertases. As a result, its homologs, including GET1 and YidC, have also been suggested to inefficiently translocate positively charged soluble domains (Rao et al., 2016; Soman et al., 2014). Indeed, the effect of charge on insertion efficiency appears to be an inherent quality of the EMC and affects both its post- and co-translational substrates. Similar to EMC's TA protein substrates, GPCRs that lack an N-terminal signal sequence and are therefore potential EMC clients, typically contain neutral or negatively charged N-terminal extracellular domains (Figure 3.5A; (Wallin & Von Heijne, 1995). Using the same strategy for discrimination of mitochondrial TA proteins, the EMC also enforces the 'positive-inside' rule (von Heijne, 1986) for a subset of co-translational multipass substrates that meet its general client criteria (i.e. those without signal sequence containing a short and unstructured N-terminal domain). For N_{cyt} multipass clients, the EMC selectivity filter imposes correct topology by limiting translocation of their typically positively charged N-terminal cytosolic domains into the ER lumen using charge repulsion. The resulting longer dwell times at the EMC for N_{cyt} clients then likely triggers transfer to Sec61 for insertion in the correct topology.

Given that signal-sequence containing proteins are delivered to the ER membrane via the same route as multi-pass membrane proteins, it is likely that signal sequences also transiently sample EMC's hydrophilic vestibule. Their frequently positively charged N-terminal region (N-region) could mediate their rejection by the EMC selectivity filter and thus trigger handover to Sec61 for insertion in the correct N_{cyt} topology, required for signal sequence cleavage. In this model, the biophysical properties of the N-region would dictate the extent of charge repulsion at the EMC and therefore modulate signal sequence topogenesis. We thus propose that the EMC might contribute to the previously observed N_{exo} misinsertion of signal sequence-containing proteins that makes them substrates of corrective quality control pathways (M. J. McKenna et al., 2022). By extension, the selectivity filter in the EMC, would play a further role in enforcing the correct topology of secreted proteins, along with TAs and multipass membrane proteins.

In summary, we have characterized the molecular logic for how the EMC contributes to selective membrane protein localization in human cells. Its function is analogous to the active role Sec61 plays in substrate selection and rejection at the ER (Trueman et al., 2012). Whether MTCH1 and 2 also confer similar contributions to substrate selectivity at the mitochondrial outer membrane is an

important question for future research. However, specificity at the membrane is only one layer of the multi-faceted approach used to regulate protein sorting. Cells employ a sieved strategy in which the overall fidelity of protein localization is the combined result of selectivity at each biogenesis step including chaperone binding in the cytosol, insertion at the membrane, and extraction of misinserted substrates (Rao et al., 2016). How specificity is imparted during the targeting and extraction steps is an area that warrants further study. Particularly in metazoans, where membrane protein mislocalization can lead to disease (Juszkiewicz & Hegde, 2018), these steps are tightly coupled to quality control machinery that ensures immediate recognition and degradation of failed intermediates. By limiting misinsertion of TA proteins and preventing topological errors in multipass membrane proteins, the EMC serves as a guardian for protein biogenesis at the ER.

3.5 Supplementary Material
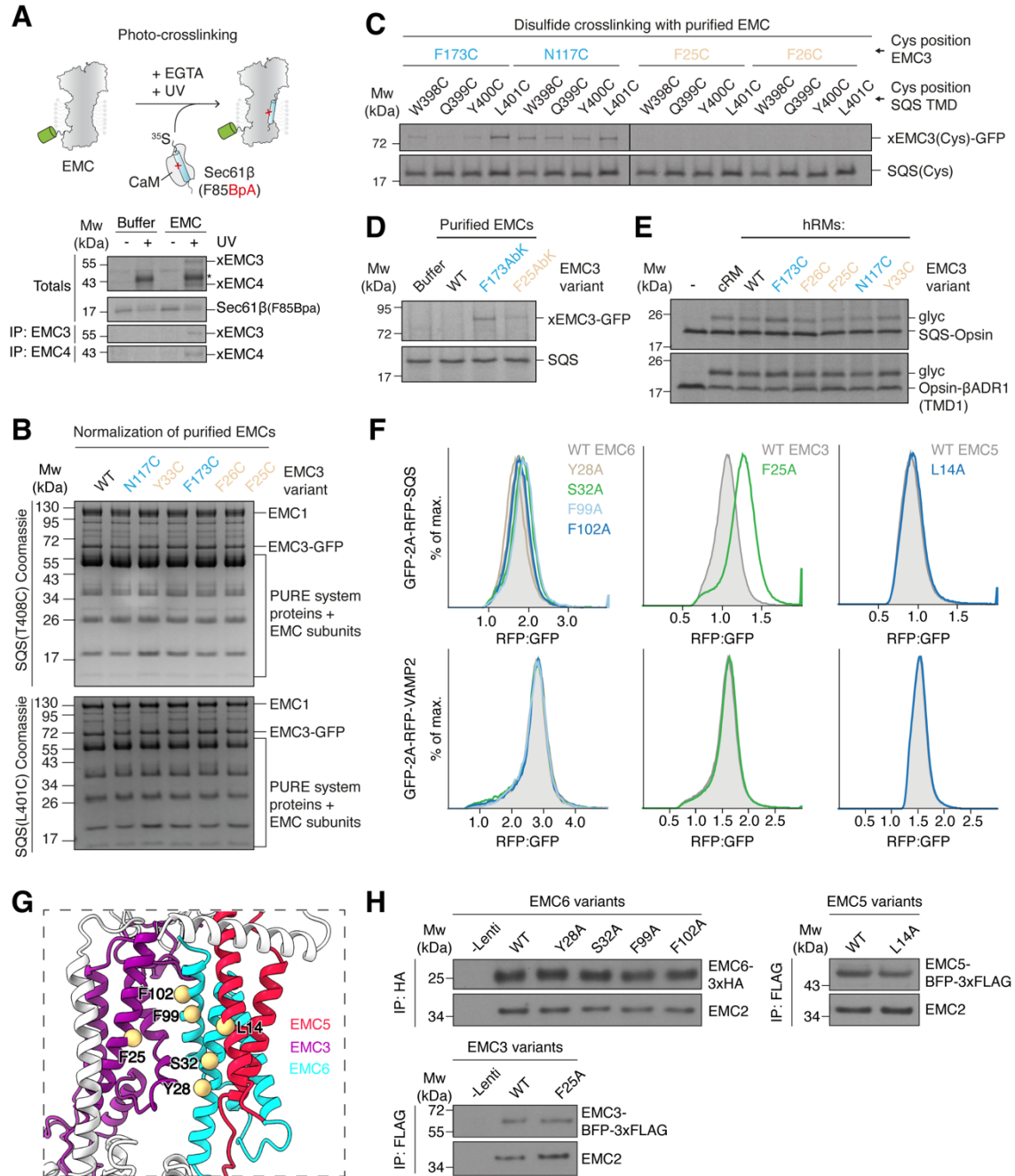
Figure S3.1 **Defining the hydrophilic vestibule as the insertase-competent side.**

(**A**) Schematic depiction of the site-specific photocrosslinking approach. The ^{35}S-methionine labeled TA protein substrate SEC61β, with a BpA photo–crosslinker incorporated into its TMD, was produced as a complex with calmodulin (CaM) in the PURE *in vitro* translation system. It was then incubated with EMC solubilized and purified in the detergent LMNG. Except for the -UV controls, all reactions were irradiated with UV light after substrate release from CaM with EGTA and then analyzed by SDS-PAGE and autoradiography. Crosslinks to EMC3 and EMC4 were identified by immunoprecipitation (IP) with anti-EMC3 and -EMC4 antibodies. The asterisk indicates the crosslinked TA protein dimer band. (**B**) Coomassie stained SDS-PAGE gel of the disulfide crosslinking experiments with purified EMC shown in Figure 2B before analysis via autoradiography. The gel shows that equal amounts of EMC were used in the different crosslinking reactions. (**C**) Disulfide crosslinking with purified EMC as in Figure 3.2B, but with cysteines positioned around a turn of the SQS TMD, showing that the observed crosslinking bias to residues on the hydrophilic vestibule (in blue) is independent of cysteine position. All crosslinking reactions were performed in parallel, and gels were exposed to the same film. (**D**) Purified EMC complexes containing the unnatural amino acid and photocrosslinker Abk incorporated into EMC3 at the indicated positions were mixed with SQS (WT)-CaM complexes prepared in the PURE system and irradiated with UV light after substrate release from CaM with EGTA. Samples were analyzed by SDS-PAGE and autoradiography. (**E**) Insertion activity of human ER-derived microsomes (hRMs) prepared from EMC3 WT or Cys mutant cell lines. Two well-characterized EMC substrates, SQS and TMD1 of the β-adrenergic receptor 1 (βADR1) (Chitwood et al., 2018; Guna et al., 2018), were translated in rabbit reticulocyte lysate in the presence of the indicated hRMs. Successful ER insertion results in the glycosylation (glyc) of the fused opsin tag. Canine pancreatic rough microsomes (cRMs) were used as a control. (**F**) HEK293 cells stably expressing RFP-SQS or -VAMP2 and cytosolic GFP as a normalization control were transduced with lentivirus to express the indicated mutants of EMC3, 5, and 6 in the hydrophobic crevice. The RFP:GFP ratio for each mutant was determined using flow cytometry and is plotted as a histogram. (**G**) Side-view of the membrane-spanning region of the EMC, focusing on the large cleft-like hydrophobic crevice. Residues on EMC3, 5, and 6 that were mutated in (F) line the cleft and are highlighted. (**H**) Incorporation of EMC subunit mutants into intact EMCs. A fraction of cells from (F) were harvested, solubilized, and subjected to anti-HA or anti-FLAG immunoprecipitation. Co-purification with the soluble subunit EMC2 indicates successful

incorporation of WT and mutant EMC3, 5, and 6 variants, suggesting that all the mutant subunits are assembled into the mature EMC.
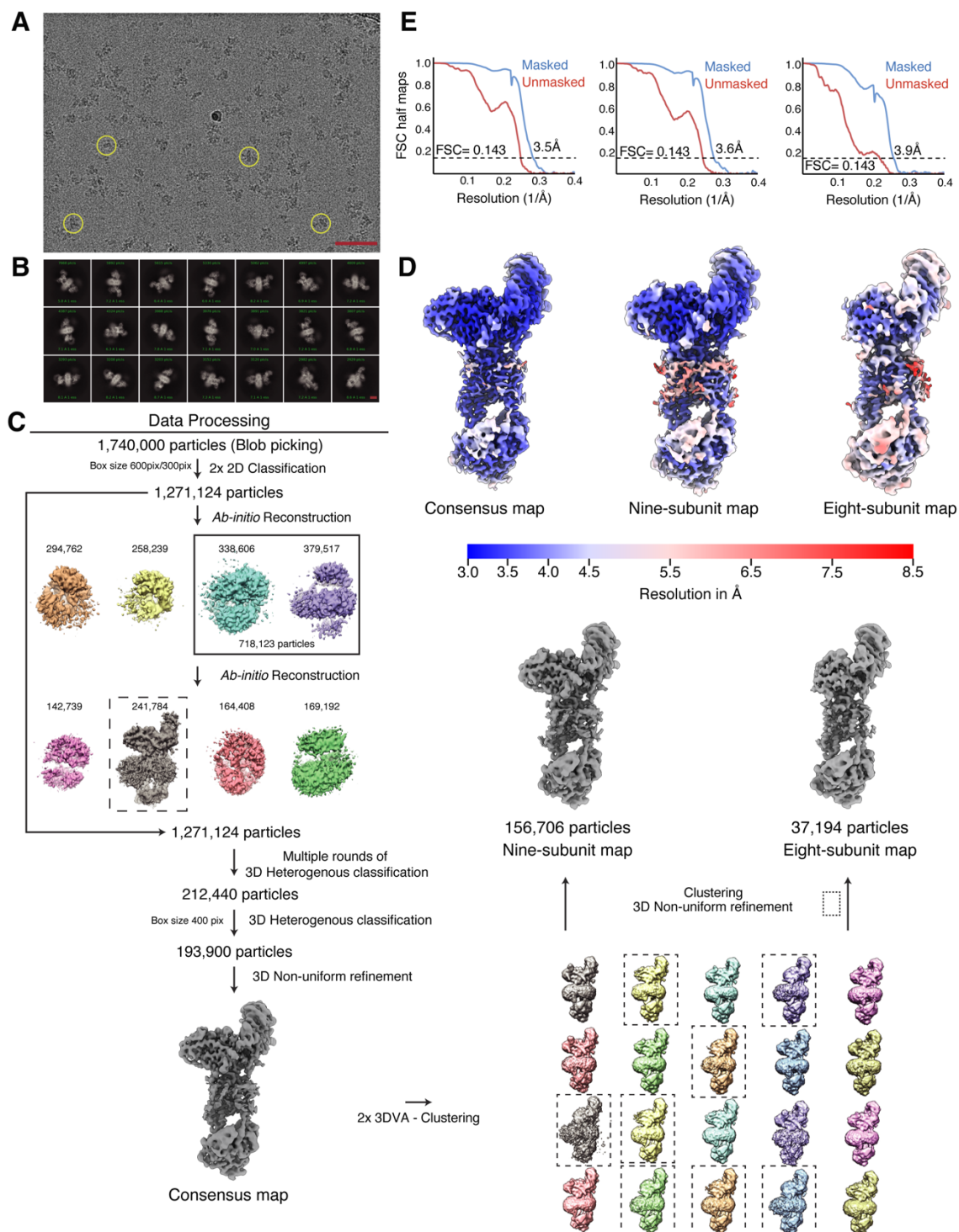
Figure S3.2 **Classification and refinement procedure of an improved model of the EMC.**

(**A**) A representative micrograph with several particles highlighted with yellow circles. Scale bar = 75 nm. (**B**) Representative 2D class averages generated during data processing. Scale bar = 5 nm. The number of particles for each class and its resolution are indicated. (**C**) Flowchart highlighting the data processing pipeline used to obtain an improved structure of the EMC. The 3D Variability Analysis (3DVA) enabled the exploration of the heterogeneity of the sample and allowed to parse out a subset of particles that lack the subunit EMC10, which provided unique insights into the placement of EMC10's TMD. Particles with all nine subunits, or those missing EMC10 (dashed boxes) were combined separately. Particles with poorly defined or low-resolution features were discarded (see Methods). (**D**) Final EM density maps colored by local resolution in Å. For clarity a dust filter was applied in ChimeraX. (**E**) Gold-standard Fourier Shell Correlation (FSC) curves for the consensus, 9- and 8-subunit complex maps generated by cryoSPARC V4.0.

Figure S3.3 **Architecture of the insertase-competent region of the EMC.**

(**A**) Updated model of the EMC, with views of the hydrophilic vestibule (left) and hydrophobic crevice side (right). (**B**) Low-pass filtered maps (5.5 Å) generated using volume tools in cryoSPARC V4.0. (Left) 9-subunit EMC complex map colored by the EMC subunits with the atomic model displayed as a superimposed cartoon. The EM density for the detergent micelle is displayed in gray. (Right) 8-subunit EMC complex (ΔEMC10) map. Due to the inherently flexible nature of EMC10's TMD we could not unambiguously model its TMD, however, comparing +/Δ EMC10 maps gave insights into localization of its TMD because the ΔEMC10 map lacks additional density (colored in brown) enclosing the hydrophilic vestibule of the EMC. (**C**) Updated schematic of the topology of

all nine EMC subunits. EMC8 and 9 are mutually exclusive paralogs. (**D**) EMC7 and EMC10 span the membrane. ^{35}S-methionine labeled EMC7 (top) or EMC10 (bottom) carrying an N-terminal signal sequence (ss) and 1xHA tag, as well as a C-terminal 3xFLAG tag were *in vitro* translated in rabbit reticulocyte lysate supplemented with canine pancreatic rough microsomes (cRMs). Nascent chains were released from the ribosome with puromycin, and non-incorporated as well as cytosolically accessible proteins were digested with proteinase K (PK) in the presence or absence of Triton-X-100 to solubilize the cRM membrane. The resulting protease protected fragments were subjected to denaturing anti-HA and anti-FLAG immunoprecipitations (IP). Note that only the N-terminal HA tags of EMC7 and EMC10 were protected (PF = protected fragment) from PK digestion, whereas the C-terminal 3xFLAG was PK-accessible, indicating a type I, single-spanning topology for both subunits. (**E**) EMC4 and EMC7, but not EMC10 are required for SQS biogenesis in human cells. WT or EMC4/7/10 KO HEK293 cells were transduced with lentivirus to express RFP-SQS or -VAMP2. The relative level of the RFP-fused TA protein to an internal GFP expression control was measured via flow cytometry and plotted as a histogram. (**F**) Purification of EMC complexes from HEK293 cells stably expressing GFP-EMC2 (WT), with or without additional knockout of EMC4, 7, or 10. Samples of total lysate and elution following an IP via GFP-E

A

GFP-2A-RFP-SQS

GFP-2A-RFP-VAMP2

% of max.

RFP:GFP

WT EMC3
M^{L2}A
M^{L2}L
M^{L2}E

B

Incorporation of EMC3 mutants

Mw (kDa)   BFP   WT   M^{L2}S   M^{L2}A   M^{L2}L   M^{L2}E

IP: FLAG

72
55   EMC3-BFP-3xFLAG

34   EMC2

M190
M194
M198
M200
Helix 1 (H1)

F213
M214
F218
Helix 2 (H2)

D

H1   Link

RFP-SQS:GFP (KO+Mut/KO+WT)

WT

KO

KO+WT   KO   ΔH1   M^{H1}E   M^{H1}A   M^{H1}L   KKR→EEE   Y136A   R152E

C

	M190 M194 M198 M200	F213 M214 F218	
Human	PLLIFVLL PKVVNTSDPD MRREMEQSMNM LNSNHELPDVSEFMTRLF SSKSSGKSSSGSSKTGKSGAGKRR---	242
Mouse	PLLIFVLL PKVVNTSDPD MRREMEQSMNM LNSNHELPDVSEFMTRLF SSKSSGKSSSGSSKTGKSGAGKRR---	241
Frog	PLLIFVLL PKVVNTSDPD MRREMEQSMNM LNTNPELPDVSEFMTRLF TSKSKSSKSSGSG-KAGKS-VGKRR---	237
Fish	PLLIIVLL PKVVNTNDPE MRKEMEQSMNM LNPNPELPDVSEFMTRLF SKGSSKSGGGN--KGSRSVATKRR---	237
Fly	PLLVMLVLL PKMI--NDPE ITKKE IES--IQFPKVSNEMPEISEMI TSFF SSAKPPEKEK---KPISSKQNKKRN--	228
Beetle	PLLLVMVLL PRMM--NDPEA RKEMEQ-LGNL-TKYDMPEMSEVI TSFF RGNQPKQQQR---CVKPAKKSSKQAAS	229
Horse	PLLIFVLL PKVVNTSDPD MRREMEQSMNM LNSNHELPDVSEFMTRLF SSKSSGKSSSGSSKTGKSGAGKRR---	241
Bat	PLLIFVLL PKVVNTSDPD MRREMEQSMNM LNSNHELPDVSEFMTRLF SSKSSGKSSSGSSKTGKSGAGKRR---	242
Bird	PLLIFVLL PKVVNTSDPD MRREMEQSMNM LNSNHELPDVSEFMTRLF SSKSSSGSSSSKAGKSSSGKRR---	246
Worm	PLVVMLIL PKMT-ANDPELKKEMEN-MQMPK--VDMPDVGEMM ANFF GGSAPAKKKA---VTGSGSGQRRK--	222

***::.::**:.:*   .:**: ::*:*.. :   ::**:.*.:: :*   .   :   .::

TMD

EMC7 C-terminus

E

GFP-2A-RFP-SQS

% of max.

RFP:GFP

WT + BFP
KO + BFP
KO + WT
KO + ΔH1
KO + ΔH2

WT + BFP
KO + BFP
KO + M^{H1}E
KO + M^{H1}A
KO + M^{H1}L

WT + BFP
KO + BFP
KO + F213E
KO + F213A
KO + F213L

WT + BFP
KO + BFP
KO + M214E
KO + M214A
KO + M214L

WT + BFP
KO + BFP
KO + F218E
KO + F218A
KO + F218L

GFP-2A-RFP-VAMP2

% of max.

RFP:GFP

F

Incorporation of EMC7 mutants

Mw (kDa)   KO+BFP   KO+WT   KO+M214E   KO+F213E   KO+F218E   KO+M^{H1}E   KO+Δ189-202   KO+Δ210-221

IP: ALFA

34   ALFA-EMC7

34   GFP-EMC2

Mw (kDa)   KO+BFP   KO   KO+K115/K117/R119E   KO+Y136A   KO+R152E

IP: ALFA

34   ALFA-EMC7

34   GFP-EMC2

G

WT   R191C   M214C   EMC7 variant

Mw (kDa)   WT   ΔTMD   WT   ΔTMD   WT   ΔTMD   SQS variant

autorad

43   xEMC7
34   WT dimer
ΔTMD dimer

SQS(L401C)
17   SQS(ΔTMD; F58C)

Coomassie

130
95
72
55   EMC3-GFP

43   PURE system proteins +
34   EMC subunits
26

H

Normalization of purified EMCs

Mw (kDa)   WT   EMC7 (M214C)   EMC3 (T102C)   EMC3 (N117C)

130
95
72
55   EMC3-GFP
43

34   ALFA-EMC7
26

I

WT   T102C   L107C   M111C   ← Cys position EMC3

Mw (kDa)   WT   +4   WT   +4   WT   +4   WT   +4   SQS C-ter charge

autorad

95
72   xEMC3-GFP

17   SQS(L401C)

Coomassie

130
95
72   EMC3-GFP
55
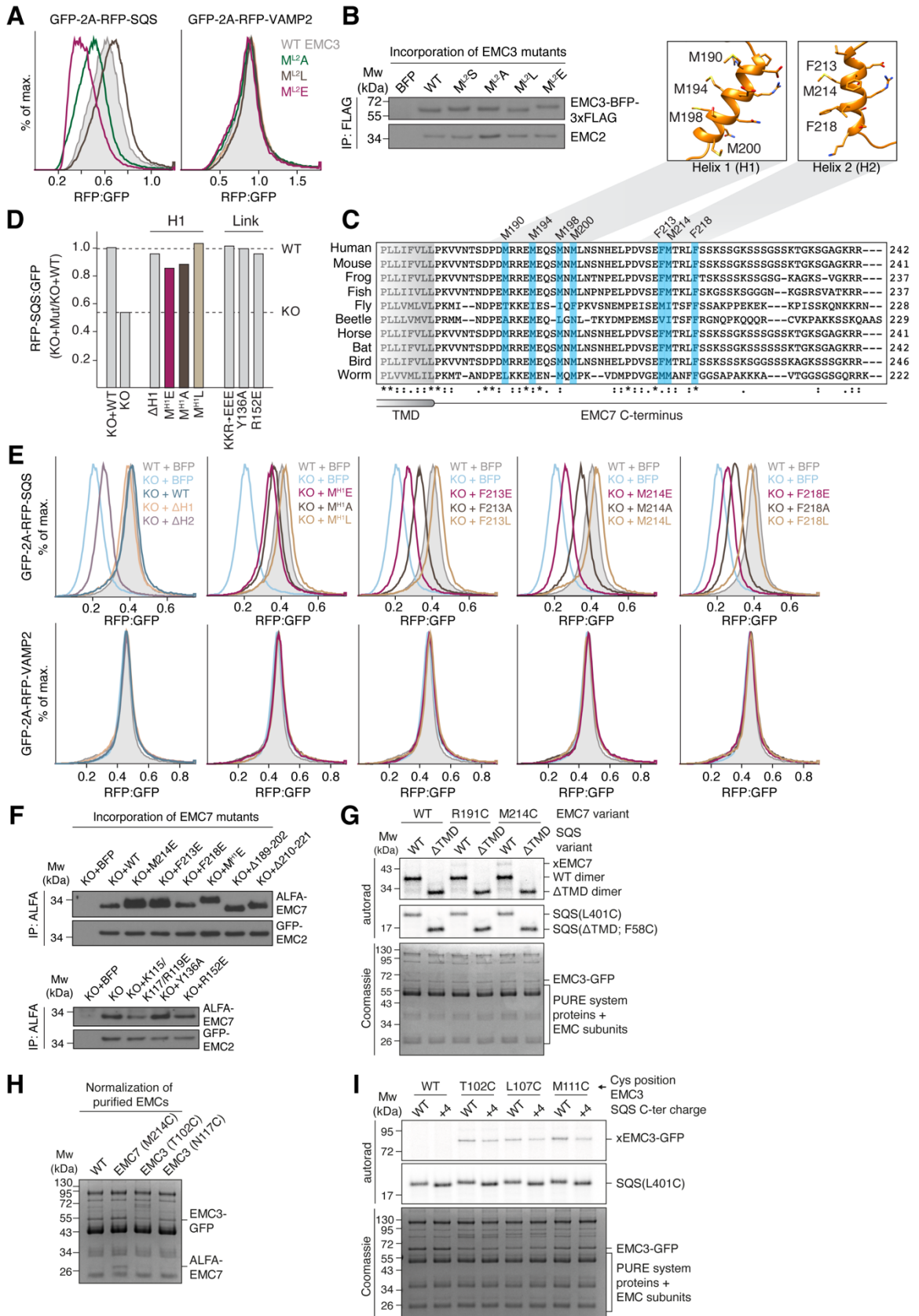43   PURE system proteins +
34   EMC subunits
26

Figure S3.4 **Substrate capture by EMC3 and 7.**

(**A**) HEK293 cells stably expressing RFP-SQS or -VAMP2 and cytosolic GFP as a normalization control were transduced with lentivirus to express the indicated EMC3 loop 2 mutants, along with BFP as a transduction marker. For each mutant, the RFP:GFP ratio of BFP-positive cells was derived via flow cytometry and is plotted as a histogram. M^{L2} refers to all four methionines in loop 2. (**B**) The indicated EMC3 loop 2 mutants were introduced into HEK293 cells via lentiviral transduction. Cells were harvested, solubilized, and subjected to anti-FLAG immunoprecipitation (IP). Eluates were analyzed by SDS-PAGE and western blotting with the indicated antibodies. (**C**) Alignment of EMC7 C-terminus sequences from various eukaryotes using Clustal Omega (Sievers et al., 2011). Two conserved sequence stretches are predicted by secondary structure algorithms to form α-helices, termed H1 and H2. Residues mutated in (E) are highlighted in blue. AlphaFold 2 models of H1 and H2 are shown. H1 is methionine-rich and H2 is predicted to form an amphipathic α-helix. (**D**) As in Figure 3C, but with the indicated mutants of H1 or the lumenal linker (link) between the EMC7's β-sandwich and TMD. M^{H1} refers to all four methionines in helix 1. KKR→EEE denotes the combined mutation of K115E, K117E, and R119E. (**E**) Wild type (WT) or EMC7 KO HEK293 cells were transduced with lentivirus to express either BFP alone or BFP plus EMC7 (WT) or the indicated mutants. 48 h after rescue construct transduction, cells were transduced with lentivirus expressing either RFP-SQS or -VAMP2, as well as a cytosolic GFP normalization control. The RFP:GFP ratio was determined by flow cytometry and is plotted as a histogram. Note that deletion of H2 strongly impaired SQS insertion in cells. Mutation of hydrophobic residues F213, M214, and F218 on H2 to either alanine or glutamate, but not leucine, similarly impaired SQS, but not VAMP2 biogenesis. (**F**) A BFP control, wild type EMC7 (WT), or the indicated mutants of EMC7 were introduced into EMC7 KO HEK293 cells via lentiviral transduction. Cells were harvested, solubilized, and subjected to anti-ALFA immunoprecipitation. Eluates were analyzed by SDS-PAGE and western blotting with antibodies against EMC2 and 7. (**G**) Purified EMC complexes containing either WT EMC7 or EMC7 with cysteines in H1 (R191C) or H2 (M214C) were incubated with purified CaM-SQS complexes with or without a TMD. The cysteine was placed either in the TMD (L401C) or the soluble linker (F58C), for the WT and ΔTMD SQS constructs, respectively. Disulfide crosslinking was carried out as in Figure 2B. (**H**) Coomassie stained SDS-PAGE gel of the disulfide crosslinking experiment shown in Figure 3D before analysis via autoradiography. The gel shows that equal amounts of EMC

were used in the different crosslinking reactions. (**I**) Purified wildtype (WT) or EMC3 Cys mutant EMC were incubated with purified CaM-SQS (L401C) complexes with wildtype (WT) or positively charged (+4) CTD. Disulfide crosslinking and analysis was carried out as above.
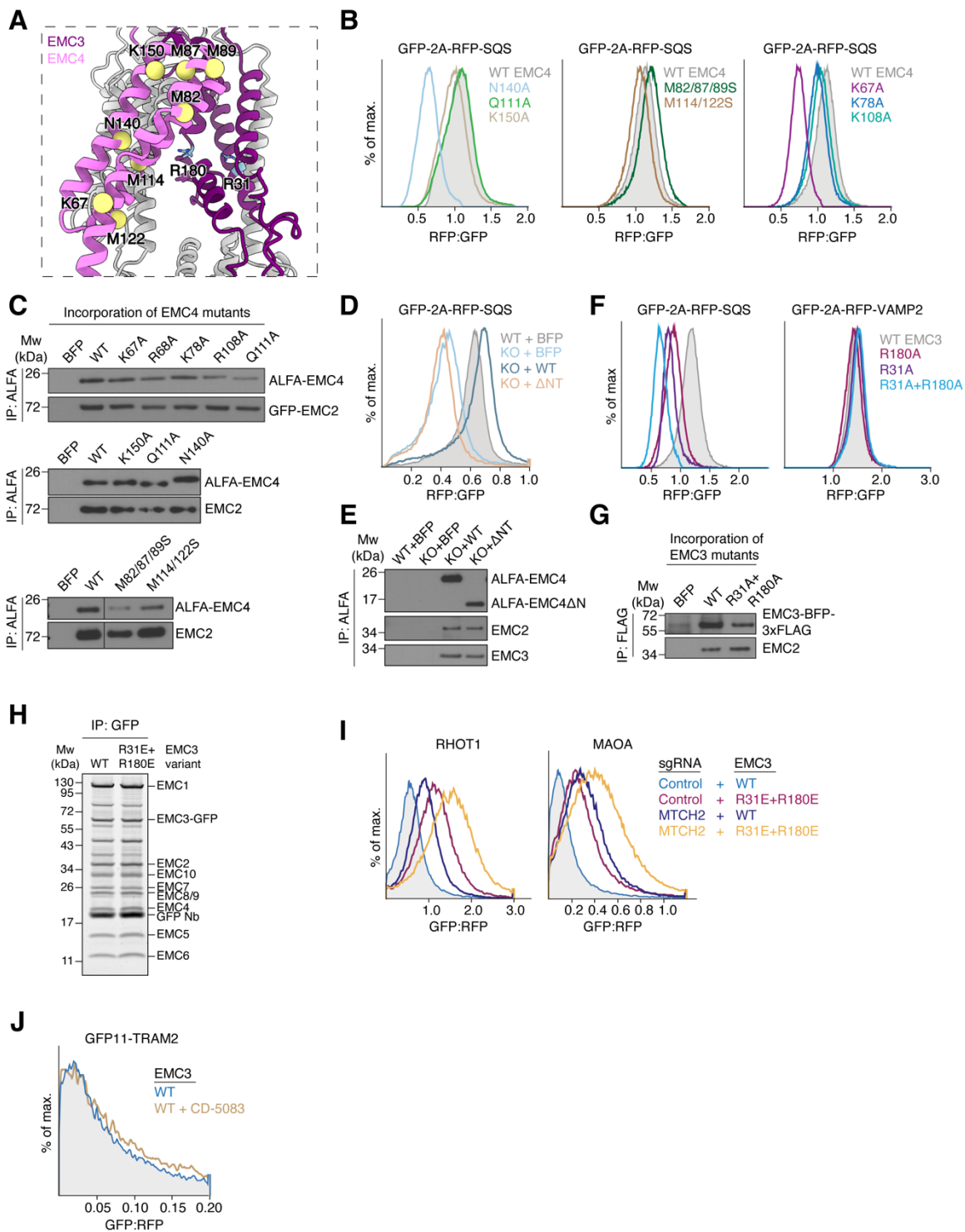
Figure S3.5 **Biophysical properties of the hydrophilic vestibule.**

(**A**) View of the insertase-competent side of the EMC. EMC7 and 10 were omitted for clarity. Residues of EMC4 mutated in (B) are highlighted. R31 and R180 of EMC3 are shown as blue sticks for reference. (**B**) HEK293 cells stably expressing RFP-SQS and cytosolic GFP as a normalization control were transduced with the indicated mutants of EMC4, along with BFP as a transduction marker. The RFP:GFP ratio of BFP-positive cells for each mutant was derived via flow cytometry and is plotted as a histogram. (**C**) The indicated EMC4 mutants from Figure 3D and (B) were introduced into HEK293 cells via lentiviral transduction. Cells were harvested, solubilized and subjected to anti-ALFA immunoprecipitation (IP). Eluates were analyzed by SDS-PAGE and Western blotting with antibodies against EMC2 and 4. (**D**) The N-terminus of EMC4 is required for TA protein biogenesis in cells. HEK293 WT or EMC4 KO cells were transduced with lentivirus to express either BFP alone or BFP plus EMC4(WT) or a ΔNT mutant (residues 57-end). 48 h after rescue construct transduction, cells were transduced with lentivirus expressing RFP-SQS, as well as a cytosolic GFP normalization control. The RFP:GFP ratio of BFP-positive cells was derived via flow cytometry and is plotted as a histogram. (**E**) A portion of the cells from (D) was harvested, solubilized, and subjected to purification of EMC4 variants via their N-terminal ALFA tag using the ALFA nanobody. The eluate was analyzed by SDS-PAGE and Western blotting with HRP-coupled ALFA nanobody or the indicated antibodies. (**F**) HEK293 cells stably expressing RFP-SQS or -VAMP2 and cytosolic GFP as a normalization control were transduced with lentivirus to express the indicated mutants of EMC3, as well as BFP. The RFP:GFP ratio of BFP-positive cells for each mutant was derived via flow cytometry and is plotted as a histogram. (**G**) A portion of the cells from (F) was harvested, solubilized and subjected to purification of EMC3 variants via their C-terminal 3xFLAG tag. Incorporation of the single mutants was described before (Pleiner et al., 2020). (**H**) Expi293 suspension cells stably expressing EMC3-GFP WT or R31E+R180E were solubilized and subjected to anti-GFP nanobody purification. The eluate was normalized by GFP fluorescence and analyzed by SDS-PAGE followed by Sypro Ruby staining. Note that both EMC3 WT and R31E+R180E mutant incorporate into EMCs with similar efficiency as they co-purify with all other EMC subunits. (**I**) Same assay as in Figure 3.7A but in cells transduced with either a non-targeting (control) or MTCH2 knockdown single guide RNA (sgRNA). (**J**) Same assay as in Figure 7D measuring the ER insertion of GFP11-TRAM2, but showing only WT EMC3 -/+ p97 inhibitor CD-5083.

Table S3.1 **Cryo-EM data collection, refinement, and validation statistics**

	Consensus (EMDB-40246)	Nine-subunit (EMDB-40245) (PDB 8S9S)	Eight-subunit (EMDB-40247)
Data collection and processing			
Microscope	FEI Titan Krios		
Voltage (kV)	300		
Camera	Gatan K3		
Magnification	105,000		
Defocus range (μm)	-1.0 to -3.0		
Pixel size (Å/pix)	0.416		
Electron exposure (e–/Å^2)	60		
Number of frames per movie	40		
Dose Rate (e-/pix/s)	16.0		
Automation software	SerialEM		
Number of micrographs	11,822		
Initial particle images (no.)	1,271,124		
Final particle images (no.)	193,900	156,706	37,194
Local resolution range	3.0 – 7.0	3.0 – 7.0	3.5 – 8.0
Map resolution range (Å, FSC=0.143)	3.5	3.6	3.9
Refinement			
Software (phenix.real_space_refine)	PHENIX 1.20.1-4487		
Initial model used (PDB code)		6WW7 + Alpha Fold	
Correlation coefficient (CC$_{mask}$)		0.83	
Map sharpening B factor (Å^2)	-112	-103	-76
Model composition			
Non-hydrogen atoms		18,012	
Protein residues		2262	
Ligands		7 NAG & 6 PCW	
B factors (Å^2)			
Protein		106	
Ligand		102	
R.M.S deviations			
Bond lengths (Å) (# > 4σ)		0.003	
Bond angles (°) (# > 4σ)		0.612	
Validation			
MolProbity score		1.92	
Clashscore		12.13	
Poor rotamers (%)		0.27	
Cß deviations (%)		0.00	
CaBLAM outliers (%)		1.60	
Ramachandran plot			
Favored (%)		95.29	
Allowed (%)		4.71	
Disallowed (%)		0.00	

3.6 Materials and Methods

<u>Plasmids and antibodies</u>

Constructs for *in vitro* translations in rabbit reticulocyte lysate were based on the pSP64 vector (Promega, USA). Constructs for *in vitro* translation in the *E. coli* PURExpress system were generated from the T7 PURExpress plasmid (New England Biolabs, USA). pSpCas9(BB)-2A-Puro (PX459) and lentiCRISPR v2 were gifts from Feng Zhang (Addgene plasmids #48139 and #52961). pLG1-puro non-targeting sgRNA 3, used for cloning CRISPRi sgRNAs, was a gift from Jacob Corn (Addgene plasmid #109003). The 2nd generation lenti-viral packaging plasmid psPAX2 (Addgene plasmid #12260) and envelope plasmid pMD2.G (Addgene #12259) were gifts from Didier Trono. The pHAGE2 lenti-viral transfer plasmid was a gift of Magnus A. Hoffmann and Pamela Bjorkman. For expression in K562 cells, a lenti-viral backbone containing a UCOE-EF-1α promoter and a 3′ WPRE element was used (Addgene #135448), which was a kind gift of Martin Kampmann and Jonathan Weissman. The expression plasmid for the SENP^{EuB} protease (Addgene ID #149333) was a gift of Dirk Görlich. Plasmids for amber suppression in mammalian cells were kind gifts of Simon Elsässer. Note that the mCherry variant of RFP was used throughout this study, but the simpler nomenclature of RFP is used in the text and figures. Similarly, EGFP is used throughout this study, but referred to as GFP.

The following antibodies were used in this study: rabbit polyclonal anti-EMC2 (25443-1-AP, Proteintech, USA); mouse polyclonal anti-EMC3 (67205-1-Ig, Proteintech, USA); rabbit polyclonal anti-EMC4 (27708-1-AP, Proteintech, USA); rabbit polyclonal anti-EMC5 (A305-833, Bethyl Laboratories, USA); rabbit polyclonal anti-EMC7 (27550-1-AP, Proteintech, USA); rabbit monoclonal anti-EMC10 (ab180148, Abcam, UK); rabbit polyclonal anti-GET2 (#359 002, Synaptic Systems, Germany); mouse monoclonal anti-HA-HRP (H6533, Millipore-Sigma, USA); mouse monoclonal anti-FLAG M2-HRP (A8592, Millipore-Sigma, USA). The rabbit polyclonal antibodies against BAG6 and GFP were gifts from Ramanujan Hegde (Chakrabarti & Hegde, 2009; Sharma et al., 2010). Secondary antibodies used for Western blotting were Goat anti-mouse- and anti-rabbit-HRP (#172-1011 and #170-6515, Bio-Rad, USA). The chemiluminescent substrates used were SuperSignal West Pico PLUS and SuperSignal West Femto Maximum Sensitivity (34580 and 34096,

Thermo Fisher Scientific, USA). The signal was detected on Blue Devil Autoradiography Film (#30-101, Genesee Scientific, USA).

The following sgRNAs were cloned into PX459 or lentiCRISPR v2 and used to generate knockout cell lines: EMC3 (AAGAAAGTGATGATAACGAT); EMC4 (TCATACACACCATCATAGTA); EMC6 (GCCGCCTCGCTGATGAACGG); EMC7 (TTCTCCGTCTACCAGCACTC); EMC10 (AGTGCCAACTTCCGGAAGCG). The following sgRNAs were cloned into pLG1 for CRISPRi knockdowns: non-targeting control (GGCTCGGTCCCGCGTCGTCG); EMC2 (GCCATCTTCCCAGAACCTAG); GET2 (ATGTTGGCCGCCGCTGCGA); MTCH2 (GACGGAGCCACCAAGCGACC).

The following siRNAs were used in this study: negative control no. 2 siRNA (#4390846) and EMC5 siRNA s41131 (both Silencer Select; Thermo Fisher Scientific, USA)

Expression and purification of biotinylated anti-GFP and anti-ALFA nanobody

Protease-cleavable biotinylated anti-GFP and anti-ALFA tag nanobodies (Götzke et al., 2019; Kirchhofer et al., 2010) that were used for EMC purifications throughout this study were expressed in *E. coli* and purified using Ni^{2+}-chelate affinity chromatography using protocols described in detail before (Pleiner et al., 2015, 2020; Stevens et al., 2023). The expression of His_{14}-Avi-SUMO^{Eu1}-anti GFP nanobody from plasmid pTP396 (Addgene #149336) was carried out with the following modification. Instead of biotinylating the nanobody *in vitro* with purified biotin ligase BirA, pTP396 was expressed in the *E. coli* strain AVB101 (Avidity, USA), which contains an IPTG-inducible plasmid for BirA co-expression. 50 µM biotin was added to the main culture 1 h before induction of nanobody and BirA expression.

The sequence of the ALFAST nanobody was derived from the original study describing its generation (Götzke et al., 2019)and cloned into pTP396. Expression was carried out in *E. coli* Rosetta-gami 2 cells (Millipore-Sigma, USA) in a 1 L scale for 6 h at 18˚C after induction of protein expression with 0.2 mM IPTG. The resulting His_{14}-Avi-SUMO^{Eu1}-anti ALFA nanobody fusion protein was purified from cell lysate using Ni^{2+}-chelate affinity chromatography for *in vitro* biotinylation with purified biotin ligase BirA as described before (Pleiner et al., 2020).

Immobilized biotinylated nanobodies were cleaved off of streptavidin magnetic beads using an engineered SUMO protease (SENP[EuB]) that recognizes the SUMO[Eu1] module (Rodriguez et al., 2019). His$_{14}$-Tev-tagged SENP[EuB] protease (Addgene ID #149333) was expressed in *E. coli* NEB express I^q as described before (Pleiner et al., 2020). For sequential immunoprecipitations, a commercial system with orthogonal cleavage sites based on the SUMOStar tag and SUMOStar protease (LifeSensors, USA) (Liu et al., 2008)was used.

Conjugation of ALFA nanobody to HRP for Western blotting

To use the ALFA nanobody in Western blotting, it was coupled to HRP-maleimide via a single engineered C-terminal cysteine residue as described for other nanobodies before (Pleiner et al., 2018).

Mammalian *in vitro* translation

In vitro translation reactions in rabbit reticulocyte lysate (RRL) were carried out with *in vitro* transcribed mRNA as described before (Sharma et al., 2010). PCR products generated from pSP64-derived plasmids or gene fragments (synthesized by Integrated DNA Technologies or Twist Biosciences, both USA) served as templates for run-off transcription and contained a 5' SP6 promoter followed by an open-reading frame and a 3´ stop codon. A 10 µl transcription reaction contained 7.6 µl T1 mix (Sharma et al., 2010), 0.2 µl SP6 polymerase (New England Biolabs, USA), 0.2 µl RNAsin (Promega, USA), 100 ng PCR product, and was carried out for 1.5 h at 37˚C. Transcriptions were added directly to RRL. Unless indicated otherwise, RRL was treated with S7 micrococcal nuclease (Roche, Germany) in the presence of CaCl$_2$ to remove endogenous hemoglobin mRNA. Nascent proteins are labeled during translation reactions of 15-30 min at 32˚C in RRL by incorporation of radioactive ^{35}S-methionine (Perkin Elmer, USA). Nascent TA proteins were released from the ribosome with 1 mM puromycin and then incubated with 5% (v/v) of either canine pancreatic rough microsomes (cRMs) (Walter & Blobel, 1983) or human ER-derived microsomes (hRMs), prepared from engineered cell lines as described below, for another 20 min at 32˚C. Samples were analyzed by SDS-PAGE and autoradiography to detect the translated ^{35}S-labeled TA protein.

Successful post-translational insertion into microsomes was monitored by glycosylation of a canonical NXS/T acceptor motif. This was appended either as part of a charged C-terminal opsin tag (MNGTEGPNFYVPFS**NKT**VD) or where no additional C-terminal domain charge was desired, an

NGT motif was placed 22 amino acids downstream of the TMD after a neutral glycine-serine linker and followed by an additional C-terminal GS dipeptide.

Protease protection assay

To assess the membrane spanning topology of EMC7 and EMC10, they were tagged with an N-terminal 1xHA and a C-terminal 3xFLAG tag and translated in RRL in the presence of cRMs as described above. Protease-accessible regions of both proteins were digested by incubation with 0.5 mg/ml Proteinase K for 1 h at 4˚C in the presence or absence of 0.05% (v/v) Triton-X-100 to solubilize cRM membranes. Proteinase K was inactivated by addition of 5 mM PMSF and quick transfer into boiling SDS buffer (100 mM Tris/HCl pH 8.4; 1% [w/v] SDS). Denatured digestion reactions were diluted tenfold with IP buffer (50 mM HEPES/KOH pH 7.5; 300 mM NaCl; 0.5 % [v/v] Triton-X-100) and incubated with anti-HA or anti-FLAG M2 resin (Millipore-Sigma, USA) for 1 h at 4˚C for immunoprecipitation of protected fragments.

Preparation of human ER-derived microsomes (hRMs)

To prepare hRMs from Expi293 suspension cell lines, cells were harvested and then washed twice in 50 ml 1x PBS. Cells were then resuspended in 4x pellet volume of sucrose buffer (10 mM HEPES/KOH pH 7.5.; 2 mM MgAc; 250 mM Sucrose, 1x Protease inhibitor cocktail [Roche, Germany]) and lysed with ˜50 strokes in a tight-fit dounce homogenizer. Complete cell lysis was verified by trypan blue staining. The lysate was then diluted twofold and spun for 30 min at 3214 g in a table-top centrifuge at 4˚C to remove nuclei and cell debris. This spin was repeated and the resulting supernatant was then centrifuged for 1 h at 75,000 g at 4˚C (TLA-100.3 rotor or Type60 Ti rotor; Beckman Coulter, USA). The supernatant was aspirated and the membrane pellet gently resuspended in microsome buffer (10 mM HEPES/KOH pH 7.5; 1 mM MgAc; 250 mM Sucrose, 0.5 mM DTT). Membranes prepared for disulfide crosslinking were resuspended in microsome buffer without DTT. The absorbance at 280 nm of the resuspended membranes was measured by boiling an aliquot in SDS buffer (100 mM Tris/HCl pH 8.4; 1% [w/v] SDS). The hRM preparation was then adjusted to an absorbance of 75 at 280 nm using microsome buffer. To remove endogenous mRNAs, the adjusted hRM preps were further treated with S7 micrococcal nuclease (Roche, Germany) at a concentration of 0.075 U/µl in the presence of 0.33 mM $CaCl_2$ for 6 min in a 25˚C water bath, then

quickly removed to ice and quenched by Ca^{2+}-chelation with 0.66 mM EGTA. Nucleased hRMs were snap-frozen in liquid nitrogen in single-use aliquots and stored until further use at -80˚C.

In vitro translation of TA proteins in the PURExpress system

Plasmids containing a 5´ T7 promoter, followed by an open-reading frame, stop codon and 3' T7 terminator were used as templates for the coupled *in vitro* transcription/translation PURExpress system (New England Biolabs, USA). The various SQS constructs used for cysteine crosslinking comprised an N-terminal 3xFLAG tag, the human Sec61β cytosolic linker (residues 2-59) with the natural cysteine at position 39 mutated to serine, as well as the five N-terminal flanking residues, TMD and complete C-terminus of human FDT1/SQS (residues 378-end). Cysteine residues were introduced at the indicated positions using site-directed mutagenesis. TA protein translations were supplemented with radioactive ^{35}S-methionine and 10 µM purified Calmodulin (CaM) (Shao et al., 2017)

For use in photocrosslinking reactions, TA protein substrates were generated that contained the unnatural amino acid and photocrosslinker 4-Benzoylphenylalanine (BpA) (Bachem, Switzerland), which was incorporated into the TMD by amber stop codon suppression in the PURExpress system lacking all release factors (ΔRF123; New England Biolabs, USA). The release factors RF2 and RF3, but not RF1 (which recognizes the UAG [amber] stop codon) were added back to the reaction. BpA was added at 100 µM and incorporated at UAG codons using purified BpA aminoacyl-tRNA synthetase and suppressor tRNA, prepared as described before (Shao et al., 2017) .

All PURE translation reactions were carried out for 2 h at 32˚C and then ribosome-associated nascent chains were released by addition of 1 mM puromycin (Thermo Fisher Scientific, USA) and further incubation for 10 min at 32˚C. To remove aggregated protein, the translation reactions were layered over a 20% (w/v) sucrose cushion prepared in physiological salt buffer (PSB) (50 mM HEPES/KOH pH 7.5; 130 mM KAc, 2 mM MgAc) that further contained 100 nM $CaCl_2$. After a 1 h spin at 55,000 rpm (TLS-55 rotor; Beckman-Coulter, USA) at 4˚C, soluble TA protein-CaM complexes were retrieved from the supernatant.

Photocrosslinking

Purified EMC complexes in detergent micelles for photocrosslinking were obtained via anti-GFP nanobody IP from stable human suspension cell lines that ectopically express GFP-EMC2. They were mixed with ^{35}S-Methionine labeled BpA-containing TA protein-CaM complexes generated in the PURExpress system as described above. TA proteins were released from CaM shortly before UV radiation by addition of 1 mM EGTA to chelate calcium. Except for the -UV control sample, all reactions were irradiated at a distance of ˜7-10 cm with a UVP B-100 series lamp (Analytik Jena, Germany) for 15 min on ice before quenching with SDS-PAGE sample buffer. Samples were adjusted to 1% (w/v) SDS and boiled. Denatured reactions were diluted tenfold with IP buffer (50 mM HEPES/KOH pH 7.5; 300 mM NaCl; 0.5 % [v/v] Triton-X-100) and incubated with Protein A sepharose beads (Thermo Fisher Scientific, USA) and EMC3 or EMC4 antibodies for immunoprecipitation. Samples were analyzed by SDS-PAGE and autoradiography.

Site specific incorporation of the photocrosslinking amino acid 3'-azibutyl-N-carbamoyl-lysine (AbK) into EMC3 in mammalian cells was performed by amber suppression using the *Methanosarcina mazei* pyrrolysyl-tRNA synthetase (PylRS)/tRNA$^{Pyl}_{CUA}$ (PylT) pair (Ai et al., 2011). Constructs for amber suppression in mammalian cells were created as follows using previously reported plasmids as template (Elsässer et al., 2016). The first plasmid encodes 4 copies of PylT (U25C), as well as WT PylRS, which was further modified by mutating Y306A and Y384F to accommodate the bulky AbK (O'donnell et al., 2020; Yanagisawa et al., 2008). The coding region of EMC3 was inserted with a C-terminal GFP-tag into a second plasmid which also encoded 4 additional copies of PylT (U25C). Selected amino acid positions in EMC3 were mutated to amber stop codons, for incorporation of AbK at these sites. To generate AbK-containing EMC, Expi293 cells (Thermo Fisher Scientific, USA) were transiently co-transfected with 4xPylT/PylRS (Y306A, Y384F) and 4xPylT/EMC3 (Amber[TAG])-GFP plasmids at a ratio of 4:1 using PEI "MAX" (Polysciences, USA). The cells were grown in the presence of 0.5 mM AbK (Iris Biotech, Germany) and harvested 72 h after transfection. EMC complexes with successfully suppressed Amber stop codons contained full length AbK-modified EMC3 and could thus be purified via the C-terminal GFP-tag as described below. The purified EMC complexes were mixed with ^{35}S-Methionine labeled SQS (WT)-CaM complexes generated in the PURExpress system and irradiated with UV as described above. Samples were analyzed by SDS-PAGE and autoradiography.

Disulfide crosslinking

EMC complexes containing wild type or cysteine mutant EMC3 or EMC7 variants were purified from stable human suspension cell lines as described below and mixed with wild type or cysteine mutant SQS-CaM complexes generated in the PURExpress system as described above. The zero-length disulfide crosslinker 4,4′-Dipyridyldisulfide (DPS) (Millipore-Sigma, USA) was added to a final concentration of 250 µM to initiate the crosslinking of cysteines in close proximity after SQS release from CaM with 1 mM EGTA. The reaction was incubated for 2 h on ice and analyzed by SDS-PAGE and autoradiography.

For disulfide crosslinking in membranes, hRMs were prepared from stable human suspension cell lines expressing wild type or cysteine mutant EMC3 variants as described above. hRMs were mixed with PURE translated SQS-CaM complexes in PSB and 500 µM DPS. After substrate release with 500 µM EGTA, reactions were incubated for 2 h on ice before quenching with 5 mM L-Cysteine (Millipore-Sigma, USA). The reactions were then adjusted to 1% (w/v) SDS and incubated at room temperature for 10 min to denature the EMC complex. The denatured reactions were diluted tenfold with IP buffer (50 mM HEPES/KOH pH 7.5; 300 mM NaCl; 0.5 % [v/v] Triton-X-100) and the EMC3-GFP subunit was specifically enriched via anti-GFP nanobody IP. After elution by boiling in sample buffer containing 0.5 M urea, the samples were analyzed by SDS-PAGE and autoradiography.

Cell culture and cell line generation

Adherent HEK293 cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal calf serum (FCS) and 2 mM L-Glutamine. For Flp-In T-Rex 293 cell lines containing integrated doxycycline-inducible reporters, tetracycline-free FCS was used and culture medium additionally supplemented with 15 µg/ml blasticidin S and 100 µg/ml hygromycin B. RPE1 cells were cultured in DMEM/F-12 (1:1) supplemented with 10% FCS and 2 mM L-Glutamine.

Flp-In 293 T-Rex cells were purchased from Thermo Fisher Scientific (USA). Stable Flp-In 293 T-Rex cell lines designated as GFP-2A-RFP-SQS/VAMP2 express the RFP-tagged transmembrane domain and flanking regions of human squalene synthase (SQS/FDFT1) or vesicle-associated membrane protein 2 (VAMP2). The generation of these cell lines was described previously (Guna et

al., 2018; Pleiner et al., 2020). In these cell lines, GFP is expressed as a soluble cytosolic protein from the same mRNA as RFP-SQS/VAMP2 using a viral 2A sequence that induces peptide-bond skipping by the ribosome (De Felipe et al., 2006). Their RFP and GFP fluorescence intensity can be measured by flow cytometry to derive a RFP:GFP ratio. Changes in this ratio after perturbation, e.g. expression of a mutant EMC subunit, reflect differences in the post-translational stability of the TA protein reporter.

The stable, doxycycline-inducible GFP-EMC2 Flp-In 293 T-Rex cell line and its adaptation to suspension growth in FreeStyle 293 Expression Medium (Thermo Fisher Scientific, USA) was described before (Pleiner et al., 2020). Clonal knockouts of EMC4, 7 and 10 in this background were obtained by transfecting the adherent parental cell line with PX459 encoding the respective sgRNA using TransIT-293 transfection reagent (Mirus, USA). 48 h post transfection, 1 µg/ml puromycin was added for three consecutive days. Medium was subsequently exchanged to allow for two days of recovery before single cell clones were seeded into 96-well plates by limiting dilution. Knockout efficiency of the selected clones was verified by Western blotting and the resulting adherent knockout cell lines were either used directly for flow cytometry experiments or adapted to suspension growth for EMC purifications.

Expi293 cells (Thermo Fisher Scientific, USA) were maintained at a concentration of 0.5-2.0 million cells per ml in Expi293 Expression Medium (Thermo Fisher Scientific, USA). An EMC3 knockdown suspension cell line was generated by transient transfection of Expi293 cells with an EMC3 sgRNA cloned into lentiCRISPR v2 using PEI "MAX" (Polysciences, USA). Transfected cells were treated with 10 µg/ml puromycin for four consecutive days. Then the medium was exchanged to allow for 10 days of recovery. The polyclonal cell population demonstrated a sufficient level of consistent downregulation of endogenous EMC3 and was thus used directly to re-introduce wild type EMC3 or various mutants tagged with a C-terminal TagBFP or GFP via lenti-viral transduction as described below. Transduced cell lines were sorted using fluorescence of the fused TagBFP or GFP to obtain a homogenous population of cells with near full replacement of endogenous EMC3 with a tagged mutant copy of interest. Wild type EMC7 or various cysteine mutants with an N-terminal ALFA tag were introduced via lenti-viral transduction into the EMC3-GFP cell line.

A K562 CRISPRi cell line, stably expressing dCas9-BFP-KRAB Tet-ON (Jost et al., 2017), was transduced with lentivirus as described below to constitutively express β-strands 1-10 of superfolder GFP (residues 2-214) (Cabantous et al., 2005) in the ER lumen via fusion to an N-terminal signal sequence and a C-terminal KDEL sequence as described previously (Guna et al., 2022).

<u>CRISPRi knockdowns</u>

K562 dCas9-BFP-KRAB Tet-ON, ER GFP1-10 cells were transduced via spinfection as described below with lentivirus containing a pLG1-puro backbone and a sgRNA targeting a gene of interest. Sequences of sgRNAs were derived from the hCRISPRi-v2 compact library (Horlbeck et al., 2016). 48 h after spinfection, 1 µg/ml puromycin was added for three consecutive days to select cells with a successfully integrated sgRNA expression cassette. After two days of recovery, cells were transduced with GFP11-tagged TA protein reporters expressed from a lentiviral backbone under control of a UCOE-EF1α promoter. Cells were analyzed 48 h after reporter spinfection by flow cytometry (8 days after sgRNA transduction).

<u>Lenti-viral transduction</u>

Lentivirus was generated by co-transfection of HEK293T cells with a desired transfer plasmid and two packaging plasmids (psPAX2 and pMD2.G) using the TransIT-293 transfection reagent (Mirus, USA). 48 h post transfection, culture supernatant was harvested, aliquoted, and flash frozen in liquid nitrogen.

For transduction of Expi293 or suspension-adapted Flp-In 293 T-Rex cells, 20 million cells were mixed with 2.5 ml freshly harvested lenti-viral supernatant (i.e. the complete supernatant from one 6-well of lenti-producing HEK293T cells 48 h after transfection) in 20 ml medium in a 125 ml vented Erlenmeyer flask (Celltreat, USA) (Stevens et al., 2023). Then the flask was transferred to a shaking incubator and transduced cells were grown for around 16 hours. Cells were then pelleted, resuspended in 50 ml of fresh medium, and grown for 2-3 days before sorting of successfully transduced cells on a SH800S cell sorter (Sony Biotechnology, USA).

K562 cells were transduced by spinfection. Briefly, 250,000 cells were mixed with 50-200 μl of lentiviral supernatant and RPMI medium in the presence of 8 μg/ml polybrene in a total volume of 1 ml in a 24-well plate. 24-well plates were spun at 1,000 g for 1.5 h at 30°C. Cells were then resuspended and transferred to a 6-well plate. Lenti-viral reporter constructs used in K562 cells for flow cytometry analysis all contained an upstream UCOE-EF1α promoter, followed by RFP, a P2A site and the full-length human coding regions for all mitochondrial TA proteins fused to GFP11 via a five residue Gly-Ser linker. SQS mutants were expressed in the same cassette but contained the cytosolic linker (residues 2-70) of human Sec61β at the N-terminus followed by the TMD, N-terminal flanking region and complete C-terminus of human FDFT1/SQS (residues 378-417 [end]). Charge mutations were introduced as shown in Figure 4B. EMC3 WT or its arginine mutants were expressed in K562 cells from a lentiviral transfer plasmid with an upstream EF1α promoter and fused to a C-terminal TagBFP-3xFLAG tag.

For lenti-viral transduction of adherent HEK293 or RPE1 cells, 50-200 μl lentiviral supernatant and 8 μg/ml polybrene (Millipore-Sigma, USA) were usually added directly to ˜70% confluent cells in 2.5 ml culture medium in a 6-well. Lenti-viral reporter constructs of SQS and VAMP2 for use in HEK293 cells (Figures 3C,E, S3E, S4A,D-E, S5A-B,D) contained an upstream CMV promoter, followed by GFP, a 2A site and RFP, which was directly fused to the TMD and flanking regions of human FDFT1/SQS or VAMP2 as described before (Guna et al., 2018; Pleiner et al., 2020). OPRK1 reporter constructs used in RPE1 cells, expressed full length human OPRK1 (WT/-5), OPRK1 (E45K, D46R, E50K) (+1 variant) or OPRK1 (E35K,D37R,E45K,D46R,E50K) (+5 variant) as N-terminal fusions to GFP, followed by a 2A site and RFP from a CMV promoter.

<u>Flow cytometry analysis of reporter cell lines</u>

All adherent cells were trypsinized, washed, and resuspended in 1xPBS for flow cytometry analysis. K562 cells were analyzed directly. Analysis was either on an Attune NxT Flow Cytometer (Thermo Fisher Scientific, USA) or a MACSQuant VYB (Miltenyi Biotec, Germany). Flow cytometry data was analyzed using FlowJo v10.8 Software (BD Life Sciences, USA). Unstained cells transiently transfected with either GFP, RFP (or BFP if needed) were analyzed separately along every run as single-color controls for multicolor compensation using the FlowJo software package.

For experiments in K562 cells, lenti-viral fluorescent reporters were introduced via spinfection as described above usually 48 h before analysis. To probe the effect on EMC2 or GET2 knockdown on reporter insertion, cells were additionally transduced with sgRNA expressing lenti-viral vectors as described under 'CRISPRi knockdowns'. To analyze the effect of EMC3 mutations on TA protein reporters, K562 cells were first spinfected with lentivirus expressing EMC3 (WT/mut)-BFP. After 48 h, mitochondrial TA protein or SQS charge mutant reporter lentivirus was spinfected. Cells were analyzed by flow cytometry after another 48 h. For experiments with p97 inhibitor (CD-5083 [Selleckchem, USA]) the cells were treated with 1.25 µM inhibitor for the last 6 h before analysis. Adherent HEK293 or RPE1 cells were analyzed 48 h after transduction as described above.

Purification of engineered EMCs from stable suspension cell lines

Stable human suspension cell lines expressing tagged wild type or mutant copies of EMC subunits were generated and grown as described above. EMC complexes were purified using anti-GFP or anti-ALFA nanobody essentially as described before (Pleiner et al., 2020; Stevens et al., 2023). Cells were harvested by centrifugation for 10 min at 3,000 g and washed in 1xPBS. Cell pellets were resuspended with 6.8 ml solubilization buffer (50 mM HEPES/KOH pH 7.5; 200 mM NaCl; 2 mM MgAc; 1% [w/v] LMNG [Anatrace, USA], 1 mM DTT, 1x complete EDTA-free protease inhibitor cocktail [Roche, Germany]) per 1 g of cell pellet and incubated for 30 min at 4°C. Lysates were cleared by centrifugation for 30 min at 4°C at 18,000 rpm (SS-34 rotor; Beckman-Coulter, USA).

In parallel, Pierce magnetic Streptavidin beads (Thermo Fisher Scientific, USA) were equilibrated in wash buffer (solubilization buffer with 0.0025% [w/v] LMNG) and then incubated with biotinylated anti-GFP or anti-ALFA tag nanobody, purified as described above. After nanobody immobilization, free biotin binding sites were blocked by incubation with wash buffer containing 10 µM dPEG$_{24}$-biotin acid (Quanta Biodesign, USA) for 10 min on ice. Blocked, nanobody-decorated beads were then added to cell lysate for binding to detergent-solubilized ALFA- or GFP-tagged EMC complexes for 1 h at 4°C with head-over-tail mixing. Magnetic beads were then collected and washed three times with wash buffer, before resuspension of the beads in wash buffer containing 250 nM SENP[EuB] protease in a volume amounting to one half of the original bead suspension volume. Protease elution was allowed to proceed for 20 min on ice. All EMC complexes purified for disulfide crosslinking were eluted in wash buffer without DTT.

EMC complexes containing fully replaced cysteine mutant EMC7 variants, were purified via a 2-step procedure using first the C-terminal GFP tag on EMC3 and then the N-terminal ALFA tag on EMC7. The GFP nanobody eluate, obtained by SENP[EuB] cleavage, was diluted twentyfold with wash buffer and incubated with beads containing immobilized ALFA nanobody. The ALFA nanobody was tagged with an orthogonal SUMOStar protease cleavage site and bound EMC was then eluted along with the ALFA nanobody in wash buffer containing 500 nM SUMOStar protease. The resulting eluate was aliquoted and flash frozen in liquid nitrogen. The concentrations of purified EMC complexes for disulfide crosslinking were normalized by measuring GFP fluorescence on a BioTek Synergy HTX plate reader (Agilent, USA). Normalization was verified by SDS-PAGE and Sypro Ruby staining (Thermo Fisher Scientific, USA). If necessary, normalizations were adjusted based on the quantification of Sypro Ruby stained EMC subunit bands in Fiji.

Purification of EMC for structure determination

A suspension-adapted GFP-EMC2 Flp-In 293 T-Rex cell line (Pleiner et al., 2020) was used to purify the EMC for structural analysis. Additionally, EMC7 carrying a C-terminal ALFA tag was introduced into this cell line via lenti-viral transduction as described above. The lenti-viral transfer plasmid encoded EMC7-ALFA fused via a viral 2A sequence to BFP (EMC7-ALFA-2A-TagBFP). BFP fluorescence was used to sort a homogenous stable suspension cell line that ectopically expresses both GFP-EMC2 and EMC7-ALFA. EMC was purified as described above, but with the following minor modifications. Cells were solubilized with solubilization buffer containing 1% glyco-diosgenin (GDN) (Anatrace, USA). The wash buffer contained 0.05% [w/v] GDN. Finally, the EMC eluate was concentrated using an Amicon Ultra 0.5 ml 100K MWCO concentrator (Millipore-Sigma, USA) and further purified via size-exclusion chromatography using a Superose 6 Increase 3.2/300 column (Cytiva, USA) equilibrated in wash buffer (50 mM HEPES/KOH pH 7.5; 200 mM NaCl; 2 mM MgAc; 0.05% [w/v] GDN and 1 mM DTT). Fractions corresponding to the EMC were pooled and concentrated as above to 0.5 mg/ml. To reduce the conformational flexibility of EMC7 at the insertase side, we added stoichiometric amounts of purified ALFA nanobody (Götzke et al., 2019), which binds the C-terminal ALFA tag on EMC7.

Grid preparation and data collection

CryoEM grids were prepared by applying 3 μl of purified EMC at 0.5 mg/mL to glow discharged (60 seconds using a Pelco easiGlow, Emeritech K100X at a plasma current of 20 mA), Holey carbon grids (Quantifoil R1.2/1.3). The sample was blotted for 4-6 sec with filter paper at 8°C, 100% humidity at a -4-blot force prior to plunging into liquid ethane for vitrification using the FEI Vitrobot Mark v4 x2 (Thermo Fisher Scientific, USA). The data set was acquired on a Titan Krios electron microscope (Thermo Fisher Scientific, USA) operated at 300 keV equipped with a K3 direct electron detector and an energy filter (Gatan, USA) with a 20-eV slit width. A total of 11,822 micrographs were collected using 3-by-3 pattern beam image shift, acquiring movies for three non-overlapping areas per hole, using an automated acquisition pipeline in SerialEM (Mastronarde, 2005). Movies were recorded with 40 frames at a magnification of 105,000x in super resolution mode at a calibrated magnification of 0.416 Å/pixel using a dose of 60 e−/Å2 at a dose rate of 16.0 e−/pixel/s and a defocus range of −1.0 to −3.0μm.

<u>Image processing</u>

The data processing workflow is summarized in Figure S2 and was performed using cryoSPARC v.3.3–v.4.0 (Punjani et al., 2017). In short, 11,822 micrographs were motion corrected, dose weighted and down sampled (two-fold to 0.832 Å/pixel) using the Patch Motion followed by patch-based CTF estimation using Patch CTF. 10,206 movies were selected and manually curated using cut-offs for CTF fit (5.0 Å) and total motion (50 pix) for further processing. The particle picking was done using the automated Blob Picker function with particle diameter of 150 to 400 Å. After two rounds of 2D classification, 1,271,124 particles were used for two rounds of heterogeneous ab initio reconstruction (4 volumes), using Maximum/Initial resolution of 9 and 7Å respectively and an Initial/Final minibatch size of 400 and 1,200 particles respectively. Once we obtained an initial map with clear features of the EMC, we reclassified the 1.2 million particles using 3D heterogeneous classification using one well-defined class of the EMC and three decoy classes, using a batch size of 5,000 particles per class and initial low-pass filter of 50 Å. Prior to the final round of classification of 212,440 particles were re-extracted in a box size of 400 pix. The final round of classification yielded a population 193,900 particles that were further refined using non-uniform refinement to obtain a reconstruction at 3.5 Å resolution.

To explore the previously observed flexibility between the lumenal, membrane, and cytoplasmic domains, the particles were subjected to two rounds of 3D-variability analysis/clustering, selecting five modes and a filter resolution ranging from 4.0-8.0 Å. After carefully analyzing each reconstruction, a mode corresponding to a missing subunit of the EMC was identified. The subset of particles was then split into 20 clusters using 3D Variability Analysis Display for this mode. Particles belonging to the nine-subunit complex (156,706 particles) that contained high-resolution features were combined and refined using non-uniform refinement. This yielded a map with a resolution of 3.6 Å, in which we detected a stronger EM density for the TMDs of EMC4 and 7.

Particles belonging to the eight-subunit complex (37,194 particles) were combined and similarly to the nine-subunit complex, the particles were refined using non-uniform refinement. This yielded a map with a resolution of 3.9 Å. All three maps (consensus, 9- and 8-subunit) were post-processed by applying a sharpening B factor of -112 $Å^2$, -103 $Å^2$, and -76 $Å^2$, respectively. Finally, for the analysis of EMC10's TMD position a low-pass filter of 5.5 Å was applied to each map using volume tools in cryoSPARC.

All map resolutions were calculated at the final round of refinement using the gold standard FSC=0.143 criterion from the half maps. Statistic details of the EMC EM maps are reported in Table S3.1.

Model building and refinement

An initial model for the nine-subunit EMC was generated by docking the EMC structure in a lipid nanodisc (PDB: 6WW7) (Pleiner et al., 2020) into the cryo-EM density using UCSF Chimera (Pettersen et al., 2004) followed by an initial round of refinement using Phenix (Liebschner et al., 2019). Next, for the not well-ordered TMDs of EMC4 and 7 high-confidence subcomplexes EMC3 (residues 5-42 and 101-209), EMC4 (59- 155), EMC6 (12-end), and EMC7 (155-178) were generated using AlphaFold2-Multimer ColabFold (AlphaFold2_advanced.ipynb) (Mirdita et al., 2022) and then rigid body fitted into the densities. Finally, all models were combined and further manual refinement was conducted in COOT (Casañal et al., 2020; Emsley et al., 2010). Next, lipids, N-glycans and disulfide bond pairs were added where justified by both the EM density and its chemical environment. Finally, the final model was refined against the 9-subunit map using phenix.real_space_refine. Although we could successfully model a backbone through the contiguous density of the TMDs of

EMC4 and 7, we could not unambiguously assign its registry and therefore these TDMs were assigned as poly-Ala/Gly in the final model. Statistic details of the EMC model are reported in Table 1. Figures were made using PyMol (Schrödinger LLC), and UCSF ChimeraX.

Online supplemental material

Figure S1 shows crosslinking and in cell reporter assay data in support of defining the hydrophilic vestibule as the insertase side of the EMC. Figure S2 shows an overview of the cryo-EM data processing pipeline. Figure S3 shows the updated atomic model of the EMC, as well as biochemical data characterizing the peripheral subunits EMC4, 7 and 10. Figure S4 shows in cell reporter assay and crosslinking data that demonstrate substrate capture by the cytosolic loops of EMC3 and 7. Figure S5 shows data demonstrating that intramembrane residues in EMC4 do not contribute significantly to TA protein insertion, as well as data highlighting the cooperative effect of mitochondrial insertase MTCH2 and the EMC selectivity filter in mitochondrial TA protein sorting. Table 1 lists cryo-EM data collection, refinement, and validation statistics.

3.7 Data availability

The data reported in this work are available in the published article and its online supplemental material. The atomic coordinates and cryo-EM maps have been deposited and openly available in the Protein Data Bank under accession code PDB 8S9S and in the Electron Microscopy Data Bank under accession codes EMDB-40245 (nine-subunit map), EMDB-40246 (consensus map), and EMDB-40247 (eight-subunit map).

3.8 Acknowledgement

Chapter 4

EMC SAMPLES THE PREHANDOVER COMPLEX AND CONNECTS TO THE MULTIPASS TRANSLOCON VIA NOMO COMPLEX

4.1 Abstract

Previous research suggested that Type III IMPs are delivered by the SRP to the ER membrane via the prehandover complex (SRP-SR interaction) and inserted co-translationally into the ER membrane through the EMC. It was speculated that once the first TMD is correctly inserted by the EMC, the ribosome docks onto the SEC61 translocon, allowing the rest of the protein to be synthesized and inserted. However, the exact mechanism of how this handover occurs and the molecular details involved remain unknown. Based on our recent structural analyses of the EMC as described in Chapters 2 and 3 of this thesis, we found that the large cytosolic domain of the EMC prevents interaction with SEC61 translocon when it is bound to the RNC. We therefore hypothesized that the EMC must interact with additional factors that hand off substrates from the SRP to the SEC61 translocon. To understand whether and how the EMC interacts with other factors to facilitate membrane protein biogenesis, we used a combination of structural and functional approaches as described herein.

4.2 Overview

A prototypical Type III IMP is composed of a short amino-terminal sequence followed by a TMD with an N_{exo} topology. Type III IMPs play key roles in many signal transduction pathways and make up the majority of mammalian G protein-coupled receptors (GPCRs) (Chou & Shen, 2007; Hartmann et al., 1989). It is noteworthy that GPCRs are a structurally diverse class of membrane receptors that play crucial roles in both cell signaling and cellular homeostasis (Hauser et al., 2018). Consistent with their essentiality, the incorrect assembly of GPCRs can result in their loss of function with significant consequences for the cell (Hauser et al., 2018). These features make GPCRs an ideal model system for investigating multi-pass IMP biosynthesis. Moreover, approximately one-third of all marketable drugs target GPCRs. (Hauser et al., 2018; Rask-Andersen et al., 2014; Santos et al., 2017). Katie

Page and Dr. Tino Pleiner conducted a recent study in our laboratory that employed genome-wide CRISPR inhibition (CRISPRi) screening and cross-linking mass spectrometry (XL-MS) to identify potential factors necessary for the biogenesis of Type III IMPs. The results identified subunits of the NOMO complex, namely NOMO, NCLN, and TMEM147, as direct interactors of the EMC.

The NOMO complex is present in all metazoans and are composed of three subunits, all of which that span the ER membrane (Figure 4.1A) (Dettmer et al., 2010; Haffner et al., 2004, 2007). NOMO was initially identified for its function in nodal growth, which is critical for the embryonic development of zebrafish, whereas NCLN and TMEM147 are homologues of Nicastrin and APH1, which are constituents of the γ-secretase complex (Dettmer et al., 2010; Haffner et al., 2007). While some evidence suggests that a small proportion of the NOMO complex interacts with a putative insertase, TMCO1, it is also possible that the complex exists independently (McGilvray et al., 2020). However, our initial data suggests that the NOMO complex plays a direct role in the early stages of membrane protein biogenesis by interacting with EMC. This discovery could have significant implications for understanding the complex's role in development.

The discovery of a potential function for the NOMO complex in the biogenesis of Type III IMPs have prompted Katie Page, Dr. Tino Pleiner, and Vy Nguyen to conduct additional experiments. Preliminary data suggests that the NOMO complex is involved in determining the insertion and topology of IMPs (data not shown). However, the observed phenotypes may not be exclusively attributable to the NOMO complex, as they may also be influenced by the EMC, which directly interact with the NOMO complex. Thus, to deepen our understanding of the function of NOMO complex in IMP insertion, an atomic model is needed. This model will uncover the intramembrane surfaces of the NOMO complex, enabling structure-guided mutagenesis to identify crucial functional regions necessary for its biogenesis role. Moreover, the atomic model will allow for the investigation of interfaces that facilitate the interaction between the EMC and NOMO, ultimately leading to the formation of a multipass translocon.

4.3 Results

To determine the structure of the NOMO complex, we established Expi293 cell lines that express a C-terminally GFP-tagged version of TMEM147. Using a GFP-nanobody based purification technique, which we developed in the Voorhees lab, we isolated the heterotrimeric NOMO complex

under native conditions with the aid of GDN detergent. Although we achieved sub-optimal particle distribution and relatively good ice thickness, the resolution of the obtained map of the NOMO complex was low, at 10-14 Å (Figure S4.1). Whereas the density of the lumenal domain was distinct enough to fit the RoseTTAFold protein model (Baek et al., 2021), density of the transmembrane domains was not visible, possibly due to their size in comparison to the GDN micelle, which mainly contributed to the EM density (Figure 4.1 B). Additionally, the presence of disordered detergent molecules surrounding the transmembrane domain along with low signal-to-noise ratio may have led to misalignment that contributed to the limited local resolution.

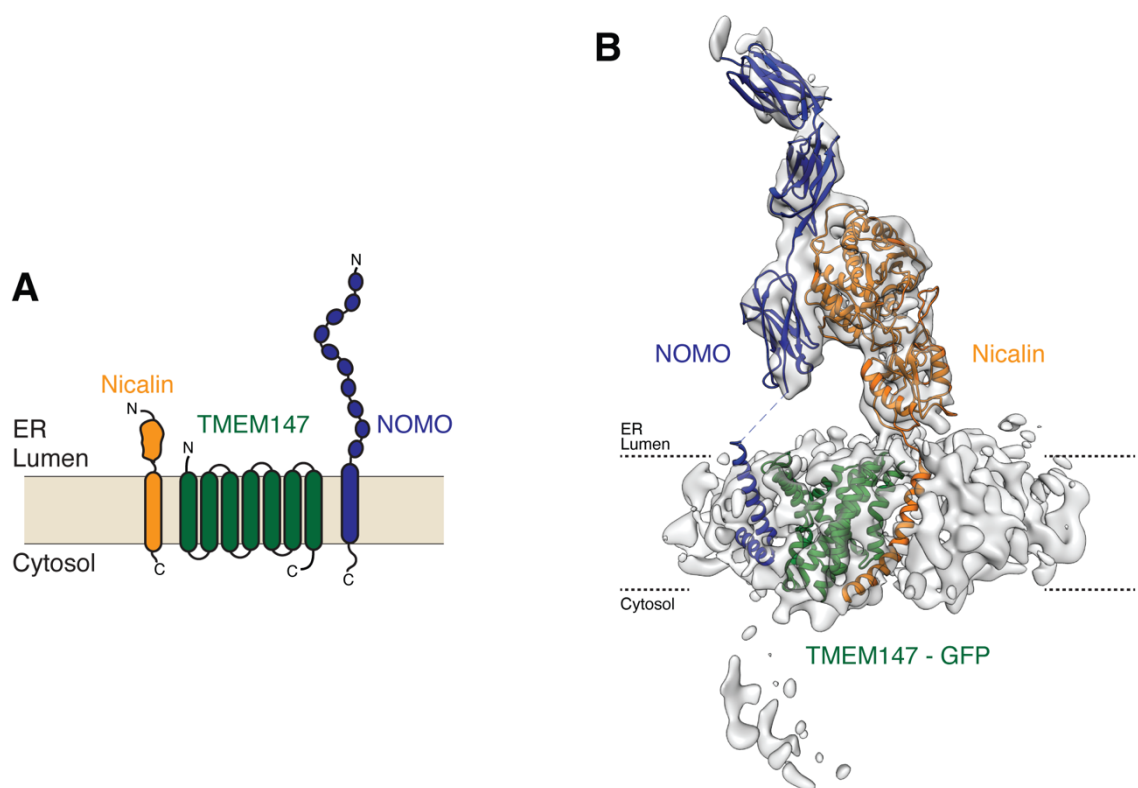Figure 4.1 **Topology model of the NOMO complex.**

(A) Schematic representation of the topology of the NOMO complex, NCLN and NOMO comprise a single transmembrane domain (TMD), while TMEM147 is composed of seven TMDs. NCLN features a lumenal domain with a globular structure, while NOMO's lumenal tail is elongated and composed of ten immunoglobulin-like repeats. (B) A low-resolution three-dimensional EM map of

the NOMO complex, produced using approximately 60,000 particles, resulting in an overall resolution ranging from 10 to 14 Å. CryoSPARC was utilized for data processing (Punjani et al., 2017). The molecular model for each protein within the complex was generated using RoseTTAFold and manually fitted into the map.

Based on our findings from the low-resolution map and RoseTTAFold model, we postulated that the extended tail of NOMO, which consists of 12 predicted IgG-like repeats, could potentially hinder protein purification and adversely impact sample behavior on EM grids due to its possible interaction with the air-water interface. Consequently, we generated a new cell line, referred to as "tNOMO," where we expressed a truncated version of NOMO with only three IgG-like domains along with TMEM147-GFP and NCLN. By co-expressing all three proteins of the NOMO complex, we anticipated that they would supplant the endogenous NOMO, TMEM147, and NCLN, enabling us to purify the tNOMO complex from cells in a stoichiometric manner. Our expectations were validated, as this sample displayed improved behavior in size-exclusion chromatography and eluted as a monodisperse peak (Figure S4.2). Following multiple rounds of sample optimization, we obtained a data set of this sample, which resulted in a reconstruction with an overall resolution of 3.5 Å (3-6 Å in different regions) (Figure 4.2).

The results of our study indicate that the EM map displayed density indicative of the presence of all three subunits. The modest resolution was sufficient to unambiguously assign and build a nearly complete atomic model for the NOMO complex, which extends approximately 160 x 55 x 35 Å (Figure 4.2). The lumenal domain and the TMD of NCLN were well resolved. NCLN contains catalytically inactive aminopeptidase domain as well as one short helix within the lumenal plane of the bilayer, which is similar to the amphipathic EH-1 helix of YidC. It may position the complex within the membrane and locally remodel the bilayer. Five out of the seven transmembrane helices of TMEM147 were at low resolution, indicating local resolution anisotropy that might be caused by some dynamics in the region.

However, for NOMO, we could only fit two IgG-like domains into its density in the lumen, and we could not unambiguously assign or build its TMD into our map. This is possibly due to the highly dynamic nature of this region, which may result in its density being averaged out during the refinement process. Moreover, unlike the low-resolution NOMO map where we were able to fit three

IgG-like domains, the higher resolution EM map only had clear density for two. We hypothesize that the third IgG-like domain became flexible in the absence of the remaining IgG-like domains.
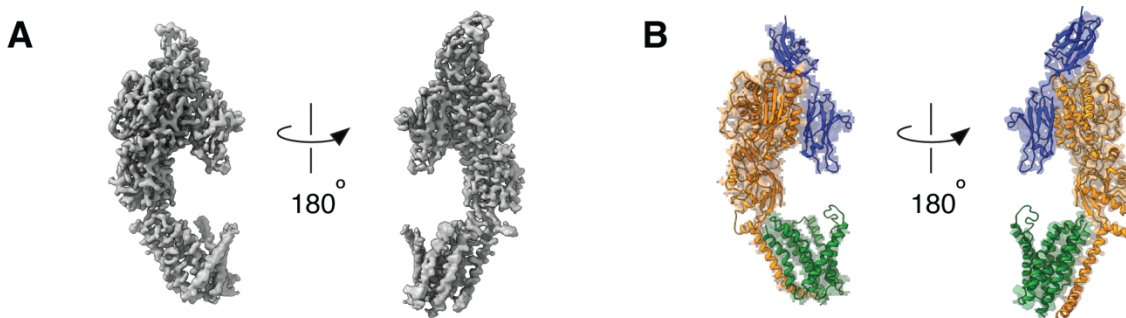
Figure 4.2 **The structure of the NOMO complex.**

(A) Two views of the sharpened EM map of the NOMO complex. (B) Atomic model in the same orientation as the EM map and colored by subunit.

The NOMO complex contains a membrane-spanning region that forms a half-channel in the membrane, which is closed from the cytosol but exposed from the lumen (Figure 4.3A). The local resolution anisotropy in this region suggests that the half-channel is dynamic in the absence of additional factors or substrates. The half-channel is highly conserved and negatively charged, with the bottom being highly conserved but positively charged (Figure 4.3B and C). This is similar to the Hrd1 protein conducting ERAD pathway, which also has a structurally similar hydrophilic half-channel that transports transmembrane segments from the bilayer to the cytosol. This suggests that the half-channel may be involved in how the complex engages substrates during membrane protein biogenesis.

Figure 4.3 **Structure analysis of NOMO complex half-channel.**

(A) Molecular model of the tNOMO construct and top view of the half-channel the TMDs are numbered accordantly. (B) Electrostatic plots (generated using PDB2PQR and APBS with range -10

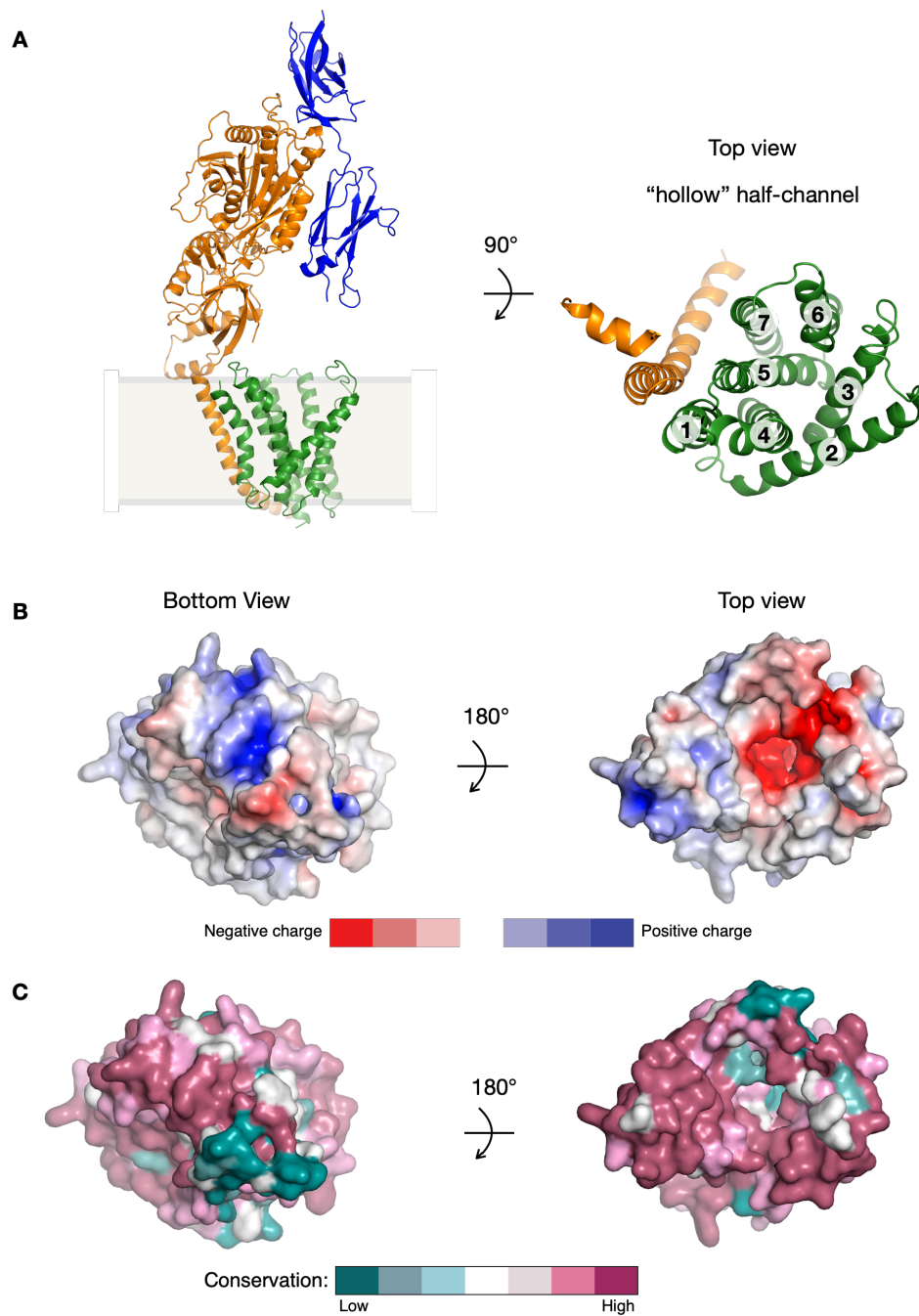to +10 kT/e) mapped onto the surface of structures TMEM147 half-channel (C) Conservation plot mapped onto the structure of NOMO complex half-channel to highlight the high conservation of the top and bottom views.

4.4 Discussion

During the course of our study, a group of researchers observed independently that the NOMO complex is part of a larger multipass translocon complex that aids in the biogenesis of multi-pass membrane proteins (McGilvray et al., 2020). The exact role of the NOMO complex in this process and its biochemical function remained undefined (Hegde & Keenan, 2011). It was postulated that in the context of the multipass translocon, the NOMO complex may function as a scaffolding protein and a molecular modifier that influences the function of Sec61 by regulating the configuration of one of its lumenal loops. Additionally, the NOMO complex may create a protective environment for multi-pass membrane folding (Smalinskaitė et al., 2022; Sundaram et al., 2022). While it was previously assumed that the components of the multipass translocon would not stably associate in the absence of ribosomes (McGilvray et al., 2020), it is possible that each component is modular and capable of acting independently in different contexts and the assembly thus may be mediated by the substrate.

In light of these hypotheses, we proposed another possible function for the NOMO complex based on our findings of both genetic and physical interactions between EMC and the NOMO complex. We propose that the NOMO complex could serve as a "bridge" between EMC and the multipass translocon. Indeed, the TMEM147 subunit of the NOMO complex interacts with a protuberant hairpin in the rRNA of the ribosome. This interaction could be how the EMC samples the prehandover complex and connect to the multipass translocon via NOMO complex. EMC would then insert the first TMD of the Type III IMPs and hand it off to the multipass translocon via interactions with the NOMO complex, enabling the ribosome to dock at SEC61 without the physical constraints presented by EMC (Figure 4.4).
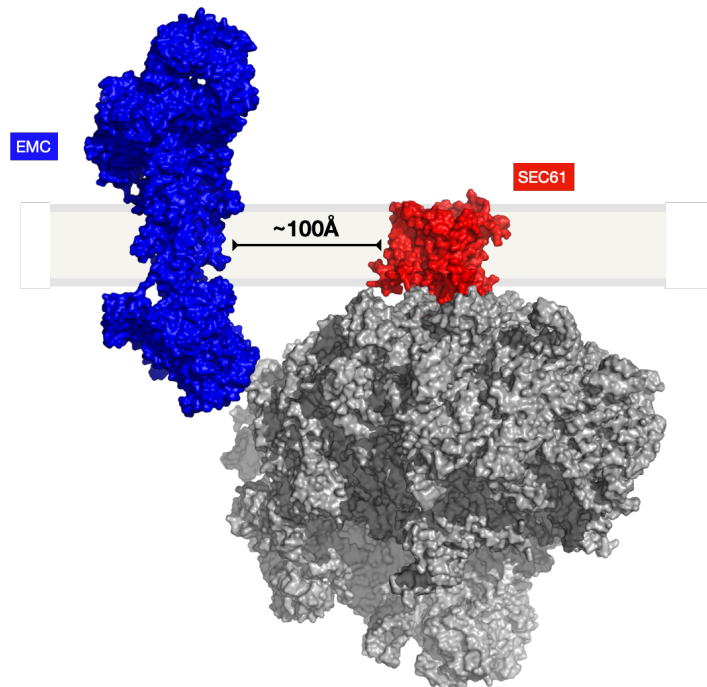
Figure 4.4 **Modelling of co-translational handover to the EMC.**

The large cytosolic domain of the EMC precludes formation of a direct ternary complex between the EMC, SEC61, and the 80S ribosome.

Further study is required to determine the functional relationship between the EMC-NOMO interaction, whether it is involved in the initial assembly of the multipass translocon, or if it facilitates multi-pass membrane protein biogenesis. This initial study establishes an interdisciplinary pipeline that uses genetics, biochemistry, and structural biology to elucidate the mechanism by which EMC interacts with other factors to facilitate membrane protein biogenesis.

4.5 Supplementary Material
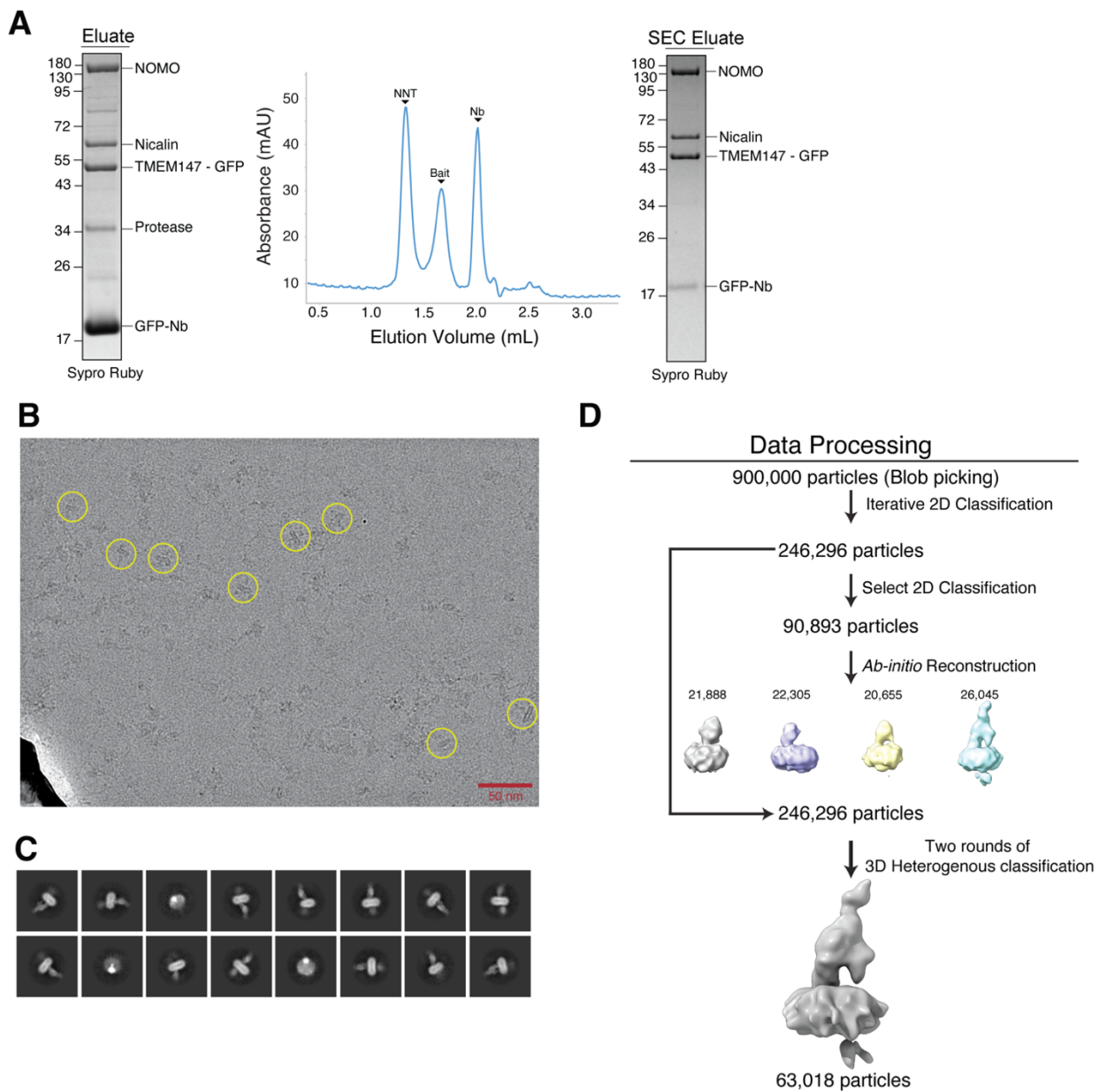
Figure S4.1 **Classification and refinement procedure of NOMO complex.**

(A) Sypro Ruby stained gel of the heterotrimeric NOMO complex purified from human cells using a GFP-nanobody based strategy followed by size exclusion chromatography in the detergent GDN. (B)

A representative micrograph with several particles circled in yellow. (C) Representative 2D class averages after rounds of classification. (D) Overall work flow for data processing to obtain a low-resolution three-dimensional EM map of the NOMO complex.
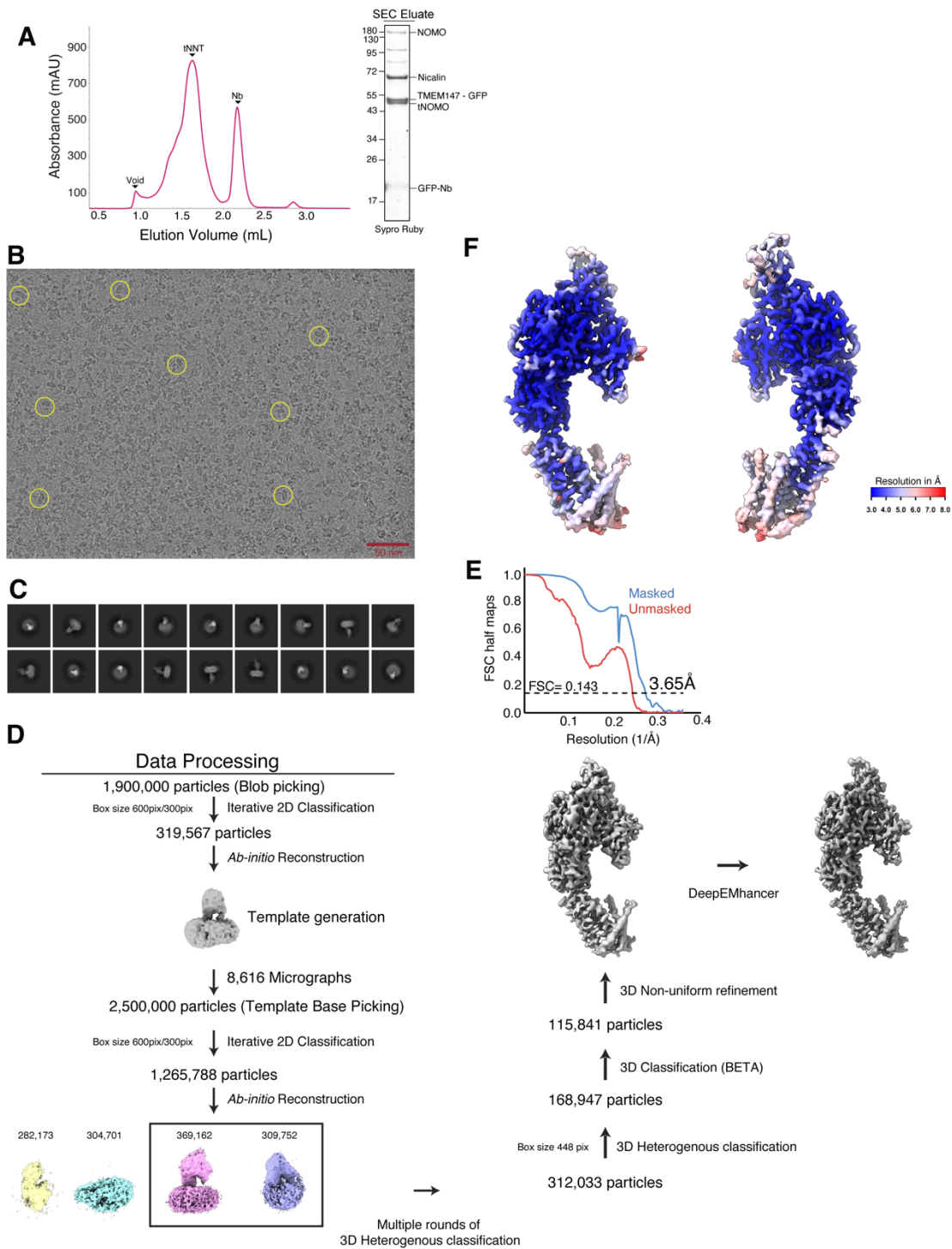
Figure S4.2 **Classification and refinement procedure of tNOMO complex.**

(A) Sypro Ruby stained gel of the heterotrimeric tNOMO complex purified from human cells using a GFP-nanobody based strategy followed by size exclusion chromatography in the detergent GDN. (B) A representative micrograph with several particles circled in yellow. (C) Representative 2D class averages after rounds of classification. (D) Overall workflow for data processing to obtain a high-resolution three-dimensional EM map of the tNOMO complex (F) Cryo-EM density maps colored by local resolution in Å as calculated by cryoSPARC.

C h a p t e r   5

CONCLUSION

Over the past 35 years, research on membrane protein biogenesis has made significant progress, advancing from a basic understanding of a linear pathway with a singular machinery that identifies hydrophobic residues emerging from the ribosome to the identification of multiple factors and pathways involved at different stages. However, our current understanding only provides a preliminary grasp of the process, and despite continued exploration, we still have much to learn about the complex mechanisms involved. Nonetheless, this thesis emphasizes the significance of recent progress in the field of structural biology, particularly the use of single-particle cryo-electron microscopy (cryo-EM). This technology has enabled visualization of one of the machineries involved, thereby providing greater insight into the stages of membrane protein biogenesis.

The study of nascent polypeptides aims to determine how they are directed towards specific machinery based on their sequence features. By understanding this process, we can begin to unravel the molecular mechanisms that govern the assembly of different classes of membrane protein. Ultimately, our goal is to gain a comprehensive understanding of the steps involved in achieving the final assembled state of these proteins. One example of this was the discovery of the EMC about 15 years ago. Since then, researchers have assigned some functions to EMC and have observed its role in a wide range of physiologically important phenotypes. As a result, several groups, including ours, have sought to determine the structure of EMC.

The past two years have been marked by a significant breakthrough in our initial understanding of the mechanism by which the EMC inserts substrates into the lipid bilayer. Through a series of cryo-EM structures and structural base mutagenesis experiments, including our own, we now have valuable insights into this process. For example, we now know that the EMC and YidC share a similar mechanism, involving membrane thinning and polar intramembrane residues, to reduce the energy required to transport a TMD from the cytosol into the membrane. Interestingly, this mechanism is not unique to these two insertases. Other protein conducting channels, such as BAM, Tat, and Hrd1, also

rely on a protein-induced thinned membrane during mammalian membrane protein biogenesis and quality control.

The research described in this thesis has provided answers to longstanding questions while also generating new ones. Specifically, the increased complexity of the EMC compared to YidC and Sec61 indicates that its well-defined insertase function represents only a small portion of its broader role in membrane protein biogenesis and quality control. Moreover, the EMC in humans appears to be more intricate than its counterparts in other species, suggesting a unique mechanism and potential connections to other pathways, indeed we have shown that the EMC serves as an interaction surface to integrate it into other pathways.

Although the EMC targeting pathway is an important mechanism for membrane protein biogenesis, several aspects of the process remain unclear. For example, whereas the lower hydrophobicity of EMC substrates may not require the same level of coordination as the GET pathway, the specific mechanism for efficiently inserting substrates that are stochastically released by CaM or SGTA in any orientation remains elusive. It is unclear whether additional co-factors are involved or if the cytosolic domains of the EMC have a more prominent role in substrate capture, orientation, and release.

We addressed these questions using an improved structural model of the human EMC, employing mutagenesis and site-specific crosslinking to map the path of a TA protein from its cytosolic capture by methionine-rich loops to its membrane insertion through a hydrophilic vestibule. Our findings suggest that positively charged residues at the EMC act as a selectivity filter, employing charge-repulsion to exclude mitochondrial TA proteins. Additionally, this selectivity filter retains the positively charged soluble domains of multipass substrates in the cytosol, ensuring correct topology and enforcing the 'positive-inside' rule.

The EMC's substrate discrimination provides a biochemical explanation for the role of charge in TA protein sorting, preventing protein misinsertion and maintaining compartment integrity. Our study sheds new light on the molecular mechanisms of EMC-mediated membrane protein biogenesis and highlights the importance of the EMC pathway in cellular physiology.

Prior to recent studies, the function of the EMC as a membrane protein chaperone beyond its TMD-insertion role remained unknown. However, the research on the voltage-gated calcium channel has demonstrated that the EMC's chaperone function involves the protection of interaction surfaces and maintenance of proper topology, thereby facilitating multicomplex assembly (Z. Chen et al., 2023). Despite these new insights into the EMC's functions and mechanisms, a comprehensive understanding of its workings remains limited. Further research into this area would help clarify the fundamental processes of membrane protein biogenesis while also shedding light on the basis for the severe consequences of its dysfunction.

BIBLIOGRAPHY

Afonine, P. V., Poon, B. K., Read, R. J., Sobolev, O. V., Terwilliger, T. C., Urzhumtsev, A., & Adams, P. D. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallographica Section D: Structural Biology*, *74*(6), 531–544. https://doi.org/10.1107/S2059798318006551

Ai, H. wang, Shen, W., Sagi, A., Chen, P. R., & Schultz, P. G. (2011). Probing Protein-Protein Interactions with a Genetically Encoded Photo-crosslinking Amino Acid. *ChemBioChem*, *12*(12), 1854–1857. https://doi.org/10.1002/CBIC.201100194

Akopian, D., Shen, K., Zhang, X., & Shan, S. O. (2013). Signal recognition particle: An essential protein-targeting machine. *Annual Review of Biochemistry*, *82*, 693–721. https://doi.org/10.1146/ANNUREV-BIOCHEM-072711-164732

Almén, M. S., Nordström, K. J. V., Fredriksson, R., & Schiöth, H. B. (2009). Mapping the human membrane proteome: A majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biology*, *7*(1), 50. https://doi.org/10.1186/1741-7007-7-50/FIGURES/7

Anghel, S. A., McGilvray, P. T., Hegde, R. S., & Keenan, R. J. (2017). Identification of Oxa1 Homologs Operating in the Eukaryotic Endoplasmic Reticulum. *Cell Reports*, *21*(13), 3708–3716. https://doi.org/10.1016/J.CELREP.2017.12.006

Aviram, N., Ast, T., Costa, E. A., Arakel, E. C., Chuartzman, S. G., Jan, C. H., Haßdenteufel, S., Dudek, J., Jung, M., Schorr, S., Zimmermann, R., Schwappach, B., Weissman, J. S., & Schuldiner, M. (2016). The SND proteins constitute an alternative targeting route to the endoplasmic reticulum. *Nature*, *540*(7631), 134–138. https://doi.org/10.1038/NATURE20169

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Dustin Schaeffer, R., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., Van Dijk, A. A., Ebrecht, A. C., … Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural

network. *Science*, *373*(6557), 871–876. https://doi.org/10.1126/SCIENCE.ABJ8754/SUPPL_FILE/ABJ8754_MDAR_REPRODUCI BILITY_CHECKLIST.PDF

Bagchi, P., Torres, M., Qi, L., & Tsai, B. (2020). Selective EMC subunits act as molecular tethers of intracellular organelles exploited during viral entry. *Nature Communications 2020 11:1*, *11*(1), 1–15. https://doi.org/10.1038/s41467-020-14967-w

Bai, L., You, Q., Feng, X., Kovach, A., & Li, H. (2020). Structure of the ER membrane complex, a transmembrane-domain insertase. *Nature*, *584*(7821), 475–478. https://doi.org/10.1038/S41586-020-2389-3

Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Bye-AJee, H., Cowley, A., Da Silva, A., De Giorgi, M., Dogan, T., Fazzini, F., Castro, L. G., Figueira, L., … Zhang, J. (2017). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, *45*(D1), D158–D169. https://doi.org/10.1093/NAR/GKW1099

Beckett, D., Kovaleva, E., & Schatz, P. J. (2008). A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Science*, *8*(4), 921–929. https://doi.org/10.1110/PS.8.4.921

Beckmann, R., Bubeck, D., Grassucci, R., Penczek, P., Verschoor, A., Blobel, G., & Frank, J. (1997). Alignment of conduits for the nascent polypeptide chain in the ribosome- Sec61 complex. *Science*, *278*(5346), 2123–2126. https://doi.org/10.1126/SCIENCE.278.5346.2123/ASSET/F7DB54DA-00D4-445F-A665-2AE6AC9521CC/ASSETS/GRAPHIC/SE5176124003.JPEG

Bircham, P. W., Maass, D. R., Roberts, C. A., Kiew, P. Y., Low, Y. S., Yegambaram, M., Matthews, J., Jack, C. A., & Atkinson, P. H. (2011). Secretory pathway genes assessed by high-throughput microscopy and synthetic genetic array analysis. *Molecular BioSystems*, *7*(9), 2589–2598. https://doi.org/10.1039/C1MB05175J

Blobel, G. (1980). Intracellular protein topogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *77*(3), 1496. https://doi.org/10.1073/PNAS.77.3.1496

Borgese, N., Brambillasca, S., & Colombo, S. (2007). How tails guide tail-anchored proteins to their destinations. *Current Opinion in Cell Biology*, *19*(4), 368–375. https://doi.org/10.1016/J.CEB.2007.04.019

Borgese, N., & Fasana, E. (2011). Targeting pathways of C-tail-anchored proteins. *Biochimica et Biophysica Acta*, *1808*(3), 937–946. https://doi.org/10.1016/J.BBAMEM.2010.07.010

Borowska, M. T., Dominik, P. K., Anghel, S. A., Kossiakoff, A. A., & Keenan, R. J. (2015). A YidC-like Protein in the Archaeal Plasma Membrane. *Structure*, *23*(9), 1715–1724. https://doi.org/10.1016/J.STR.2015.06.025

Boyd, D., Schierle, C., & Beckwith, J. (1998). How many membrane proteins are there? *Protein Science*, *7*(1), 201–205. https://doi.org/10.1002/PRO.5560070121

Cabantous, S., Terwilliger, T. C., & Waldo, G. S. (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nature Biotechnology*, *23*(1), 102–107. https://doi.org/10.1038/NBT1044

Carvalho, H. J. F., Del Bondio, A., Maltecca, F., Colombo, S. F., & Borgese, N. (2019). The WRB Subunit of the Get3 Receptor is Required for the Correct Integration of its Partner CAML into the ER. *Scientific Reports*, *9*(1). https://doi.org/10.1038/S41598-019-48363-2

Casañal, A., Lohkamp, B., & Emsley, P. (2020). Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Science*, *29*(4), 1069–1078. https://doi.org/10.1002/PRO.3791

Casson, J., McKenna, M., Haßdenteufel, S., Aviram, N., Zimmerman, R., & High, S. (2017). Multiple pathways facilitate the biogenesis of mammalian tail-anchored proteins. *Journal of Cell Science*, *130*(22), 3851–3861. https://doi.org/10.1242/JCS.207829

Chakrabarti, O., & Hegde, R. S. (2009). Functional Depletion of Mahogunin by Cytosolically Exposed Prion Protein Contributes to Neurodegeneration. *Cell*, *137*(6), 1136–1147. https://doi.org/10.1016/j.cell.2009.03.042

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2010). MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, *66*(1), 12–21. https://doi.org/10.1107/S0907444909042073

Chen, Y., Capponi, S., Zhu, L., Gellenbeck, P., Freites, J. A., White, S. H., & Dalbey, R. E. (2017). YidC Insertase of Escherichia coli: Water Accessibility and Membrane Shaping. *Structure*, *25*(9), 1403-1414.e3. https://doi.org/10.1016/J.STR.2017.07.008

Chen, Y., Umanah, G. K. E., Dephoure, N., Andrabi, S. A., Gygi, S. P., Dawson, T. M., Dawson, V. L., & Rutter, J. (2014). M sp1/ ATAD 1 maintains mitochondrial function by facilitating the degradation of mislocalized tail-anchored proteins . *The EMBO Journal*, *33*(14), 1548–1564. https://doi.org/10.15252/EMBJ.201487943

Chen, Z., Mondal, A., Ali, F. A., Jang, S., Niranjan, S., Montaño, J. L., Zaro, B. W., & Minor, D. L. (2023). EMC chaperone-CaV structure reveals an ion channel assembly intermediate. *Nature 2023*, 1–3. https://doi.org/10.1038/s41586-023-06175-5

Chio, U. S., Cho, H., & Shan, S. O. (2017). Mechanisms of tail-anchored membrane protein targeting and insertion. *Annual Review of Cell and Developmental Biology*, *33*, 417–438. https://doi.org/10.1146/ANNUREV-CELLBIO-100616-060839

Chitwood, P. J., & Hegde, R. S. (2019). The Role of EMC during Membrane Protein Biogenesis. *Trends in Cell Biology*, *29*(5), 371–384. https://doi.org/10.1016/j.tcb.2019.01.007

Chitwood, P. J., Juszkiewicz, S., Guna, A., Shao, S., & Hegde, R. S. (2018). EMC Is Required to Initiate Accurate Membrane Protein Topogenesis. *Cell*, *175*(6), 1507-1519.e16. https://doi.org/10.1016/J.CELL.2018.10.009

Chou, K. C., & Shen, H. Bin. (2007). MemType-2L: A Web server for predicting membrane proteins and their types by incorporating evolution information through Pse-PSSM. *Biochemical and*

Biophysical Research Communications, *360*(2), 339–345. https://doi.org/10.1016/J.BBRC.2007.06.027

Christianson, J. C., Olzmann, J. A., Shaler, T. A., Sowa, M. E., Bennett, E. J., Richter, C. M., Tyler, R. E., Greenblatt, E. J., Wade Harper, J., & Kopito, R. R. (2012). Defining human ERAD networks through an integrative mapping strategy. *Nature Cell Biology*, *14*(1), 93–105. https://doi.org/10.1038/NCB2383

Costello, J. L., Castro, I. G., Camões, F., Schrader, T. A., McNeall, D., Yang, J., Giannopoulou, E. A., Gomes, S., Pogenberg, V., Bonekamp, N. A., Ribeiro, D., Wilmanns, M., Jedd, G., Islinger, M., & Schrader, M. (2017). Predicting the targeting of tail-anchored proteins to subcellular compartments in mammalian cells. *Journal of Cell Science*, *130*(9), 1675–1687. https://doi.org/10.1242/JCS.200204

Croll, T. I. (2018). ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallographica Section D: Structural Biology*, *74*(6), 519–530. https://doi.org/10.1107/S2059798318002425

Cymer, F., Von Heijne, G., & White, S. H. (2015). Mechanisms of Integral Membrane Protein Insertion and Folding. *Journal of Molecular Biology*, *427*(5), 999–1022. https://doi.org/10.1016/J.JMB.2014.09.014

De Felipe, P., Luke, G. A., Hughes, L. E., Gani, D., Halpin, C., & Ryan, M. D. (2006). E unum pluribus: Multiple proteins from a self-processing polyprotein. *Trends in Biotechnology*, *24*(2), 68–75. https://doi.org/10.1016/J.TIBTECH.2005.12.006

Demangel, C., & High, S. (2018). Sec61 blockade by mycolactone: A central mechanism in Buruli ulcer disease. *Biology of the Cell*, *110*(11), 237–248. https://doi.org/10.1111/BOC.201800030

Dettmer, U., Kuhn, P. H., Abou-Ajram, C., Lichtenthaler, S. F., Krüger, M., Kremmer, E., Haass, C., & Haffner, C. (2010). Transmembrane protein 147 (TMEM147) is a novel component of the Nicalin-NOMO protein complex. *The Journal of Biological Chemistry*, *285*(34), 26174–26181. https://doi.org/10.1074/JBC.M110.132548

Drin, G., & Antonny, B. (2010). Amphipathic helices and membrane curvature. *FEBS Letters*, *584*(9), 1840–1847. https://doi.org/10.1016/J.FEBSLET.2009.10.022

Elsässer, S. J., Ernst, R. J., Walker, O. S., & Chin, J. W. (2016). Genetic code expansion in stable cell lines enables encoded chromatin modification. *Nature Methods*, *13*(2), 158–164. https://doi.org/10.1038/NMETH.3701

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, *66*(4), 486–501. https://doi.org/10.1107/S0907444910007493

Fairhead, M., & Howarth, M. (2015). Site-specific biotinylation of purified proteins using BirA. *Methods in Molecular Biology*, *1266*, 171–184. https://doi.org/10.1007/978-1-4939-2272-7_12

Favaloro, V., Spasic, M., Schwappach, B., & Dobberstein, B. (2008). Distinct targeting pathways for the membrane insertion of tail-anchored (TA) proteins. *Journal of Cell Science*, *121*(11), 1832–1840. https://doi.org/10.1242/JCS.020321

Frey, S., & Görlich, D. (2014). A new set of highly efficient, tag-cleaving proteases for purifying recombinant proteins. *Journal of Chromatography A*, *1337*, 95–105. https://doi.org/10.1016/J.CHROMA.2014.02.029

Görlich, D., & Rapoport, T. A. (1993). Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. *Cell*, *75*(4), 615–630. https://doi.org/10.1016/0092-8674(93)90483-7

Götzke, H., Kilisch, M., Martínez-Carranza, M., Sograte-Idrissi, S., Rajavel, A., Schlichthaerle, T., Engels, N., Jungmann, R., Stenmark, P., Opazo, F., & Frey, S. (2019). The ALFA-tag is a highly versatile tool for nanobody-based bioscience applications. *Nature Communications 2019 10:1*, *10*(1), 1–12. https://doi.org/10.1038/s41467-019-12301-7

Guna, A., Hazu, M., Tomaleri, G. P., & Voorhees, R. M. (2023). A TAle of Two Pathways: Tail-Anchored Protein Insertion at the Endoplasmic Reticulum. *Cold Spring Harbor Perspectives in Biology*, *15*(3), a041252. https://doi.org/10.1101/CSHPERSPECT.A041252

Guna, A., & Hegde, R. S. (2018). Transmembrane Domain Recognition during Membrane Protein Biogenesis and Quality Control. *Current Biology*, *28*(8), R498–R511. https://doi.org/10.1016/J.CUB.2018.02.004

Guna, A., Stevens, T. A., Inglis, A. J., Replogle, J. M., Esantsi, T. K., Muthukumar, G., Shaffer, K. C. L., Wang, M. L., Pogson, A. N., Jones, J. J., Lomenick, B., Chou, T. F., Weissman, J. S., & Voorhees, R. M. (2022). MTCH2 is a mitochondrial outer membrane protein insertase. *Science*, *378*(6617), 317–322. https://doi.org/10.1126/SCIENCE.ADD1856

Guna, A., Volkmar, N., Christianson, J. C., & Hegde, R. S. (2018). The ER membrane protein complex is a transmembrane domain insertase. *Science*, *359*(6374), 470–473. https://doi.org/10.1126/SCIENCE.AAO3099

Haffner, C., Dettmer, U., Weiler, T., & Haass, C. (2007). The Nicastrin-like protein Nicalin regulates assembly and stability of the Nicalin-nodal modulator (NOMO) membrane protein complex. *The Journal of Biological Chemistry*, *282*(14), 10632–10638. https://doi.org/10.1074/JBC.M611033200

Haffner, C., Frauli, M., Topp, S., Irmler, M., Hofmann, K., Regula, J. T., Bally-Cuif, L., & Haass, C. (2004). Nicalin and its binding partner Nomo are novel Nodal signaling antagonists. *The EMBO Journal*, *23*(15), 3041–3050. https://doi.org/10.1038/SJ.EMBOJ.7600307

Halic, M., Becker, T., Pool, M. R., Spahn, C. M. T., Grassucci, R. A., Frank, J., & Beckmann, R. (2004). Structure of the signal recognition particle interacting with the elongation-arrested ribosome. *Nature 2004 427:6977*, *427*(6977), 808–814. https://doi.org/10.1038/nature02342

Halic, M., & Beckmann, R. (2005). The signal recognition particle and its interactions during protein targeting. *Current Opinion in Structural Biology*, *15*(1 SPEC. ISS.), 116–125. https://doi.org/10.1016/J.SBI.2005.01.013

Harel, T., Yesil, G., Bayram, Y., Coban-Akdemir, Z., Charng, W. L., Karaca, E., Al Asmari, A., Eldomery, M. K., Hunter, J. V., Jhangiani, S. N., Rosenfeld, J. A., Pehlivan, D., El-Hattab, A. W., Saleh, M. A., Leduc, C. A., Muzny, D., Boerwinkle, E., Gibbs, R. A., Chung, W. K., … Lupski, J. R. (2016). Monoallelic and Biallelic Variants in EMC1 Identified in Individuals with

Global Developmental Delay, Hypotonia, Scoliosis, and Cerebellar Atrophy. *American Journal of Human Genetics*, *98*(3), 562–570. https://doi.org/10.1016/J.AJHG.2016.01.011

Hartmann, E., Rapoport, T. A., & Lodish, H. F. (1989). Predicting the orientation of eukaryotic membrane-spanning proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(15), 5786–5790. https://doi.org/10.1073/PNAS.86.15.5786

Haßdenteufel, S., Sicking, M., Schorr, S., Aviram, N., Fecher-Trost, C., Schuldiner, M., Jung, M., Zimmermann, R., & Lang, S. (2017). hSnd2 protein represents an alternative targeting factor to the endoplasmic reticulum in human cells. *FEBS Letters*, *591*(20), 3211–3224. https://doi.org/10.1002/1873-3468.12831

Hauser, A. S., Chavali, S., Masuho, I., Jahn, L. J., Martemyanov, K. A., Gloriam, D. E., & Babu, M. M. (2018). Pharmacogenomics of GPCR Drug Targets. *Cell*, *172*(1–2), 41-54.e19. https://doi.org/10.1016/j.cell.2017.11.033

Hegde, R. S., & Keenan, R. J. (2011). Tail-anchored membrane protein insertion into the endoplasmic reticulum. *Nature Reviews Molecular Cell Biology*, *12*(12), 787–798. https://doi.org/10.1038/NRM3226

Hegde, R. S., & Ploegh, H. L. (2010). Quality and quantity control at the endoplasmic reticulum. *Current Opinion in Cell Biology*, *22*(4), 437–446. https://doi.org/10.1016/j.ceb.2010.05.005

Hein, M. Y., Hubner, N. C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I. A., Weisswange, I., Mansfeld, J., Buchholz, F., Hyman, A. A., & Mann, M. (2015). A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell*, *163*(3), 712–723. https://doi.org/10.1016/J.CELL.2015.09.053

Hennon, S. W., Soman, R., Zhu, L., & Dalbey, R. E. (2015). YidC/Alb3/Oxa1 family of insertases. *Journal of Biological Chemistry*, *290*(24), 14866–14874. https://doi.org/10.1074/JBC.R115.638171

Hessa, T., Meindl-Beinker, N. M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S. H., & Von Heijne, G. (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature*, *450*(7172), 1026–1030. https://doi.org/10.1038/NATURE06387

Ho, C. M., Li, X., Lai, M., Terwilliger, T. C., Beck, J. R., Wohlschlegel, J., Goldberg, D. E., Fitzpatrick, A. W. P., & Zhou, Z. H. (2020). Bottom-up structural proteomics: cryoEM of protein complexes enriched from the cellular milieu. *Nature Methods*, *17*(1), 79–85. https://doi.org/10.1038/S41592-019-0637-Y

Horie, C., Suzuki, H., Sakaguchi, M., & Mihara, K. (2002). Characterization of signal that directs C-tail-anchored proteins to mammalian mitochondrial outer membrane. *Molecular Biology of the Cell*, *13*(5), 1615–1625. https://doi.org/10.1091/MBC.01-12-0570

Horlbeck, M. A., Gilbert, L. A., Villalta, J. E., Adamson, B., Pak, R. A., Chen, Y., Fields, A. P., Park, C. Y., Corn, J. E., Kampmann, M., & Weissman, J. S. (2016). Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *ELife*, *5*(September2016). https://doi.org/10.7554/ELIFE.19760

Inglis, A. J., Page, K. R., Guna, A., & Voorhees, R. M. (2020). Differential Modes of Orphan Subunit Recognition for the WRB/CAML Complex. *Cell Reports*, *30*(11), 3691-3698.e5. https://doi.org/10.1016/J.CELREP.2020.02.084

Isenmann, S., Khew-Goodall, Y., Gamble, J., Vadas, M., & Wattenberg, B. W. (1998). A splice-isoform of vesicle-associated membrane protein-1 (VAMP-1) contains a mitochondrial targeting signal. *Molecular Biology of the Cell*, *9*(7), 1649–1660. https://doi.org/10.1091/MBC.9.7.1649

Itakura, E., Zavodszky, E., Shao, S., Wohlever, M. L., Keenan, R. J., & Hegde, R. S. (2016). Ubiquilins Chaperone and Triage Mitochondrial Membrane Proteins for Degradation. *Molecular Cell*, *63*(1), 21–33. https://doi.org/10.1016/J.MOLCEL.2016.05.020

Jomaa, A., Gamerdinger, M., Hsieh, H. H., Wallisch, A., Chandrasekaran, V., Ulusoy, Z., Scaiola, A., Hegde, R. S., Shan, S. O., Ban, N., & Deuerling, E. (2022). Mechanism of signal sequence handover from NAC to SRP on ribosomes during ER-protein targeting. *Science*, *375*(6583), 839–844. https://doi.org/10.1126/SCIENCE.ABL6459/SUPPL_FILE/SCIENCE.ABL6459_MDAR_REPRODUCIBILITY_CHECKLIST.PDF

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*, *292*(2), 195–202. https://doi.org/10.1006/JMBI.1999.3091

Jonikas, M. C., Collins, S. R., Denic, V., Oh, E., Quan, E. M., Schmid, V., Weibezahn, J., Schwappach, B., Walter, P., Weissman, J. S., & Schuldiner, M. (2009). Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science*, *323*(5922), 1693–1697. https://doi.org/10.1126/SCIENCE.1167983

Jost, M., Chen, Y., Gilbert, L. A., Horlbeck, M. A., Krenning, L., Menchon, G., Rai, A., Cho, M. Y., Stern, J. J., Prota, A. E., Kampmann, M., Akhmanova, A., Steinmetz, M. O., Tanenbaum, M. E., & Weissman, J. S. (2017). Combined CRISPRi/a-Based Chemical Genetic Screens Reveal that Rigosertib Is a Microtubule-Destabilizing Agent. *Molecular Cell*, *68*(1), 210-223.e6. https://doi.org/10.1016/J.MOLCEL.2017.09.012

Juszkiewicz, S., & Hegde, R. S. (2018). Quality Control of Orphaned Proteins. *Molecular Cell*, *71*(3), 443–457. https://doi.org/10.1016/J.MOLCEL.2018.07.001

Kalbfleisch, T., Cambon, A., & Wattenberg, B. W. (2007). A bioinformatics approach to identifying tail-anchored proteins in the human genome. *Traffic*, *8*(12), 1687–1694. https://doi.org/10.1111/J.1600-0854.2007.00661.X

Karpenahalli, M. R., Lupas, A. N., & Söding, J. (2007). TPRpred: A tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. *BMC Bioinformatics*, *8*. https://doi.org/10.1186/1471-2105-8-2

Keenan, R. J., Freymann, D. M., Stroud, R. M., & Walter, P. (2003). The Signal Recognition Particle. *Https://Doi.Org/10.1146/Annurev.Biochem.70.1.755*, *70*, 755–775. https://doi.org/10.1146/ANNUREV.BIOCHEM.70.1.755

Kirchhofer, A., Helma, J., Schmidthals, K., Frauer, C., Cui, S., Karcher, A., Pellis, M., Muyldermans, S., Casas-Delucchi, C. S., Cardoso, M. C., Leonhardt, H., Hopfner, K. P., & Rothbauer, U. (2010). Modulation of protein properties in living cells using nanobodies. *Nature Structural and Molecular Biology*, *17*(1), 133–139. https://doi.org/10.1038/NSMB.1727

Krissinel, E., & Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*, *372*(3), 774–797. https://doi.org/10.1016/J.JMB.2007.05.022

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*, *305*(3), 567–580. https://doi.org/10.1006/JMBI.2000.4315

Kumazaki, K., Chiba, S., Takemoto, M., Furukawa, A., Nishiyama, K. I., Sugano, Y., Mori, T., Dohmae, N., Hirata, K., Nakada-Nakura, Y., Maturana, A. D., Tanaka, Y., Mori, H., Sugita, Y., Arisaka, F., Ito, K., Ishitani, R., Tsukazaki, T., & Nureki, O. (2014). Structural basis of Sec-independent membrane protein insertion by YidC. *Nature*, *509*(7501), 516–519. https://doi.org/10.1038/NATURE13167

Kuroda, R., Ikenoue, T., Honsho, M., Tsujimoto, S., Mitoma, J. Y., & Ito, A. (1998). Charged amine acids at the carboxyl-terminal portions determine the intracellular locations of two isoforms of cytochrome b5. *Journal of Biological Chemistry*, *273*(47), 31097–31102. https://doi.org/10.1074/JBC.273.47.31097

Kutay, U., Ahnert-Hilger, G., Hartmann, E., Wiedenmann, B., & Rapoport, T. A. (1995). Transport route for synaptobrevin via a novel pathway of insertion into the endoplasmic reticulum membrane. *EMBO Journal*, *14*(2), 217–223. https://doi.org/10.1002/J.1460-2075.1995.TB06994.X

Kutay, U., Hartmann, E., & Rapoport, T. A. (1993). A class of membrane proteins with a C-terminal anchor. *Trends in Cell Biology*, *3*(3), 72–75. https://doi.org/10.1016/0962-8924(93)90066-A

Lahiri, S., Chao, J. T., Tavassoli, S., Wong, A. K. O., Choudhary, V., Young, B. P., Loewen, C. J. R., & Prinz, W. A. (2014). A conserved endoplasmic reticulum membrane protein complex (EMC) facilitates phospholipid transfer from the ER to mitochondria. *PLoS Biology*, *12*(10). https://doi.org/10.1371/JOURNAL.PBIO.1001969

Lakshminarayan, R., Phillips, B. P., Binnian, I. L., Gomez-Navarro, N., Escudero-Urquijo, N., Warren, A. J., & Miller, E. A. (2020). Pre-emptive Quality Control of a Misfolded Membrane

Protein by Ribosome-Driven Effects. *Current Biology*, *30*(5), 854-864.e5. https://doi.org/10.1016/J.CUB.2019.12.060

Laskowski, P. R., Pluhackova, K., Haase, M., Lang, B. M., Nagler, G., Kuhn, A., & Müller, D. J. (2021). Monitoring the binding and insertion of a single transmembrane protein by an insertase. *Nature Communications 2021 12:1*, *12*(1), 1–11. https://doi.org/10.1038/s41467-021-27315-3

Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., & Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. *Protein Science*, *27*(1), 129–134. https://doi.org/10.1002/PRO.3289

Li, G. W., Burkhardt, D., Gross, C., & Weissman, J. S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, *157*(3), 624–635. https://doi.org/10.1016/j.cell.2014.02.033

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., … Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology*, *75*, 861–877. https://doi.org/10.1107/S2059798319011471

Lin, K. F., Fry, M. Y., Saladi, S. M., & Clemons, W. M. (2021). Molecular basis of tail-anchored integral membrane protein recognition by the cochaperone Sgt2. *Journal of Biological Chemistry*, *296*. https://doi.org/10.1016/J.JBC.2021.100441

Liu, L., Spurrier, J., Butt, T. R., & Strickler, J. E. (2008). Enhanced protein expression in the baculovirus/insect cell system using engineered SUMO fusions. *Protein Expression and Purification*, *62*(1), 21–28. https://doi.org/10.1016/J.PEP.2008.07.010

Louie, R. J., Guo, J., Rodgers, J. W., White, R., Shah, N. A., Pagant, S., Kim, P., Livstone, M., Dolinski, K., McKinney, B. A., Hong, J., Sorscher, E. J., Bryan, J., Miller, E. A., & Hartman IV, J. L. (2012). A yeast phenomic model for the gene interaction network modulating CFTR-ΔF508 protein biogenesis. *Genome Medicine*, *4*(12). https://doi.org/10.1186/GM404

Mariappan, M., Li, X., Stefanovic, S., Sharma, A., Mateja, A., Keenan, R. J., & Hegde, R. S. (2010). A ribosome-associating factor chaperones tail-anchored membrane proteins. *Nature*, *466*(7310), 1120–1124. https://doi.org/10.1038/NATURE09296

Mariappan, M., Mateja, A., Dobosz, M., Bove, E., Hegde, R. S., & Keenan, R. J. (2011). The mechanism of membrane-associated steps in tail-anchored protein insertion. *Nature*, *477*(7362), 61–69. https://doi.org/10.1038/NATURE10362

Mastronarde, D. N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *Journal of Structural Biology*, *152*(1), 36–51. https://doi.org/10.1016/J.JSB.2005.07.007

Matlack, K. E. S., Mothes, W., & Rapoport, T. A. (1998). Protein translocation: Tunnel vision. *Cell*, *92*(3), 381–390. https://doi.org/10.1016/S0092-8674(00)80930-7

McDowell, M. A., Heimes, M., Fiorentino, F., Mehmood, S., Farkas, Á., Coy-Vergara, J., Wu, D., Bolla, J. R., Schmid, V., Heinze, R., Wild, K., Flemming, D., Pfeffer, S., Schwappach, B., Robinson, C. V., & Sinning, I. (2020). Structural Basis of Tail-Anchored Membrane Protein Biogenesis by the GET Insertase Complex. *Molecular Cell*, *80*(1), 72–86. https://doi.org/10.1016/J.MOLCEL.2020.08.012

McGilvray, P. T., Anghel, S. A., Sundaram, A., Zhong, F., Trnka, M. J., Fuller, J. R., Hu, H., Burlingame, A. L., & Keenan, R. J. (2020). An ER translocon for multi-pass membrane protein biogenesis. *ELife*, *9*, 1–43. https://doi.org/10.7554/ELIFE.56889

McKenna, M. J., Adams, B. M., Chu, V., Paulo, J. A., & Shao, S. (2022). ATP13A1 prevents ERAD of folding-competent mislocalized and misoriented proteins. *Molecular Cell*, *82*(22), 4277-4289.e10. https://doi.org/10.1016/j.molcel.2022.09.035

McKenna, M. J., Sim, S. I., Ordureau, A., Wei, L., Wade Harper, J., Shao, S., & Park, E. (2020). The endoplasmic reticulum P5A-ATPase is a transmembrane helix dislocase. *Science*, *369*(6511). https://doi.org/10.1126/SCIENCE.ABC5809

McKenna, M., Simmonds, R. E., & High, S. (2016). Mechanistic insights into the inhibition of Sec61-dependent co- and post-translational translocation by mycolactone. *Journal of Cell Science*,

129(7), 1404–1415. https://doi.org/10.1242/JCS.182352/260250/AM/MECHANISTIC-INSIGHTS-INTO-THE-INHIBITION-OF-SEC61

McKenna, M., Simmonds, R. E., & High, S. (2017). Mycolactone reveals the substrate-driven complexity of Sec61-dependent transmembrane protein biogenesis. *Journal of Cell Science*, *130*(7), 1307–1320. https://doi.org/10.1242/JCS.198655/265276/AM/MYCOLACTONE-REVEALS-SUBSTRATE-DRIVEN-COMPLEXITY-OF

Meusser, B., Hirsch, C., Jarosch, E., & Sommer, T. (2005). ERAD: the long road to destruction. *Nature Cell Biology 2005 7:8*, *7*(8), 766–772. https://doi.org/10.1038/ncb0805-766

Michael Gromiha, M., & Ou, Y. Y. (2014). Bioinformatics approaches for functional annotation of membrane proteins. *Briefings in Bioinformatics*, *15*(2), 155–168. https://doi.org/10.1093/BIB/BBT015

Miller-Vedam, L. E., Bräuning, B., Popova, K. D., Oakdale, N. T. S., Bonnar, J. L., Prabu, J. R., Boydston, E. A., Sevillano, N., Shurtleff, M. J., Stroud, R. M., Craik, C. S., Schulman, B. A., Frost, A., & Weissman, J. S. (2020). Structural and mechanistic basis of the EMC-dependent biogenesis of distinct transmembrane clients. *ELife*, *9*, 1–124. https://doi.org/10.7554/ELIFE.62611

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature Methods 2022 19:6*, *19*(6), 679–682. https://doi.org/10.1038/s41592-022-01488-1

Mitra, K., Ubarretxena-Belandia, I., Taguchi, T., Warren, G., & Engelman, D. M. (2004). Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(12), 4083–4088. https://doi.org/10.1073/PNAS.0307332101

Moriarty, N. W., Grosse-Kunstleve, R. W., & Adams, P. D. (2009). Electronic ligand builder and optimization workbench (eLBOW): A tool for ligand coordinate and restraint generation. *Acta Crystallographica Section D: Biological Crystallography*, *65*(10), 1074–1080. https://doi.org/10.1107/S0907444909029436

Ng, D. P., Poulsen, B. E., & Deber, C. M. (2012). Membrane protein misassembly in disease. *Biochimica et Biophysica Acta*, *1818*(4), 1115–1122. https://doi.org/10.1016/J.BBAMEM.2011.07.046

Nicolaus, F., Metola, A., Mermans, D., Liljenström, A., Krč, A., Abdullahi, S. M., Zimmer, M., Miller, T. F., & von Heijne, G. (2021). Residue-by-residue analysis of cotranslational membrane protein integration in vivo. *ELife*, *10*, 1–16. https://doi.org/10.7554/ELIFE.64302

O'donnell, J. P., Phillips, B. P., Yagita, Y., Juszkiewicz, S., Wagner, A., Malinverni, D., Keenan, R. J., Miller, E. A., & Hegde, R. S. (2020). The architecture of EMC reveals a path for membrane protein insertion. *ELife*, *9*, 1–56. https://doi.org/10.7554/ELIFE.57887

Okreglak, V., & Walter, P. (2014). 8019-8024 the conserved AAA-ATPase Msp1 confers organelle specificity to tail-anchored proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(22), 8019–8024. https://doi.org/10.1073/PNAS.1405755111

Overington, J. P., Al-Lazikani, B., & Hopkins, A. L. (2006). How many drug targets are there? *Nature Reviews Drug Discovery 2006 5:12*, *5*(12), 993–996. https://doi.org/10.1038/nrd2199

Park, E., & Rapoport, T. A. (2012). Mechanisms of Sec61SecY-mediated protein translocation across membranes. *Annual Review of Biophysics*, *41*(1), 21–40. https://doi.org/10.1146/ANNUREV-BIOPHYS-050511-102312

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/JCC.20084

Pleiner, T., Bates, M., & Görlich, D. (2018). A toolbox of anti-mouse and anti-rabbit IgG secondary nanobodies. *Journal of Cell Biology*, *217*(3), 1143–1154. https://doi.org/10.1083/JCB.201709115

Pleiner, T., Bates, M., Trakhanov, S., Lee, C. T., Schliep, J. E., Chug, H., Böhning, M., Stark, H., Urlaub, H., & Görlich, D. (2015). Nanobodies: Site-specific labeling for super-resolution

imaging, rapid epitope- mapping and native protein complex isolation. *ELife*, *4*(DECEMBER2015). https://doi.org/10.7554/ELIFE.11349

Pleiner, T., Hazu, M., Tomaleri, G. P., Januszyk, K., Oania, R. S., Sweredoski, M. J., Moradian, A., Guna, A., & Voorhees, R. M. (2021). WNK1 is an assembly factor for the human ER membrane protein complex. *Molecular Cell*, *81*(13), 2693-2704.e12. https://doi.org/10.1016/J.MOLCEL.2021.04.013

Pleiner, T., Tomaleri, G. P., Januszyk, K., Inglis, A. J., Hazu, M., & Voorhees, R. M. (2020). Structural basis for membrane insertion by the human ER membrane protein complex. *Science*, *369*(6502), 433–436. https://doi.org/10.1126/SCIENCE.ABB5008

Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, *14*(3), 290–296. https://doi.org/10.1038/NMETH.4169

Rao, M., Okreglak, V., Chio, U. S., Cho, H., Walter, P., & Shan, S. O. (2016). Multiple selection filters ensure accurate tail-anchored membrane protein targeting. *ELife*, *5*(DECEMBER2016). https://doi.org/10.7554/ELIFE.21301

Rapoport, T. A., Li, L., & Park, E. (2017). Structural and mechanistic insights into protein translocation. *Annual Review of Cell and Developmental Biology*, *33*, 369–390. https://doi.org/10.1146/ANNUREV-CELLBIO-100616-060439

Rask-Andersen, M., Masuram, S., & Schiöth, H. B. (2014). The Druggable Genome: Evaluation of Drug Targets in Clinical Trials Suggests Major Shifts in Molecular Class and Indication. *Https://Doi.Org/10.1146/Annurev-Pharmtox-011613-135943*, *54*, 9–26. https://doi.org/10.1146/ANNUREV-PHARMTOX-011613-135943

Richard, M., Boulin, T., Robert, V. J. P., Richmond, J. E., & Bessereau, J. L. (2013). Biosynthesis of ionotropic acetylcholine receptors requires the evolutionarily conserved ER membrane complex. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(11), E1055–E1063. https://doi.org/10.1073/PNAS.1216154110/SUPPL_FILE/PNAS.201216154SI.PDF

Ritchie, T. K., Grinkova, Y. V., Bayburt, T. H., Denisov, I. G., Zolnerciks, J. K., Atkins, W. M., & Sligar, S. G. (2009). Chapter 11 Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. *Methods in Enzymology*, *464*(C), 211–231. https://doi.org/10.1016/S0076-6879(09)64011-8

Rodriguez, A. V., Frey, S., & Görlich, D. (2019). Engineered SUMO/protease system identifies Pdr6 as a bidirectional nuclear transport receptor. *Journal of Cell Biology*, *218*(6), 2006–2020. https://doi.org/10.1083/JCB.201812091

Sanders, C. R., & Myers, J. K. (2004). Disease-Related Misassembly of Membrane Proteins. *Https://Doi.Org/10.1146/Annurev.Biophys.33.110502.140348*, *33*, 25–51. https://doi.org/10.1146/ANNUREV.BIOPHYS.33.110502.140348

Santos, R., Ursu, O., Gaulton, A., Bento, A. P., Donadi, R. S., Bologa, C. G., Karlsson, A., Al-Lazikani, B., Hersey, A., Oprea, T. I., & Overington, J. P. (2017). A comprehensive map of molecular drug targets. *Nature Reviews. Drug Discovery*, *16*(1), 19. https://doi.org/10.1038/NRD.2016.230

Satoh, T., Ohba, A., Liu, Z., Inagaki, T., & Satoh, A. K. (2015). dPob/EMC is essential for biosynthesis of rhodopsin and other multi-pass membrane proteins in Drosophila photoreceptors. *ELife*, *4*(4). https://doi.org/10.7554/ELIFE.06306

Schaffitzel, C., Oswald, M., Berger, I., Ishikawa, T., Abrahams, J. P., Koerten, H. K., Koning, R. I., & Ban, N. (2006). Structure of the E. coli signal recognition particle bound to a translating ribosome. *Nature 2006 444:7118*, *444*(7118), 503–506. https://doi.org/10.1038/nature05182

Schlebach, J. P., Narayan, M., Alford, C., Mittendorf, K. F., Carter, B. D., Li, J., & Sanders, C. R. (2015). Conformational Stability and Pathogenic Misfolding of the Integral Membrane Protein PMP22. *Journal of the American Chemical Society*, *137*(27), 8758–8768. https://doi.org/10.1021/JACS.5B03743/SUPPL_FILE/JA5B03743_SI_001.PDF

Schuldiner, M., Collins, S. R., Thompson, N. J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J. F., Weissman, J. S., & Krogan, N. J. (2005). Exploration

of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell*, *123*(3), 507–519. https://doi.org/10.1016/J.CELL.2005.08.031

Schuldiner, M., Metz, J., Schmid, V., Denic, V., Rakwalska, M., Schmitt, H. D., Schwappach, B., & Weissman, J. S. (2008). The GET Complex Mediates Insertion of Tail-Anchored Proteins into the ER Membrane. *Cell*, *134*(4), 634–645. https://doi.org/10.1016/J.CELL.2008.06.025

Shao, S., & Hegde, R. S. (2011a). A calmodulin-dependent translocation pathway for small secretory proteins. *Cell*, *147*(7), 1576–1588. https://doi.org/10.1016/J.CELL.2011.11.048

Shao, S., & Hegde, R. S. (2011b). Membrane protein insertion at the endoplasmic reticulum. *Annual Review of Cell and Developmental Biology*, *27*, 25–56. https://doi.org/10.1146/ANNUREV-CELLBIO-092910-154125

Shao, S., Rodrigo-Brenni, M. C., Kivlen, M. H., & Hegde, R. S. (2017). Mechanistic basis for a molecular triage reaction. *Science*, *355*(6322), 298–302. https://doi.org/10.1126/SCIENCE.AAH6130

Sharma, A., Mariappan, M., Appathurai, S., & Hegde, R. S. (2010). In vitro dissection of protein translocation into the mammalian endoplasmic reticulum. *Methods in Molecular Biology (Clifton, N.J.)*, *619*, 339–363. https://doi.org/10.1007/978-1-60327-412-8_20

Shurtleff, M. J., Itzhak, D. N., Hussmann, J. A., Schirle Oakdale, N. T., Costa, E. A., Jonikas, M., Weibezahn, J., Popova, K. D., Jan, C. H., Sinitcyn, P., Vembar, S. S., Hernandez, H., Cox, J., Burlingame, A. L., Brodsky, J. L., Frost, A., Borner, G. H. H., & Weissman, J. S. (2018). The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. *ELife*, *7*. https://doi.org/10.7554/ELIFE.37018

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*. https://doi.org/10.1038/MSB.2011.75

Smalinskaitė, L., Kim, M. K., Lewis, A. J. O., Keenan, R. J., & Hegde, R. S. (2022). Mechanism of an intramembrane chaperone for multipass membrane proteins. *Nature*, *611*(7934), 161. https://doi.org/10.1038/S41586-022-05336-2

Soman, R., Yuan, J., Kuhn, A., & Dalbey, R. E. (2014). 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*(2), 1023–1032. https://doi.org/10.1074/JBC.M113.522250

Stefanovic, S., & Hegde, R. S. (2007a). Identification of a Targeting Factor for Posttranslational Membrane Protein Insertion into the ER. *Cell*, *128*(6), 1147–1159. https://doi.org/10.1016/J.CELL.2007.01.036

Stefanovic, S., & Hegde, R. S. (2007b). Identification of a Targeting Factor for Posttranslational Membrane Protein Insertion into the ER. *Cell*, *128*(6), 1147–1159. https://doi.org/10.1016/J.CELL.2007.01.036

Stevens, T. A., Tomaleri, G. P., Hazu, M., Wei, S., Nguyen, V. N., DeKalb, C., Voorhees, R. M., & Pleiner, T. (2023). A nanobody-based strategy for rapid and scalable purification of native human protein complexes. *BioRxiv*.

Sundaram, A., Yamsek, M., Zhong, F., Hooda, Y., Hegde, R. S., & Keenan, R. J. (2022). Substrate-driven assembly of a translocon for multipass membrane proteins. *Nature*, *611*(7934), 167–172. https://doi.org/10.1038/S41586-022-05330-8

Terwilliger, T. C., Adams, P. D., Afonine, P. V., & Sobolev, O. V. (2018). A fully automatic method yielding initial models from high-resolution cryo-electron microscopy maps. *Nature Methods*, *15*(11), 905–908. https://doi.org/10.1038/S41592-018-0173-1

Tian, S., Wu, Q., Zhou, B., Choi, M. Y., Ding, B., Yang, W., & Dong, M. (2019). Proteomic Analysis Identifies Membrane Proteins Dependent on the ER Membrane Protein Complex. *Cell Reports*, *28*(10), 2517-2526.e5. https://doi.org/10.1016/j.celrep.2019.08.006

Trueman, S. F., Mandon, E. C., & Gilmore, R. (2012). A gating motif in the translocation channel sets the hydrophobicity threshold for signal sequence function. *Journal of Cell Biology*, *199*(6), 907–918. https://doi.org/10.1083/JCB.201207163

Van Den Berg, B., Clemons, W. M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C., & Rapoport, T. A. (2003). X-ray structure of a protein-conducting channel. *Nature 2004 427:6969*, *427*(6969), 36–44. https://doi.org/10.1038/nature02218

Volkmar, N., Thezenas, M. L., Louie, S. M., Juszkiewicz, S., Nomura, D. K., Hegde, R. S., Kessler, B. M., & Christianson, J. C. (2019). The ER membrane protein complex promotes biogenesis of sterol-related enzymes maintaining cholesterol homeostasis. *Journal of Cell Science*, *132*(2). https://doi.org/10.1242/JCS.223453

von Heijne, G. (1986). The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *The EMBO Journal*, *5*(11), 3021–3027. https://doi.org/10.1002/J.1460-2075.1986.TB04601.X

Voorhees, R. M., Fernández, I. S., Scheres, S. H. W., & Hegde, R. S. (2014). Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution. *Cell*, *157*(7), 1632–1643. https://doi.org/10.1016/J.CELL.2014.05.024

Voorhees, R. M., & Hegde, R. S. (2016). Structure of the Sec61 channel opened by a signal sequence. *Science*, *351*(6268), 88–89. https://doi.org/10.1126/SCIENCE.AAD4992/SUPPL_FILE/AAD4992S2.MPG

Wallin, E., & Von Heijne, G. (1995). Properties of n-terminal tails in g-protein coupled receptors: A statistical study. *Protein Engineering, Design and Selection*, *8*(7), 693–698. https://doi.org/10.1093/PROTEIN/8.7.693

Walter, P., & Blobel, G. (1983). Preparation of microsomal membranes for cotranslational protein translocation. *Methods in Enzymology*, *96*(C), 84–93. https://doi.org/10.1016/S0076-6879(83)96010-X

Wang, F., Brown, E. C., Mak, G., Zhuang, J., & Denic, V. (2010). A chaperone cascade sorts proteins for posttranslational membrane insertion into the endoplasmic reticulum. *Molecular Cell*, *40*(1), 159–171. https://doi.org/10.1016/J.MOLCEL.2010.08.038

White, S. H., & Von Heijne, G. (2005). Transmembrane helices before, during, and after insertion. *Current Opinion in Structural Biology*, *15*(4), 378–386. https://doi.org/10.1016/J.SBI.2005.07.004

White, S. H., & Wimley, W. C. (1999). Membrane protein folding and stability: Physical principles. *Annual Review of Biophysics and Biomolecular Structure*, *28*, 319–365. https://doi.org/10.1146/ANNUREV.BIOPHYS.28.1.319

Wideman, J. G. (2015). The ubiquitous and ancient ER membrane protein complex (EMC): tether or not? *F1000Research*, *4*, 624. https://doi.org/10.12688/F1000RESEARCH.6944.1

Wu, X., Siggel, M., Ovchinnikov, S., Mi, W., Svetlov, V., Nudler, E., Liao, M., Hummer, G., & Rapoport, T. A. (2020). Structural basis of ER-associated protein degradation mediated by the Hrd1 ubiquitin ligase complex. *Science (New York, N.Y.)*, *368*(6489). https://doi.org/10.1126/SCIENCE.AAZ2449

Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K., & Yokoyama, S. (2008). Multistep Engineering of Pyrrolysyl-tRNA Synthetase to Genetically Encode Nε-(o-Azidobenzyloxycarbonyl) lysine for Site-Specific Protein Modification. *Chemistry and Biology*, *15*(11), 1187–1197. https://doi.org/10.1016/J.CHEMBIOL.2008.10.004

Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., & Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(3), 1496–1503. https://doi.org/10.1073/PNAS.1914677117

Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J. H., Lindahl, E., & Scheres, S. H. W. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *ELife*, *7*. https://doi.org/10.7554/ELIFE.42166