# Using functional genomics to characterize biogenesis and quality control pathways in the mammalian ER

Thesis by

Katharine Rose Page

In Partial Fulfillment of the Requirements for the

Degree of

Doctor of Philosophy


CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2024

Defended September 12, 2023

Katharine Rose Page
ORCID: 0000-0001-8904-1244

# ACKNOWLEDGEMENTS

During my time at Caltech, I have had the privilege of getting to connect and work with so many incredible individuals. The best part of science truly is the people, and I could not have made it through my time here without the support, guidance, and community of many.

First, I want to thank my advisor, Rebecca Voorhees. I joined the lab as one of her first graduate students, and Rebecca gave me the freedom to explore and start new projects in the lab, which strengthened me as a scientist. She gave me the space to learn, to grow, and to make mistakes. She also thoughtfully built the lab, filling it with kind and intelligent people with whom I was able to work alongside of during my graduate career. I would also like to thank my committee, Shu-ou Shan, Doug Rees, and Mitch Guttman, for their support and scientific guidance over the years.

I would not have made it to Caltech without the support of scientific mentors from earlier in my career. Jane Liu was my advisor at Pomona College, and she helped me build a solid foundation in scientific research. She inspired in me the joys of doing science, which I brought with me to Caltech. I had the opportunity to work with Jacob Berlin and Greg Copeland at City of Hope prior to graduate school. I am grateful to Jacob for recommending that I apply to Caltech and for supporting me during my year between Pomona College and Caltech. I am grateful to Greg for making lab work exciting and for sharing his general scientific knowledge and advice with me over the years.

I owe a heartfelt thank you to all of the scientists that I worked alongside of during the course of my time at Caltech. I would like to give a special thank you to members of the Voorhees lab that I worked most closely with. Maxine, Vy, Tino, Alina, and Alison made doing science so incredibly fun. I would also like to thank other members of the Voorhees lab, past and present, including Masami, Gio, Taylor, Erini, Gina, Charlene, Natalie, Akshaye, Ángel, Kurt, and Rob. A big thank you to Mike from the Zinn lab, an honorary Voorhees lab member, who would always share his snacks and great conversation. Though the pandemic made lab work a bit chaotic and challenging, it brought me closer to my lab mates. This is something I am immensely grateful for.

At Caltech I was so lucky to find a community of people that supported me through both the good and the difficult times. I owe a debt of gratitude to my good friends Reina, Sean, Marta, Lincoln,

and Mark. Through all of the birthday celebrations, Thanksgivings, and April Fool's Day hijinks (Reina, I am looking at you), these friends provided me a solid support system and reminded me of what is most important. Living through a lockdown and pandemic while also trying to do lab work was difficult, but Reina, Sean, and Marta as roommates made all the difference. Our daily NYT crosswords, nightly Jeopardy, and the endless quality time together provided a buffer in those uncertain times.

I was fortunate to find support in family friends from my childhood, who provided a home away from home during the holidays when I couldn't be with my family in Texas. Thank you to the Gallizzi and Secrest families for being so kind and welcoming.

I would also like to thank Casey, a friend and roommate, who never fails to make me laugh and gives amazing book recommendations.

Thank you to Wendy, who provided a space for me to learn and grow over the past few years. I would not have been able to do this without you.

I would not have come this far it were not for the support of my family. To my parents, thank you for your patience, for encouraging me to pursue my interests, and for always sharing in my joys when I find something I love doing. To my brother, thank you for being a genuine friend, keeping me grounded, and reminding me that life is to be enjoyed. Thank you to my aunt and uncle, who have been like a second set of parents to me and who give me unconditional love and support. To my grandmother who always keeps me on my toes, thank you for being so authentically you and for the love you give to our family.

And finally, thank you to Rachel. I do not think words can do justice to how lucky I feel to have you by my side. Your kindness, patience, and compassion are immeasurable. You have this uncanny ability to see the light in any situation and you inspire me to appreciate all of the small moments of life. I have so much fun when I am with you, whether that's taking an evening walk together, making breakfast together, or even adopting a cat (or two) together. You have been a pillar of support these past few years, especially recently. Joey, Jamie, and I are so lucky to have you.

# ABSTRACT

Cells are tasked with ensuring the proper synthesis, localization, and insertion of membrane proteins at the mammalian endoplasmic reticulum (ER). Recent advances have shown that the insertion of transmembrane domains into the ER and the translocation of their associated soluble domains across the ER can be facilitated by members of the Oxa1 superfamily of insertases. The ER membrane protein complex (EMC) contains an Oxa1 insertase that facilitates the co-translational insertion of the first TMD of $N_{exo}$ proteins, which have their N-terminal soluble domains localized in the ER lumen or extracellular space. Additionally, recent work has described the multipass translocon, a supercomplex at the Sec61 translocation channel that facilitates insertion of multipass membrane proteins. During my PhD and in collaboration with other scientists in the Voorhees lab, I elucidated how the EMC cooperates with the multipass translocon to facilitate biogenesis of multipass membrane proteins. We took a systematic approach, applying a combination of functional genomics, biochemistry, structural biology, and mechanistic cell biology to understand how the biophysical properties of the TMDs and the intervening soluble domains of multipass membrane proteins influence their path into the ER bilayer. We show that the EMC is epistatic with members of the multipass translocon, including the BOS, GEL, and PAT complexes. We structurally characterize the EMC•BOS holocomplex, showing that these complexes directly interact, and that this interaction is mutually exclusive to the interaction of BOS and Sec61. Further, we demonstrate that $N_{exo}$ proteins that contain a net positive charge in their N-terminal soluble domain are difficult for the EMC to insert and thus also rely on Sec61 or TMCO1, the Oxa1 insertase of the GEL complex, for insertion. We overturn the prevailing model for multipass membrane protein insertion and show that how this diverse class of membrane proteins utilizes the suite of ER biogenesis machinery depends on their distinct biophysical properties. In addition to biogenesis, during my PhD I also studied how the cell surveils multi-subunit complex assembly. I focused on the model ER-resident and obligate complex GET1/GET2. GET1 is unstable and degraded in the absence of GET2, but how it is recognized for degradation was unknown. I used unbiased functional genomics approaches to identify the quality control components that regulate GET1's degradation in the cell, uncovering a novel component in the MARCHF6 ER-associated degradation (ERAD) pathway, TXNDC15. Further, I used proteomics to identify endogenous substrates of the MARCHF6-TXNDC15 ERAD pathway, demonstrating that this pathway regulates proteins critical for maintaining the cytoskeleton and ciliary structure. This work expands upon our understanding of how multi-subunit complexes are regulated by machinery in the cell.

# PUBLISHED CONTENT AND CONTRIBUTIONS

Inglis, A.J., **Page, K.R.,** Guna, A., Voorhees, R.M. (2020) *Differential Modes of Orphan Subunit Recognition for the WRB/CAML Complex*. Cell Rep, 30(11): 3691-3698.e5. doi: 10.1016/j.celrep.2020.02.084
Contributions: K.P. performed *in vitro* and cell experiments, analyzed data, and helped edit and revise the manuscript.

Guna, A.*, **Page, K.***, Replogle, J. R., Esantsi, T. K., Wang, M. L., Weissman, J. S., Voorhees, R.M. (2023) *A dual sgRNA library design to probe genetic modifiers using genome-wide CRISPRi screens*. BMC Genomics, 24(651). doi: 10.1186/s12864-023-09754-y
Contributions: K.P. participated in conceptual design of the project, performed experiments and data analysis, and wrote the manuscript.

**Page, K.R.*,** Nguyen, V.,*, Pleiner, T.*, Tomaleri, G.P., Wang, M.L., Guna, A., Wang, T., Chou, T., and Voorhees, R.M. (2023) *Role of a holo-insertase complex in the biogenesis of biophysically diverse ER membrane proteins*. (Manuscript in revision)
Contributions: K.P. participated in conceptual design of the project, performed experiments and data analysis, and helped assemble the manuscript.

# TABLE OF CONTENTS

# LIST OF ILLUSTRATIONS AND TABLES

*Page*

*Chapter 1*

# INTRODUCTION

## 1.1 Overview

Membrane proteins comprise approximately 25% of the eukaryotic proteome and are essential for a range of functions, including neuronal excitation, intracellular trafficking, and cell signaling (Krogh et al., 2001; von Heijne, 2007). As a result, integral membrane proteins represent major pharmaceutical targets, and defects in their biogenesis underlie a variety of human diseases (Cheng et al., 1990; Chiti & Dobson, 2006). Moreover, nearly 50% of integral membrane proteins must be further assembled into multi-subunit complexes in order to function (Juszkiewicz & Hegde, 2018).

Membrane proteins are characterized by having one or more transmembrane domains (TMDs) that must be integrated into a hydrophobic bilayer within the cell. The number of TMDs, their orientation across the bilayer (i.e., topology), and the biophysical properties within the TMDs vary widely across the membrane proteome (von Heijne, 2007). For example, many membrane proteins contain charged or polar residues in their TMDs, which are required for function, but are otherwise energetically unfavorable in the bilayer. An additional challenge is that nearly all membrane proteins throughout the cell are initially synthesized by ribosomes in the aqueous cytosol, where their hydrophobic TMDs must be shielded by targeting factors prior to their integration into a bilayer (Shao & Hegde, 2011). From their site of synthesis in the cytosol, they must then be localized to their specific destination in the cell via dedicated targeting factors, which can occur co- or post-translationally (Guna, Hazu, et al., 2023; Shan & Walter, 2005). Further, many of these proteins must then assemble into multi-subunit complexes within the ER (Babu et al., 2012; Havugimana et al., 2012).

## 1.2 Membrane protein biogenesis at the ER

The largest site of membrane protein synthesis in the cell is the endoplasmic reticulum (ER), where nearly 5000 unique proteins mature (Krogh et al., 2001; Rapoport et al., 1996). This encompasses a topologically diverse set of proteins, including those targeted co-translationally and post-

translationally. Tail-anchored (TA) proteins contain a single TMD within approximately 30-50 amino acids of their C-terminus and thus are fully released from the ribosome before targeting factors can engage the nascent protein's TMD (Guna, Hazu, et al., 2023). This necessitates that all TAs are targeted to the ER post-translationally, via distinct pathways. In contrast, membrane proteins that are targeted co-translationally vary in number and topology of their TMDs. Singlepass (containing a single TMD) and multipass (containing greater than one TMD) membrane proteins can be further classified based on topology: (i) Type I membrane proteins contain a signal peptide and adopt an $N_{exo}$ topology, where the N-terminus is localized to the ER lumen or extracellular space; (ii) Type II membrane proteins adopt an $N_{cyt}$ topology, where their N-terminus is localized to the cytosol; and (iii) Type III membrane proteins do not contain a signal peptide, but adopt the $N_{exo}$ topology (The UniProt Consortium, 2017; Wallin & von Heijne, 1995).

The signal recognition particle (SRP) recognizes nascent TMDs of multipass membrane proteins and aids in their co-translational delivery to the ER membrane (Halic & Beckmann, 2005; Shan & Walter, 2005). At the ER, two things need to occur during membrane protein insertion: i) the TMD must be inserted and integrated into the lipid bilayer, and ii) the associated soluble domains to either side of the TMDs must get translocated across the bilayer, which is an energetically unfavorable process. These two processes are facilitated by a set of machinery known as insertases (Smalinskaitė & Hegde, 2023). For decades, it was believed that the majority of membrane proteins are inserted via the ribosomally-associated Sec61 translocation channel (Rapoport et al., 2017). Sec61 forms a heterotrimeric complex containing Sec61α, Sec61β, and Sec61γ. The ten TMDs of Sec61α form the protein-conducting pore, where translocating segments can access the ER lumen and TMDs can exit into the lipid bilayer through the lateral gate (Van den Berg et al., 2004). However, even early on, it was clear that not all TMDs were sufficiently hydrophobic to open the lateral gate of Sec61 (Enquist et al., 2009; Schorr et al., 2020).

Advances in the past decade have shed light on the function of additional insertases at the ER membrane. Notably, members of the Oxa1 Superfamily of Insertases have been identified in the ER and have been implicated in biogenesis of membrane proteins (Anghel et al., 2017; Hegde & Keenan, 2022). Oxa1 Superfamily members at the mammalian ER include GET1, EMC3, and TMCO1, each containing similar structural features that enable insertion of TMDs (Bai et al., 2020; McDowell et al., 2020; McGilvray et al., 2020; Miller-Vedam et al., 2020; Pleiner et al., 2020). Highlighting their importance, the Oxa1 Superfamily of Insertases is highly conserved across

kingdoms of life and also includes the bacterial YidC and the archael Ylp1 (Anghel et al., 2017; Hennon et al., 2015). How these insertases facilitate insertion of their client proteins in the context of other biogenesis machinery at the ER, including Sec61, has been a major area of research in recent years.

It has been well known for decades that Sec61 engages a variety of accessory factors during protein translocation and insertion (Rapoport et al., 2017). More recent work, however, has expanded upon these observations and demonstrated that Sec61 engages three additional complexes to specifically facilitate multipass membrane protein biogenesis in higher eukaryotes (McGilvray et al., 2020). This larger complex is termed the multipass translocon (MPT) and consists of the PAT, BOS, and GEL complexes (McGilvray et al., 2020; Smalinskaitė et al., 2022; Sundaram et al., 2022). The PAT complex is composed of CCDC47 and Asterix, which are thought to guide nascent TMDs from the ribosome to the back of Sec61 and to chaperone the TMDs within the bilayer during protein synthesis (Chitwood & Hegde, 2020; Smalinskaitė et al., 2022). The Get and EMC-like (GEL) complex is composed of TMCO1 and its binding partner C20Orf24/OPT1. TMCO1 is a member of the Oxa1 superfamily and is thought to act as the dedicated insertase at the MPT. The back of Sec61 (BOS) complex is an obligate heterotrimeric complex composed of NOMO, Nicalin, and TMEM147 (Dettmer et al., 2010; Haffner et al., 2004). The function of the BOS complex is not fully understood, but it is believed to, at least in part, act as a scaffold of the MPT. In this model, the MPT uses Sec61 as a scaffold and nascent TMDs are instead inserted through TMCO1 (Smalinskaitė et al., 2022; Sundaram et al., 2022). However, how TMCO1 can accommodate the diversity of multipass membrane proteins was unclear from this model.

Prior to the discovery of the MPT, the role of an additional Oxa1 insertase, EMC3, was elucidated. EMC3 is a member of the ER membrane protein complex (EMC), a highly conserved nine subunit complex at the ER composed of EMC1-7, EMC8/9 and EMC10 (Christianson et al., 2012; Jonikas et al., 2009). The EMC was demonstrated to play a critical role in the biogenesis of both tail anchored proteins and $N_{exo}$ multipass membrane proteins (Chitwood et al., 2018; Guna et al., 2018). The EMC inserts the first TMD of $N_{exo}$ multipass membrane proteins, including most G-protein coupled receptors (GPCRs), before they are handed off to the MPT (Chitwood et al., 2018). During this process, the EMC must also translocate the N-terminal soluble domain preceding the first TMD. Recent structural analysis has fueled our understanding of how EMC is able to insert both its co- and post-translational clients (Bai et al., 2020; Miller-Vedam et al., 2020; Pleiner et al., 2020,

2023). The EMC contains a hydrophilic vestibule along the path of TMD insertion, allowing it to translocate N-terminal soluble domains of $N_{exo}$ proteins as well as C-terminal soluble domains of TAs. However, how the ribosome-nascent chain complex containing the remaining TMDs of $N_{exo}$ multipass membrane proteins is handed off and inserted by the MPT after TMD1 insertion by the EMC remained unclear.

### 1.3 Multi-subunit complex assembly

Nearly half of the proteome is organized into multi-subunit complexes. However, how the cell coordinates assembly and quality control (QC) of multimers remains poorly understood. To briefly contextualize the problem, individual subunits often need their binding partners to achieve a properly folded and functional state. In the case of post-translational assembly, subunits are translated individually and then must be co-localized with their partners to enable correct folding. During this biosynthetic process, subunits may require the aid of chaperones or other accessory factors to avoid engaging in unproductive interactions (Juszkiewicz & Hegde, 2018). If the assembly process fails for any reason, for example supernumerary subunit production, the unassembled subunits become QC substrates and are shuttled to degradation machinery. How the cell determines whether and when subunits have failed to assemble is generally unknown. It is likely that the cell has dedicated factors that can (1) chaperone nascent proteins prior to and during assembly; and (2) specifically recognize the subunits and deliver them for degradation if assembly fails. While we lack a detailed mechanistic picture of the biosynthetic and QC processes and the extent to which they are coupled, there is ample support for the active regulation of complexes in cells.

Some of the key evidence for the regulation of complex assembly relies on the observation that individual subunits within the cell are maintained with precise stoichiometry (Harper & Bennett, 2016). One way that the cell accomplishes this task is by degrading subunits that are produced in excess of their cognate partners. In doing so, aggregation of unassembled subunits and any unproductive interactions that might occur with other proteins in their vicinity is avoided. McShane *et al.* demonstrated this cellular strategy by using a quantitative proteomics approach to monitor the decay kinetics of thousands of proteins in the cell (McShane et al., 2016). They found that more than 10% of proteins decay non-exponentially, indicating that this subset of proteins is particularly unstable when first made but then stabilize with age. Many of the proteins exhibiting non-

exponential decay (NED) profiles are supernumerary subunits and comprise the core rather than the periphery of their complexes. This observation is consistent with the idea that once incorporated into a complex, subunits are significantly more stable. Moreover, the findings of this study suggest that these orphan subunits must appear aberrant to the cell and contain molecular features that are flags for their disposal.

A number of studies focusing on the assembly of specific complexes support this model of non-exponential decay. A prime example is the degradation of ribosomal subunits produced in super-stoichiometric amounts relative to their complex by the ubiquitin proteasome system (UPS) (M.-K. Sung et al., 2016). Additionally, individual components of the endoplasmic reticulum (ER)-resident oligosaccharyltransferase (OST) complex and the mitochondrial chaperones Prohibitin 1 and 2 (PHB1 and PHB2) have been shown to be actively degraded when synthesized in excess of their partners (Merkwirth & Langer, 2009; Mueller et al., 2015). Reciprocally, when subunits of the OST and PHB1/2 complexes are genetically knocked down, their partners are degraded. These observations have been made for numerous other multi-subunit complexes, as well, clearly demonstrating that the cell has the ability to post-translationally balance levels of individual subunits (Christianson et al., 2012; Graf et al., 2003; Schulz et al., 2017). However, we still have a limited understanding of the mechanisms, machineries, and pathways within the cell that achieve this function.

### 1.4 Membrane protein quality control at the ER

Membrane protein biosynthesis must be coupled to cellular quality control in order to maintain the proteostasis of the cell (Juszkiewicz & Hegde, 2018). When membrane proteins fail to insert, fold, or assemble properly, they become quality control substrates and are recognized by specific and dedicated machinery. The branch of quality control at the ER is termed ER-associated degradation (ERAD), and its associated machinery is responsible for surveilling the proteome that matures and traffics through the ER (Meusser et al., 2005). ERAD is essential in organisms, and is not only important in the elimination of misfolded proteins, but also functions in the regulation of lipid biosynthesis machinery and in the regulation of calcium ion channels, thereby influencing ion homeostasis in the cell (Krshnan et al., 2022).

Multiple branches of ERAD exist to accommodate the diversity of protein substrates within the ER. ERAD branches, or pathways, function within the ubiquitin proteasome system (UPS), where an E3 ubiquitin ligase complex defines each specific ERAD pathway (Hampton & Sommer, 2012). The E3 ubiquitin ligase complex is involved in substrate recognition and ubiquitination. The catalytic addition of ubiquitin to protein substrates involves three enzymes: ubiquitin-activating (E1), ubiquitin-conjugating (E2), and ubiquitin-ligating (E3) enzymes (Christianson & Ye, 2014). E3 ligases transfer ubiquitin from the E2 conjugating enzyme to the substrate protein, which then marks it for extraction and degradation. Once targeted for degradation by these E3 ligase complexes, substrates are dislocated from the ER membrane by the ATPase p97/VCP (Meyer et al., 2012). The dislocation step is energetically costly and is thus catalyzed. Once in the cytosol, substrates are subject to turnover by the proteasome. Most of what we currently understand about ERAD comes from studies performed in yeast. However, while yeast have three E3 ubiquitin ligase complexes involved ERAD, humans have at least ten (Krshnan et al., 2022). This expansion likely reflects the increased diversity within the human proteome and the need to recognize a broader set of proteins that otherwise could not be efficiently recognized by only three E3 ligase complexes. Further, the greater number of E3 ligase complexes in humans would allow for redundancy, particularly important in post-mitotic cell types that must persist throughout the lifespan (B. Chen et al., 2011; Lansbury & Lashuel, 2006).

Historically in yeast, ERAD has been categorized into three branches: ERAD-L, ERAD-C, and ERAD-M, depending on the location of the degron, or degradation signal (Vembar & Brodsky, 2008). ERAD-L deals with protein substrates in the ER lumen, ERAD-M deals with proteins that are misfolded or contain degrons within the membrane, and ERAD-C deals with proteins with cytosolic degrons. The predominant and most-studied E3 ligase in ERAD in yeast is Hrd1, which performs both ERAD-L and ERAD-M, coordinating with specific cofactors in the ER lumen and membrane (Krshnan et al., 2022). In contrast, the yeast E3 ligase Doa10 which exists both in the ER membrane and inner nuclear membrane functions in ERAD-C and in some cases ERAD-M (Adle & Lee, 2008; Habeck et al., 2015; Smith et al., 2016). Additionally, Doa10 can perform quality control on a set of substrates in the inner nuclear membrane (Swanson et al., 2001). In some cases, there is cooperation within these pathways, whereby substrates require more than just the E3 ligase to be retrotranslocated. For example, the Dfm1 Derlin is required for the degradation and aids in the retrotranslocation of ERAD-M substrates of Doa10 (Neal et al., 2018). However, in higher eukaryotes, ERAD pathways have expanded and thus the designations of which E3 ligases

handle ERAD-L, ERAD-M, and ERAD-C substrates are much less clear. Furthermore, the suite of factors that exist with each E3 ligase complex and how these protein complexes facilitate recognition of their substrates is not fully understood.

The most conserved ERAD branch in eukaryotes is the HRD1 pathway, which functions in ERAD of both ER lumenal and membrane substrates, as in yeast. HRD1, also known as synoviolin (SYVN1), is a membrane-embedded E3 ubiquitin ligase that functions in concert with other co-factors in the ER membrane and ER lumen, including SEL1L/HRD3, Derlin1, OS9, AUP1, and HERP. It functions in concert with the E2 ubiquitin conjugating enzymes UBE2J1 and UBE2G2, both ER-resident membrane proteins. While HRD1 in mammals functions similarly and contains nearly identical cofactors to the yeast Hrd1, the full repertoire of HRD1 substrates in humans has not been elucidated.

Another major ERAD pathway is MARCHF6/TEB4. MARCHF6 (membrane-associated RING C3HC4 finger 6) is homologous to the yeast Doa10, which exists both in the ER and inner nuclear membranes of yeast (Hassink et al., 2005; Kreft et al., 2006). Both MARCHF6 and Doa10 contain fourteen TMDs and are involved in regulation of sterol biosynthesis, tuning their activity to sterol levels in the cell using a sterol sensing domain (Sharpe et al., 2019; Zattas & Hochstrasser, 2015). Substrates of MARCHF6, like Doa10, include soluble cytosolic and membrane proteins. The cytosolic substrates must associate with the ER membrane in order to be substrates of MARCHF6 and Doa10, and they typically engage the ER through amphipathic helices (Chua et al., 2017; Furth et al., 2011). Additionally, the yeast homolog Doa10 has been shown to function in quality control of unassembled subunits of multi-subunit complexes (Habeck et al., 2015). Similar to that in yeast, MARCHF6 requires two E2 conjugating enzymes, UBE2J2 and UBE2G2, homologous to Ubc6 and Ubc7, respectively (Krshnan et al., 2022). UBE2J2 is an integral membrane protein whereas UBE2G2 is a soluble protein that associates with MARCHF6 in the cytosol.

While the HRD1 and MARCHF6 ERAD branches are conserved across eukaryotes, how they enable the quality control of their diverse protein substrates in humans is not well understood. Indeed, most of what we know about how these E3 ligases function has been derived from the interrogation of a small number of well-studied substrates (Hirsch et al., 2009; Krshnan et al., 2022). Moreover, much of the research focus has been on the E3 ligases themselves, but how these enzymes engage substrate and are able to recognize a diverse clientele is poorly understood. We

are only starting to identify and mechanistically dissect the repertoire of co-factors required in each branch of ERAD in humans.

### 1.5 Overview of the thesis

One aim of my thesis research is to elucidate how the suite of biogenesis machinery at the mammalian ER accommodates the diversity of membrane proteins in the proteome to enable their accurate synthesis. We used a combination of genetic, biochemical, structural, and cell biological techniques, including a novel functional genomics tool that we developed. Chapter Two discusses our development of the novel functional genomics tool that we used to query epistatic and parallel biological pathways in the ER. As a proof-of-principle, we used this tool to study parallel biosynthetic pathways for tail-anchored proteins at the ER. Building from this, Chapter Three focuses on how the EMC coordinates with components of the multipass translocon to insert a diverse set of multipass membrane proteins. We employ genome-wide CRISPRi screens of membrane protein substrates, structural characterization of the EMC-BOS holocomplex, biochemical examination of the interaction of the EMC and the multipass translocon, and mechanistic analysis of membrane protein insertion both in cells and in vitro. This research culminates in a unified framework of how the biogenesis machinery at the ER cooperates to insert biophysically and topologically diverse membrane proteins.

A second major aim of my thesis research is understanding how the cell machinery enables membrane protein complex assembly. Chapter Four focuses on the assembly of the model ER-resident complex, GET1/GET2 (formerly WRB/CAML). This work describes how GET2's topology is dependent on GET1 and how GET2 is recognized in the absence of GET1. It also serves as the foundation for Chapter Five, where I identify and characterize a novel quality control component for unassembled subunits at the ER. I utilize a combination of genetic, biochemical, and cell biological approaches to place this newly identified factor within a specific ERAD pathway and further link known pathologies associated with this gene to its role in quality control.

Finally, Chapter Six summarizes the work in this thesis, highlighting how the field of membrane protein biogenesis has changed in the most recent decade and how our understanding of complex assembly and quality control has evolved.

*C h a p t e r  2*

# A DUAL sgRNA LIBRARY DESIGN TO PROBE GENETIC MODIFIERS USING GENOME-WIDE CRISPRi SCREENS

The following chapter is adapted from Guna and Page et al., 2023 and modified according to the Caltech Thesis format.

Guna, A.*, **Page, K.***, Replogle, J. R., Esantsi, T. K., Wang, M. L., Weissman, J. S., Voorhees, R.M. (2023) *A dual sgRNA library design to probe genetic modifiers using genome-wide CRISPRi screens*. BMC Genomics, 24(651). doi: 10.1186/s12864-023-09754-y

## 2.1 Abstract

Mapping genetic interactions is essential for determining gene function and defining novel biological pathways. We report a simple to use CRISPR interference (CRISPRi) based platform, compatible with Fluorescence Activated Cell Sorting (FACS)-based reporter screens, to query epistatic relationships at scale. This is enabled by a flexible dual-sgRNA library design that allows for the simultaneous delivery and selection of a fixed sgRNA and a second randomized guide, comprised of a genome-wide library, with a single transduction. We use this approach to identify epistatic relationships for a defined biological pathway, showing both increased sensitivity and specificity than traditional growth screening approaches.

## 2.2 Introduction

In higher eukaryotes, complex phenotypes are facilitated not only by genetic expansion, but by the combinatorial effects of genes working in concert (Badano & Katsanis, 2002; Hartman IV et al., 2001). Evolutionarily, this complexity affords both genetic redundancy and the ability to undergo rapid cellular adaptation, which ensures phenotypic robustness upon loss or mutation of any particular gene. Indeed, most fundamental processes are buffered by components with partially overlapping function including protein quality control (i.e., protein folding chaperones and E3 ubiquitin ligases), cellular stress response (i.e., the heat shock response and the ubiquitin-proteasome system), and protein biogenesis (i.e., targeting and insertion to the endoplasmic reticulum [ER]) (Morishima et al., 2008; Itakura et al., 2016; Rodina et al., 2016; Rutherford &

Lindquist, 1998; Lehner et al., 2006). However, this creates technical challenges to genetically interrogating biological pathways and assigning gene function in mammalian cells. For example, loss of only ~1/4 of the ~10,000 genes expressed in a typical cell will result in any detectable growth phenotype (Winzeler et al., 1999; Costanzo et al., 2010; Tsherniak et al., 2017; Behan et al., 2019).

To address these challenges, genetic modifier screens have traditionally been a powerful tool for defining gene function, identifying missing components of known pathways, establishing disease mechanisms, and pinpointing new drug targets (Eshed et al., 2001; Eshed et al., 1999; Ding et al., 2016; Hannan et al., 2016; Ahmad et al., 2009; Hurd et al., 2013; Ham et al., 2008; Najm et al., 2018; Diehl et al., 2021; Wong et al., 2016). Forward genetic modifier screens rely on genetic 'anchor points' as a baseline for determining whether subsequent mutations, generally induced through random mutagenesis, result in buffering or synthetic phenotypes. In practice, this 'anchor' is established in a model organism or cell, often requiring extensive manipulation to generate a specific knockout in either organisms or cells, or isogenic mutant cell lines (Soldner et al., 2011; Perreault et al., 2005; Johnston, 2002). Apart from being technically cumbersome, classic forward approaches lack the ability to systematically assess genetic interactions on a genome-wide scale. The advent of CRISPR-based techniques has expanded this ability by allowing for (i) the generation of specific genetic perturbations in the form of knock-outs or knock-downs and (ii) the performance of unbiased genome-wide forward genetic screens to identify the genetic basis of an observed phenotype.

The majority of genetic modifier screens in human cells leverage a CRISPR cutting based approach (Zeng et al., 2019; Hickey et al., 2020; Kramer et al., 2018; Chai et al., 2020; DeWeirdt et al., 2020). However, Cas9-mediated DNA cutting is toxic to cells because it activates the DNA damage response, which is fundamentally problematic for genetic interaction analysis where multiple genomic sites are targeted (Fu et al., 2013; Hsu et al., 2014). Additionally, cells readily adapt and compensate for loss-of-function mutations over time, diminishing observed phenotypes when isogenic knockout cell lines are required (Norman et al., 2019). Moreover, relying on a genetic knock-out approach is often not amenable to the study of essential genes. A more acute strategy, CRISPR interference (CRISPRi), circumvents many of these issues and offers several advantages, notably the ability to create homogenous, titratable knock-down of genes without generating double-stranded DNA breaks (Doench, 2018). CRISPRi relies on a catalytically dead Cas9 (dCAS9) fused to a repressor domain, which, when guided by a sgRNA targeted to a particular

promoter, results in the recruitment of endogenous modulators that lead to epigenetic modifications and subsequently gene knock down (Gilbert et al., 2013; Gilbert et al., 2014; Horlbeck et al., 2016; Qi et al., 2013).

We therefore envision that a strategy to query epistatic relationships acutely and systematically at scale, compatible with the sensitive phenotypic read-out afforded by a fluorescent reporter, would be a powerful tool for assigning genetic function. Towards this goal, we coupled existing CRISPRi technology with a simple and flexible dual-sgRNA library design that is compatible with multi-color FACS-based reporter screens. Our library design, which acutely delivers both a genetic 'anchor point' guide and a second randomized guide in a single plasmid, allows us to perform genetic modifier screens for essential and non-essential genes on a genome-wide scale. As a proof of principle, we applied this approach to dissecting the complex parallel pathways that mediate tail-anchored protein insertion into the endoplasmic reticulum (ER). This approach will be broadly applicable for (i) identifying functional redundancy, (ii) assigning factors to parallel or related biological pathways, and (iii) systematically reveal genetic interactions on a genome-wide scale for a given biological process.

## 2.3 Results

*Dual sgRNA library design and construction*

We developed a strategy to construct and deliver a library containing a fixed pre-determined guide, our genetic anchor point, with a second randomized CRISPRi guide from a single lentiviral backbone at scale (Figure 2.1A). The basis of our second guide is the CRISPRi-v2 library, a compact, validated 5 sgRNA/gene library targeting protein-coding genes in the human genome (Horlbeck et al., 2016). Ease of use was a primary focus of the library design which we addressed by (i) ensuring library construction relied on straightforward and inexpensive restriction enzyme cloning, (ii) developing a sequencing strategy that serves as a failsafe to ensure both guides are present, eliminating potential background, and (iii) designing the library such that the resulting data could be analyzed using an existing computational pipeline.

A necessary requirement of our library design is identification of a pre-verified sgRNA that efficiently targets and depletes your gene of interest. This guide is first introduced by standard restriction enzyme cloning into a human U6 (hU6) and constant region 3 protospacer (CR3), hU6-CR3 cassette using sgRNA DNA oligos that can be inexpensively synthesized and purchased

(Figure 2.1B). Using complementary restriction enzyme sites, the resulting hU6-CR3 cassette is ligated into the CRISPRi-v2 library at scale, resulting in an mU6-CR1-hU6-CR3 guide design (Replogle et al., 2020; Norman et al., 2019). As in the single element CRISPRi-v2 library, BFP and puromycin resistance genes are constitutively expressed, acting as fluorescent and selectable markers to identify guide containing cells.

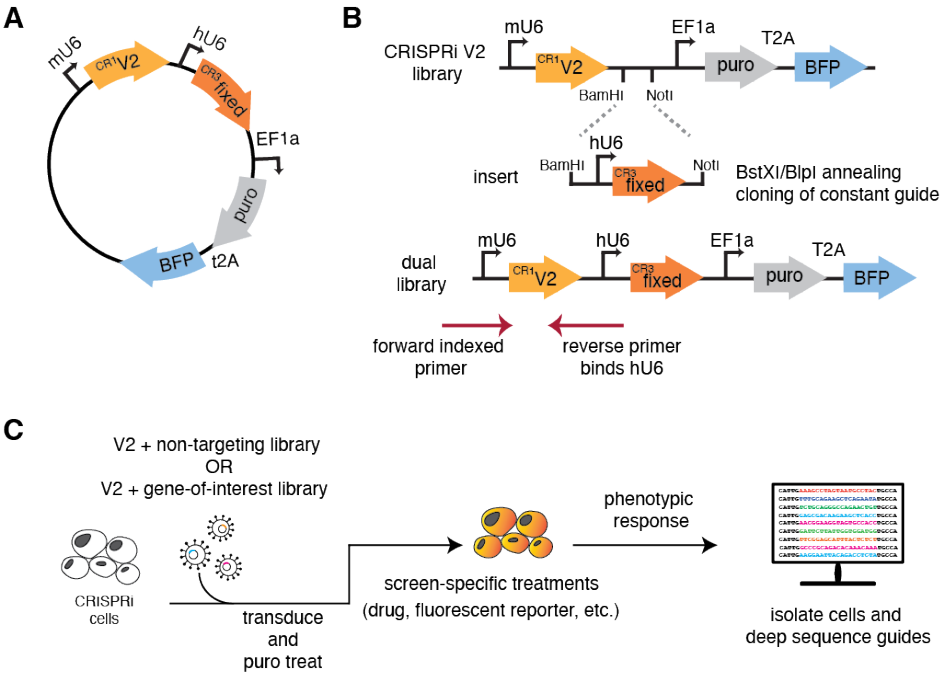



**Figure 2.1 Dual-guide library design and construction. (A)** Schematic of the dual sgRNA vector. Expression of the randomized CRISPRi-v2 sgRNA is driven by a mU6 promoter and the fixed guide is driven by a hU6 promoter, each flanked by unique guide constant regions (CR). Downstream, the EF1a promoter drives the expression of the puromycin resistance selectable marker and BFP. **(B)** Cloning a dual genome-wide library is comprised of two steps. First, a guide of interest is inserted using standard oligo annealing and ligation into a BstXI/BlpI cut backbone. Second, both CRISPRi-v2 library and the fixed guide are digested with complementary restriction sites (BamHI/NotI) and ligated at scale, resulting in an mU6- 'V2 guide'-hU6-'fixed guide' library design. To sequence the resulting library, a standard 5' indexed primer is coupled with a reverse primer that anneals to the hU6 region upstream of the inserted fixed guide. This strategy ensures only guides containing the fixed region are amplified for sequencing. **(C)** A general workflow for using our library design in any CRISPRi machinery containing cell.

Sequencing of the resulting library couples standard barcoded 5' CRISPRi-v2 index primers with a new reverse primer complementary to the hU6 region, thereby only amplifying vectors containing the fixed sgRNA insert. This is important because during library construction, it is possible to produce a small fraction (we estimate <2%) that lack the fixed guide. Additionally, because this cloning strategy involves restriction enzyme digest of the CRISPRi-v2 library, there is loss of a small number of guides that contain these cut sites (~1%, see Supplementary Table 1).

*Putative use of this dual sgRNA library for genetic modifier screening*

To test this procedure, we first generated a library with a verified 'non-targeting' sequence as the fixed guide. Comparison with the standard CRISPRi-v2 library shows that we maintain similar guide coverage across the genome after accounting for expected loss of the restriction site containing guides (Figure S2.1A) (Horlbeck et al., 2016). The resulting sgRNA library allows for the acute knock-down of two separate targets without the need for additional selection markers, which simplifies both growth screens and the more sensitive fluorescent reporter-based flow cytometry screens. This design also removes the need to first make a cell line constitutively expressing a targeting or non-targeting sgRNA, thereby ensuring both the gene-of-interest and the genome-wide library are knocked down for the same period of time, diminishing the possibility of adaptation. Our library design is therefore compatible with a workflow that permits querying epistatic relationships with a variety of phenotypic readouts in any cells expressing the CRISPRi machinery (Figure 2.1C).

To test for genetic interactors at scale, one would conduct a screen using both the non-targeting library we have generated (available from Addgene, Library 197348), and a second library targeting a validated genetic 'anchor point' for your pathway of interest. Comparison of the results of these two screens, in the presence or absence of a characterized pathway component, will uncover and place factors in their respective pathway. We expect three possibilities. (i) Enhanced phenotypes in the 'anchor point' screen suggest synthetic effects, which would be indicative of factors in a parallel pathway. (ii) In contrast, diminished phenotypes in the anchor point screen would suggest factors in the same pathway. (iii) Finally, factors with phenotypes independent of our genetic 'anchor point' likely represent orthogonal genes.

*Developing a reporter assay to assess tail-anchored (TA) protein insertion at the endoplasmic reticulum (ER)*

As a proof of principle, we tested the utility of our dual library by interrogating genetic interactors using a biological system known to contain at least two partially redundant pathways: tail-anchored membrane protein biogenesis. Tail anchored proteins (TAs) carry out essential functions including vesicle trafficking, organelle biogenesis, and cell-to-cell communication (Guna et al., 2022). This family of integral membrane proteins are characterized by a single transmembrane domain (TMD) within 30-50 amino acids of their C terminus (Kutay et al., 1993). The proximity of the TMD to the stop codon necessitates that TAs be targeted and inserted into the membrane post-translationally. Though found in all cellular membranes, the majority of TAs are targeted to the ER using two parallel pathways: the Guided Entry of Tail-anchored protein (GET) and ER membrane complex (EMC) pathways (Stefanovic & Hegde, 2007; Schuldiner et al., 2008; Guna et al., 2018; Guna et al., 2022).

In mammalian cells, the central components of the GET system are the targeting factor GET3, and the ER resident insertase composed of the heterooligomeric GET1/GET2 complex (Vilardi et al., 2011; Vilardi et al., 2014). The EMC pathway relies on targeting by the cytosolic chaperone, Calmodulin to the nine-subunit EMC insertase (Guna et al., 2018). The dependency of a particular TA on either set of factors is determined by hydrophobicity of its TMD, with more hydrophobic substrates relying on the GET, and those less hydrophobic relying on the EMC (Wang et al., 2010; Rao et al., 2016; Shao et al., 2017; Guna et al., 2018). However, substrates of intermediate hydrophobicity can utilize both pathways for targeting and insertion into the ER, potentially obscuring genetic relationships (Guna et al., 2018). We therefore reasoned that our dual-guide screening platform would be ideally suited to identify epistatic relationships between factors in these two pathways (Figure 2.2A).

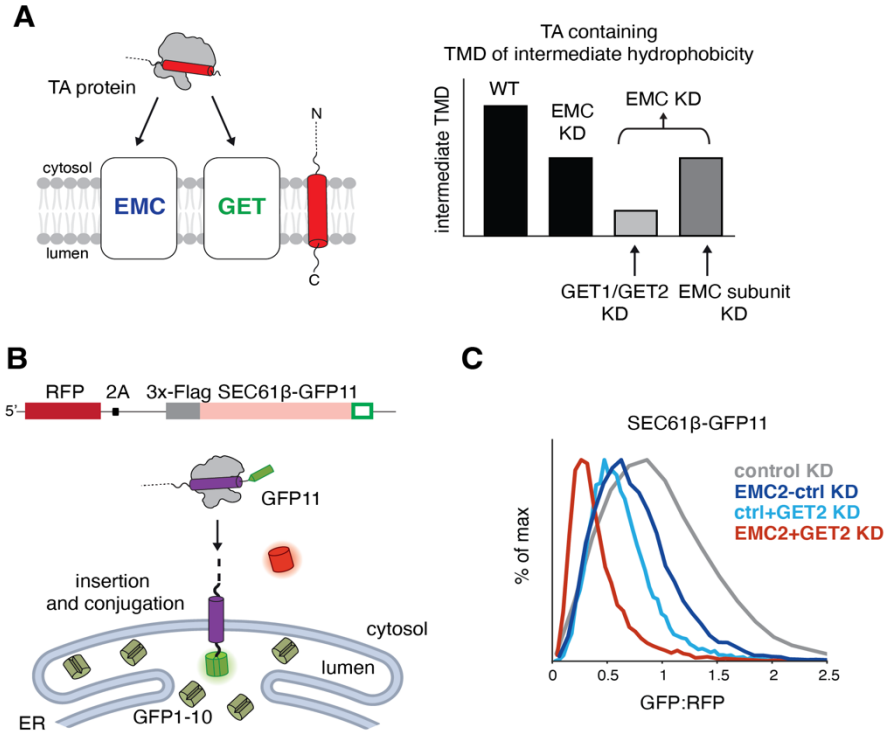



**Figure 2.2 Querying tail-anchored (TA) protein biogenesis at the endoplasmic reticulum (ER).** **(A)** (Left) TA proteins can be inserted into the lipid bilayer by either the EMC or GET insertases. (Right) TAs containing a moderately hydrophobic transmembrane domain such as SEC61β can use either EMC or GET1/GET2 to insert, obscuring strong effects on insertion when obstructing only one of these partially redundant pathways. Therefore, use of an EMC2 fixed guide dual library should uncover defined epistatic relationships between factors operating in either the GET or EMC pathways. **(B)** Schematic of the split GFP reporter system used to assess insertion of SEC61β into the ER. K562 cells expressing CRISPRi machinery were engineered to constitutively express GFP1-10 in the ER lumen. The 11$^{th}$ β -strand of GFP is fused to the C-terminus of SEC61β, allowing for conjugation and fluorescence of the full GFP upon insertion into the ER membrane. RFP is expressed as a normalization marker, separated by a viral P2A sequence. **(C)** Depletion of EMC and GET pathway components in the SEC61β reporter cell line. The SEC61β cell line was separately transduced with dual guides targeting EMC2 alone, GET2 alone, EMC2 and GET2, or a non-targeting control. The GFP:RFP ratio, a measure of SEC61β insertion at the ER, is plotted for each dual guide.

To assess TA biogenesis using a FACS-based approach, we adapted a fluorescent split GFP reporter system to specifically query insertion into the ER (Figure 2.2B). For our reporter substrate, we chose SEC61β, which is an ER-localized TA that normally forms part of the heterotrimeric SEC61 complex (along with Sec61α, and γ) (Görlich et al., 1992; Hartmann et al., 1994; Görlich & Rapoport, 1993; Esnault et al., 1993). SEC61β contains a TMD of intermediate hydrophobicity and is known to use both the EMC and GET pathways for biogenesis (Guna et al., 2018). We constitutively expressed the first 10 β-strands of GFP (GFP1-10) in the ER lumen and appended the 11$^{th}$ β -strand onto the C-terminal of the endogenous sequence of SEC61β (SEC61β-GFP11) (Inglis et al., 2020; Guna et al., 2022). Successful insertion of SEC61β into the ER membrane would therefore result in complementation (GFP11 + GFP1-10) and GFP fluorescence. To generate cell lines compatible for screening, we engineered K562 cells to stably express ER GFP1-10 and the dCas9-KRAB(Kox1) CRISPRi machinery. Under an inducible promoter, we integrated the SEC61β-GFP11 reporter alongside a normalization marker (RFP) separated by a viral 2A sequence (Figure S2.1B). Expression of both the TA and RFP from the same open reading frame allows us to use the GFP:RFP ratio to identify factors involved in TA biogenesis while discriminating against those that have a non-specific effect on protein expression levels (i.e., transcription or translation).

*Interrogating TA insertion into the ER using dual sgRNA libraries*

To permit screening with our dual-guide library design, we constructed a library using a previously validated EMC2 sgRNA as our 'fixed' guide (Figure S2.1C). EMC2 is a core, soluble subunit of the EMC complex, whose depletion leads to the post-translational degradation of the entire EMC via the ubiquitin-proteasome system (Volkmar et al., 2019; Pleiner et al., 2021). Therefore, targeting EMC2 is sufficient to disrupt the EMC pathway for TA insertion, and therefore as our 'genetic anchor.' Using our reporter line, we confirmed using programmed dual guides that loss of both the EMC complex and GET2 resulted in an additive effect of SEC61β insertion (Figure 2.2C). The enhanced effect of loss of GET2 in an EMC knockdown background validate the conceptual premise of our dual-guide screening approach at scale.

We therefore separately used both the EMC2 and a NT control library to transduce our K562 SEC61β reporter cell line, isolated cells that had perturbed GFP:RFP ratios by FACS, and identified the associated guides by deep sequencing. In parallel for comparison, we conducted a traditional growth screen with both the NT and EMC2 libraries in uninduced K562 SEC61β-GFP11 reporter cell lines (Figure S2.2A, Supplementary Table 2). As expected, in the NT-FACS screen loss of

GET pathway components (GET2, GET3, and GET1) and all EMC subunits led to decreased SEC61β-GFP fluorescence, consistent with their established role in TA biogenesis. However, the EMC2-FACS screen showed markedly different results indicative of the genetic relationships between the EMC and GET pathway components (Figure 2.3A). First, when depleted on top of EMC2, the phenotype effects of loss of the main GET pathway factors is enhanced when compared to the NT screen. Second, the majority of guides targeting EMC subunits no longer have significant effects on SEC61β-GFP, consistent with their synergistic role with EMC2 (Volkmar et al., 2019; Pleiner et al., 2021). The exceptions are EMC2, likely because two guides targeting the same gene leads to a greater degree of knockdown, and EMC10, which has been suggested to have a separate regulatory role in TA biogenesis compared to the rest of the EMC complex (Coukos et al., 2021). Conversely, in both screens we also identified several novel ER-resident factors (RNF185, TMEM259 and FAF2) whose depletion leads to increased stability of SEC61β-GFP. Presumably, these putative quality control factors are responsible for recognizing and degrading over-expressed SEC61β from the membrane, but are agnostic to which biogenesis pathway was initially used for its insertion.

To facilitate the comparison of individual screens for identification of genetic interactors, we computed a discriminant rank as a single metric that combines both the phenotype score and the statistical significance of the effects of loss of each gene (Figure 2.3B). This allowed us to visualize the effects of a specific gene on SEC61β stability in the absence or presence of EMC2. Comparison of the NT- and EMC2-genome-wide FACS screens using the discriminant score highlighted the three broad categories of factors we anticipated: members of the GET pathway which show a synthetic effect with EMC, members of the EMC pathway which effectively ‘drop out’ in absence of EMC2, and factors which operate orthogonally from both pathways and are therefore unchanged in the two conditions (Figure 2.3C).

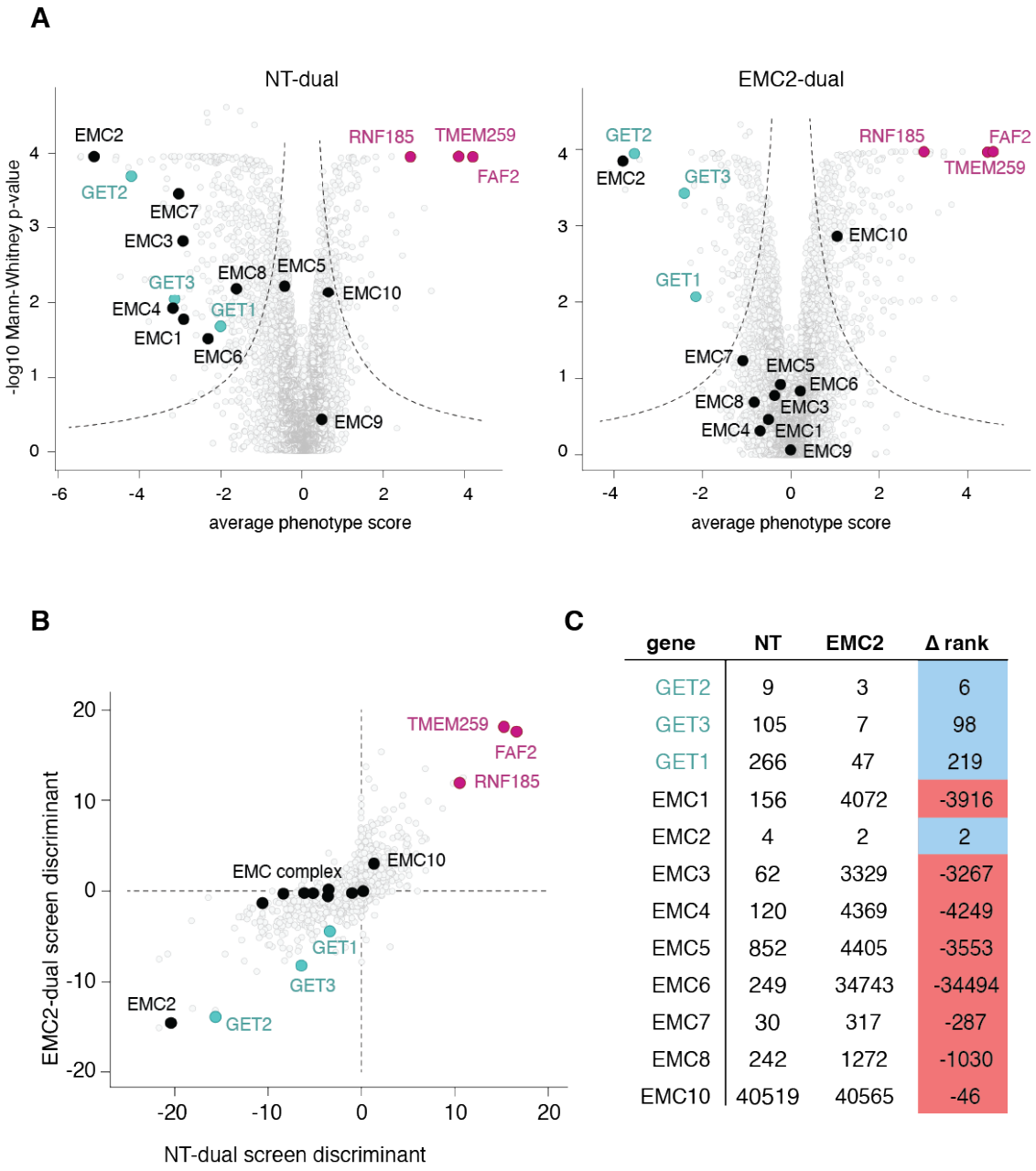



**Figure 2.3 Dual-guide CRISPRi screen with SEC61β reveals genetic interactions between GET and EMC pathway components. (A)** Volcano plot illustrating the phenotype for the three strongest guide RNAs versus Mann-Whitney p-values from two independent replicates of a genome-wide screen with either non-targeting dual (NT) or EMC2-dual libraries using the SEC61β -GFP11 reporter. Individual genes are displayed in gray, core factors of the GET pathway are highlighted in green, EMC subunits are highlighted in black, while putative stabilization factors are in pink. **(B)** A single discriminant score was computed for each gene in the screens investigating

SEC61β -GFP11 stability, representative of the average phenotype score and significance of the hit in the respective screen. This metric allows direct comparison of both NT-dual and EMC2-dual screens. **(C)** Comparison of genes ranked by discriminant score in NT and EMC2-dual screens.

To confirm a subset of the observations predicted by our reporter-based screens, we conducted arrayed assays with programmed dual guides. Using our SEC61β -GFP11 reporter, we show that depletion of both EMC2 and GET3 has an enhanced effect on biogenesis compared to obstructing either pathway individually. This effect is likely specific to substrates of intermediate TMD hydrophobicity, as squalene synthase (SQS), a TA with known EMC dependency is only affected in the absence of EMC2 (Figure 2.4A) (Volkmar et al., 2019). Additionally, depletion of the putative quality control components RNF185, TMEM259 or FAF2 have affects the stability of SEC61β (Figure 2.4B), but not SQS or the GET substrate VAMP (Figure S2.3A). Indeed, RNF185 and TEMEM259 have been recently identified as members of a novel arm of ER-associated degradation (ERAD), while FAF2 has been previously associated with ERAD (van de Weijer et al., 2020; Xu et al., 2013; Lee et al., 2008).

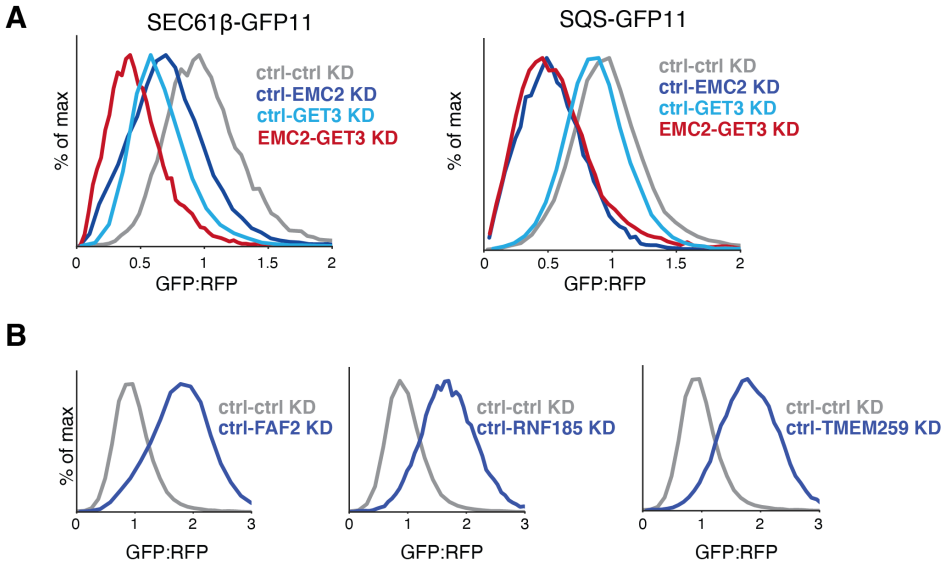



**Figure 2.4 Validating effects of factors on TA biogenesis. (A)** Integration of the TA proteins SEC61β -GFP11 or SQS-GFP11 into the ER was assessed in K562 cells that expressed the indicated programmed dual guides. GFP fluorescence is shown relative to a normalization marker (RFP) as determined by flow cytometry, and the results displayed as a histogram. **(B)** Biogenesis

of SEC61β -GFP11 was assessed as in (A) with the presence of guides targeting the indicated genetic targets.

To illustrate the efficacy of our strategy, we compared the results of our FACS-based dual-guide library screen to a more traditional growth screening approach. Using growth as the metric, there is no increased genetic reliance on GET pathway components in the absence of EMC2 (Figure S2.2B, Supplementary Table 3). This is consistent with the observation that a substantial number of genes with transcriptional phenotypes have negligible growth phenotypes (Replogle et al., 2020). The significant number of hits both in the presence and absence of EMC2 are essential genes, occluding the possibility of detecting significant factors in the context of a particular biological pathway (Figure S2.2C). Given these results, hits identified from the growth screening approach would be particularly prone to off-pathway false positive and false negatives, necessitating substantial more follow-up to identify bona fide genetic interactors of the EMC. If we assume no previous knowledge of the relationship between the EMC and GET pathways, the growth-based approach clearly fails to identify genetic interactions that are crucial to elucidating its biological function. Thereby illustrating the efficacy and potential utility of our dual-guide screening approach.

**2.4 Discussion**

We have developed a flexible, straightforward strategy to rapidly assess genetic interactions genome wide with high efficiency. Successful implementation of this approach does require sufficient prior knowledge of pathway or candidate gene of interest both to identify the fixed guide and design and validate an appropriate fluorescent reporter. However, the dual-guide strategy offers several practical advantages over existing genetic modifier screening strategies. Our approach eliminates the need to create and characterize a knock-out line for a particular gene of interest (Hickey et al., 2020; Feng et al., 2022; Westermann et al., 2022; Fu et al., 2013; Rossi et al., 2015). It also allows for the simultaneous delivery and selection of both targeted and genome-wide elements, resulting in less cell line construction and manipulation. The dual-guide library approach is compatible with multiple screening modalities while allowing for genome-wide perturbations, notably flow cytometry-based approaches where number of fluorophores may be limited. Finally, construction and use of new libraries is easy and rapid, with a two-step cloning process and reliance on existing sequencing and analysis pipelines. However, one minor caveat of the dual-guide system is that the addition of a second guide delivered on the same plasmid diminishes the efficiency of

the fixed guide, but not by a significant amount (Figure S2.3B-C). This is evident in our system, with EMC2 coming out as a significant hit in the EMC2-dual-guide reporter screen. This can be ameliorated by the selection of a fixed guide that independently results in efficient knock-down, and the use of the recently described Zim3-Cas9 effector system, which has been shown to have stronger on-target knockdown compared to KOX1-Cas9 while maintaining minimal non-specific genome-wide effects (Alerasool et al., 2020; Replogle et al., 2022).

Recent studies have highlighted the success of FACS-based CRISPRi screens for the discovery of new factors (Guna et al., 2022; Leto et al., 2019; Tsai et al., 2022; Morita et al., 2018). Our library extends the use case of this approach, for example allowing the study of processes that have parallel compensatory pathways, such as protein biosynthesis and degradation. Though a single screen with a programmed guide containing dual library is sufficient for most applications, performing an additional screen with a NT guide containing dual library provides additional data that could reveal critical genetic interactions. Another implication of our work is the relative paucity of information in traditional growth-based screens, with no additional perturbations. Moving forward, we propose that additional up-front investment in developing a more targeted phenotypic read out, whether it be sensitivity to a compound or a reporter, is worthwhile when trying to establish genome-wide genetic interactions.

**2.5 Conclusion**

The ability to genetically interrogate a biological process in mammalian cells on a genome-wide scale is a powerful tool to determine gene function. Here, we propose a simple advance to current CRISPRi sgRNA library construction that couples a genome-wide library with the simultaneous knock-down of a particular gene of interest. As a proof of principle, we use this design with a FACS-based reporter screen to show the relationships between the parallel pathways that mediate the insertion of TA proteins into the endoplasmic reticulum (ER). We not only faithfully reveal the known factors involved in this process, but can place them in either the GET or EMC pathways. We envision that these screening approaches represent a powerful strategy to unbiased and systematic identification of genetic interactors, capable of de-orphaning

## 2.6 Supplementary Material

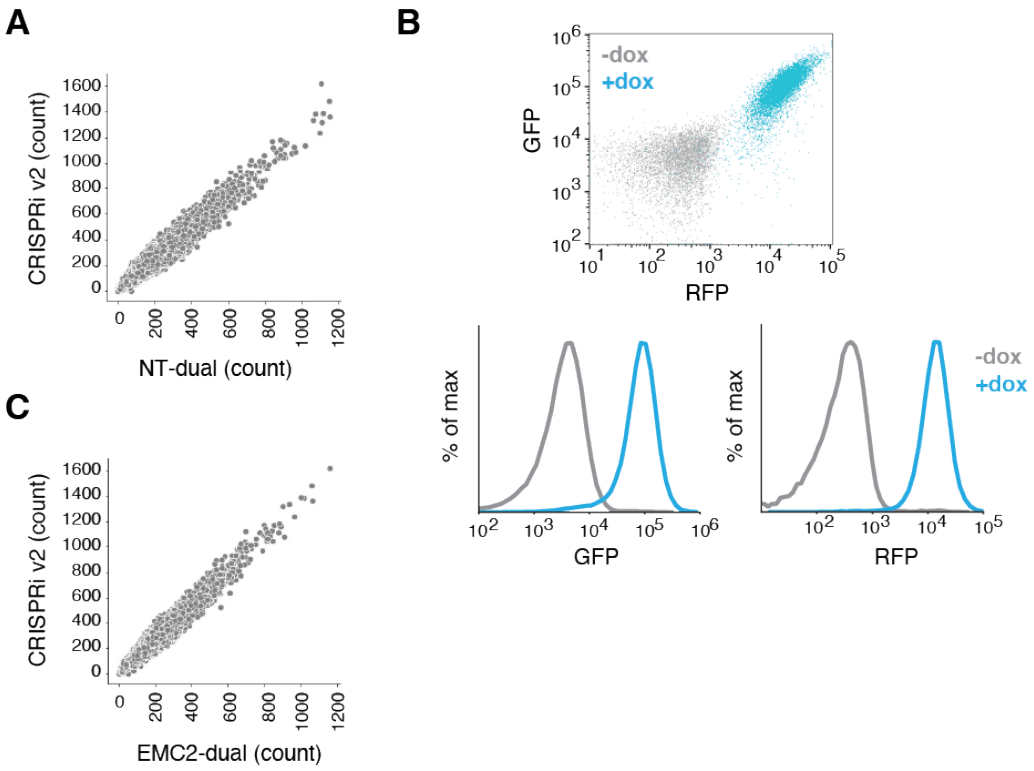



**Figure S2.1. Dual library guide coverage and reporter line characterization. (A)** Coverage of genome-wide guides in the NT dual library. Comparison of guide counts from the single CRISPRi-v2 library and the NT-dual library, after excluding guides which drop out due to restriction enzyme cutting during library construction. **(B)** K562 cells expressing GFP1-10 in the ER lumen and the SEC61β -GFP11 reporter under an inducible promoter are treated with doxycycline and analyze by flow cytometry. Green and red channels are shown separately. **(C)** As in (A) for the EMC2 dual library.

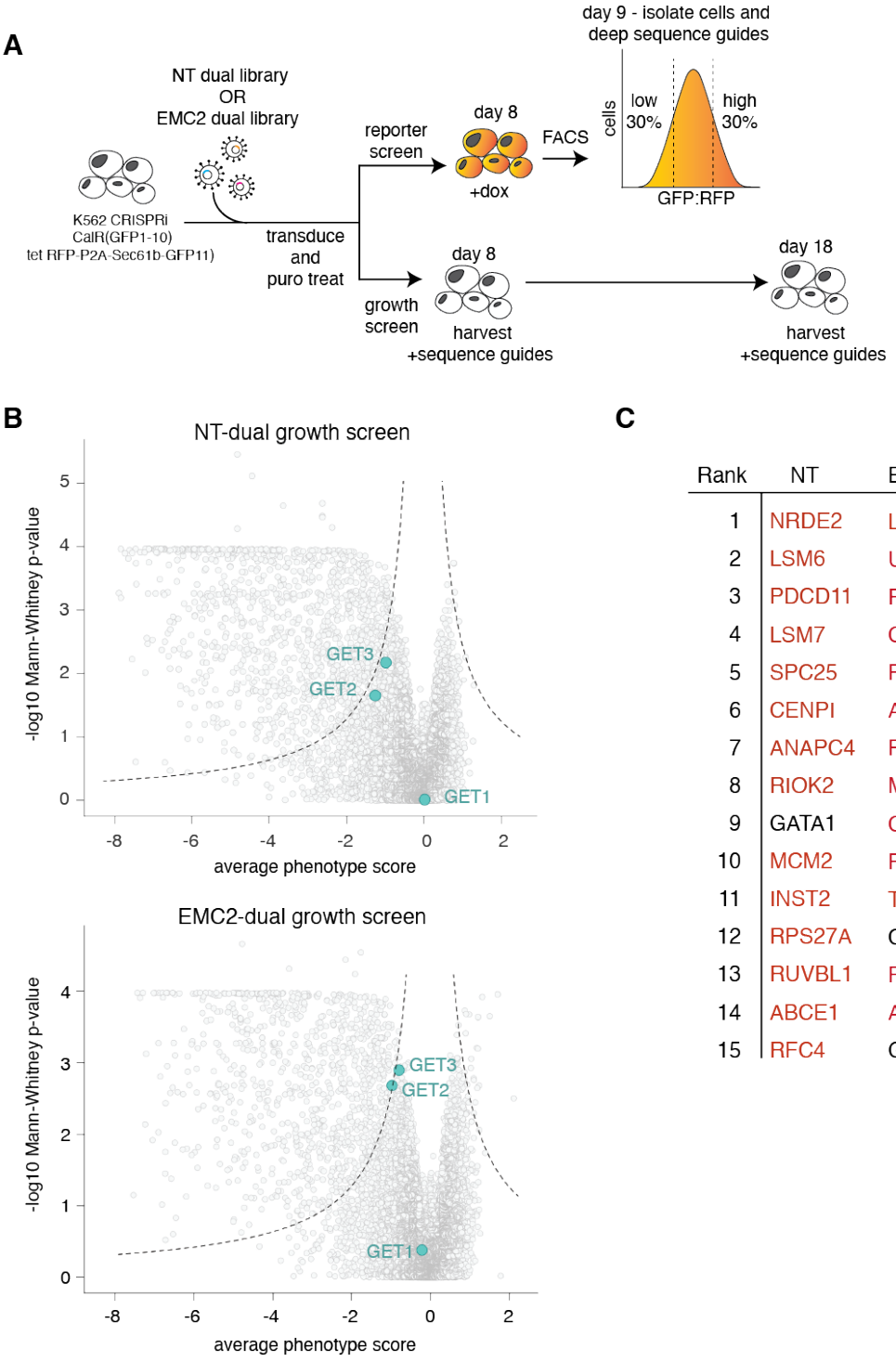



**Figure S2.2. Growth screens with dual-guide libraries. (A)** Schematic and timeline of CRISPRi fluorescent and growth screens with dual-guide libraries. **(B)** Volcano plots of growth screens for the three strongest guide RNAs versus Mann-Whitney p-values from two independent replicates of

growth screens with the indicated libraries. Individual guides are displayed in gray, while core factors of the GET pathway are highlighted in pink. (**C**) Top ranked hits, as measured from discriminant scores, from (B), essential genes are highlighted in red (Tsherniak et al., 2017; Behan et al., 2019).

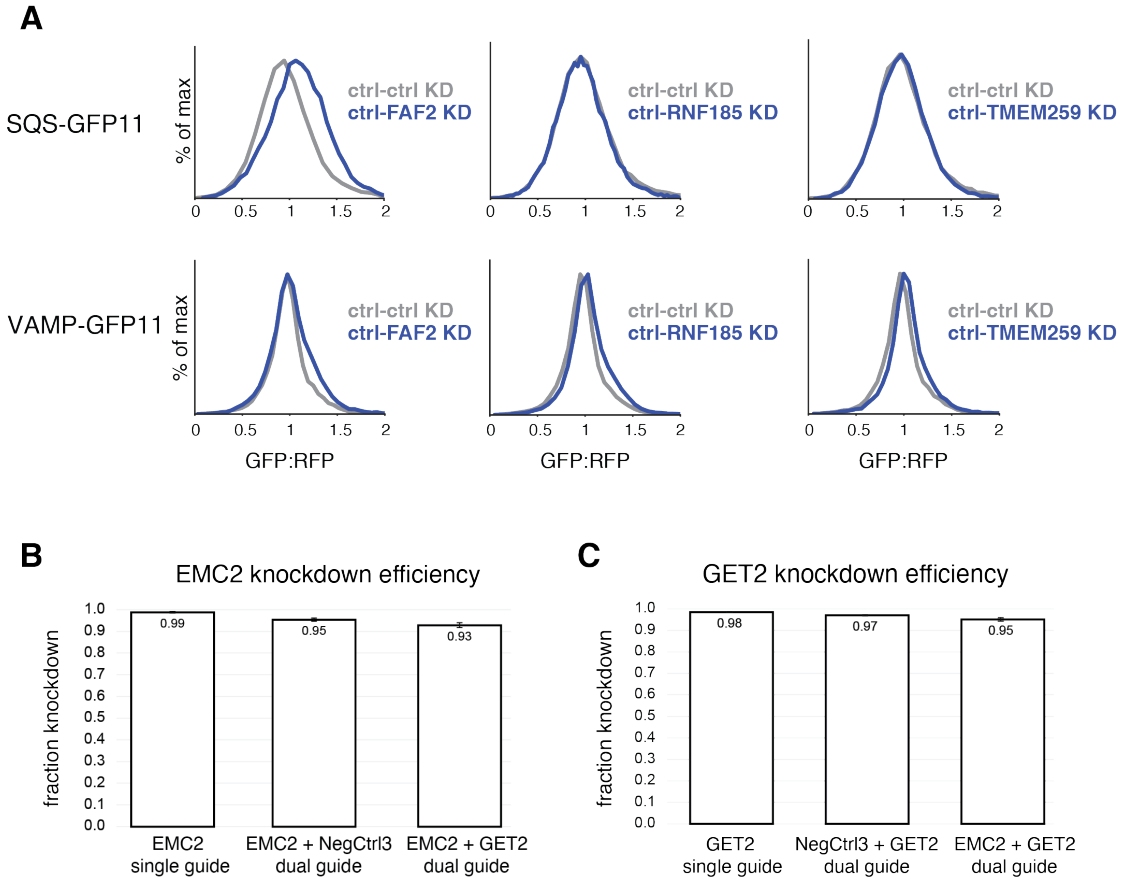



**Figure S2.3**. **Investigating the specificity of putative TA quality control factors and assessing knock-down of dual-guides**. (A) The stability of SQS-GFP11 and VAMP-GFP11 was assayed as in Figure 2.4B. (B) qPCR of EMC2 was assessed in K562 CRISPRi cells expressing the indicated guides to determine whether the context of a dual-guide affected knock-down efficiency (normalized to the housekeeping gene HPRT1). See methods for specific details. (C) As in (B) for the GET2 targeting guide.

## 2.7 Materials and Methods

*Plasmids*

Sequences used for in vivo analysis were derived from UniProtKB/Swiss-Prot and included: squalene synthase isoform 1 (SQS/FDFT1; **Q6IAX1**), vesicle associated membrane protein 2 (VAMP; *P51809-1*), and SEC61β (SEC61B, NP_006799.1). For expression in K562 cells, the transmembrane domain (TMD) and flanking regions of respective ER localized proteins were inserted into a backbone containing a UCOE-EF-1α promoter and a 3′ WPRE element (Addgene #135448) (Jost et al., 2017). The exception was the SEC61β construct used for the CRISPRi screens (RFP-P2A-Sec61b-GFP11) which was integrated into an SFFV-tet3G backbone (Jost et al., 2017). The GFP:RFP reporter system has previously been described (Chitwood et al., 2018) (Guna et al., 2018) and used in the context of CRISPRi screens (Guna et al., 2022). The mCherry variant of RFP was used in all constructs, but is referred to as RFP throughout the text and figures for simplicity. For VAMP2, SQS and SEC61β, directly upstream of the TMD and flanking regions, the first 70 residues of the flexible cytosolic domain of SEC61β was inserted. Downstream, the GFP11 tag (RDHMVLHEYVNAAGIT) was inserted at the C-terminal separated by a 2-4X GS linker to allow for complementation with GFP1-10. In order to express GFP1-10 in the ER lumen, the human calreticulin signal sequence was appended preceding GFP1-10-KDEL as previously described (Cabantous et al., 2005) (Kamiyama et al., 2016) (Inglis et al., 2020).

Programmed dual sgRNA guide vectors were used to allow for the simultaneous depletion of genes (Replogle et al., 2020). Dual guide pairs included: EMC2-Control (GGAGTACGCGTCCGGGCCAA, GACGACTAGTTAGGCGTGTA), Control-GET2 (GACGACTAGTTAGGCGTGTA, GATGTTGGCCGCCGCTGCGA), EMC2-GET2 (GGAGTACGCGTCCGGGCCAA, GATGTTGGCCGCCGCTGCGA), Control-Control (GACGACTAGTTAGGCGTGTA, GACGACTAGTTAGGCGTGTA), Control-GET3 (GACGACTAGTTAGGCGTGTA, GGCTCCAGCGGCTCCACATC), EMC2-GET3 (GGAGTACGCGTCCGGGCCAA, GGCTCCAGCGGCTCCACATC), Control-FAF2 (GACGACTAGTTAGGCGTGTA, GCGGGTCAGGAGCGTAGAGG), Control-RNF185 (GACGACTAGTTAGGCGTGTA, GGCTGGCGTTAACTGTGCGG), Control-TMEM259 (GACGACTAGTTAGGCGTGTA, GCGGACGAGAAAGCGGAAGA). All reporter constructs and programmed dual guides are available upon request.

pCMV-VSV-G was a gift from Bob Weinberg (Addgene plasmid # 8454 ; http://n2t.net/addgene:8454 ; RRID:Addgene_8454).

*CRISPRi dual-guide library construction*

Following selection and verification of a fixed guide, it is cloned into a hU6-CR3 cassette flanked by BamHI/NotI restriction cut sites (pJR152, Addgene 196280). The pJR152 backbone is compatible with standard BstXI/BlpI ligation with annealed oligos. Guide oligos must be ordered with custom overhangs (forward oligo: "ATG"-guide sequence-"GTTTCAGAGC"; reverse oligo: "TTAGCTCTGAAAC" – reverse complement of guide sequence – "CATGTTT"). For the NT and EMC2 libraries, the fixed guides were "GACGACTAGTTAGGCGTGTA" and "GGAGTACGCGTCCGGGCCAA", respectively.

The two components of the dual-guide library are pJR152 containing the fixed guide of interest, and the CRISPRi-v2 library (Addgene Pooled Libraries #83969) (Horlbeck et al., 2016)(https://www.addgene.org/pooled-library/weissman-human-crispri-v2/). Construction of the dual-guide library essentially consists of restriction digesting both elements with BamHI/NotI and inserting the hU6-CR3-fixed guide element into the CRISPRi-v2 library at scale, resulting in an mU6-CR1-hU6-CR3 design previously described (Landisman & Connors, 2005).

Specifically, the pJR152 containing either the NT or EMC2 targeting guide was restriction digested at 37C for 3 hours with BamHI/NotI, and the resulting 400 bp gene fragment (containing the hU6-CR3-fixed guide element) gel purified. Approximately 30 ug of the CRISPRi-v2 top5 library was restriction digested with BamHI/NotI in the presence of shrimp alkaline phosphatase (rSAP) for 6 hours at 37C followed by heat inactivation at 65C for 5 minutes. Smaller amounts of the CRISPRi-v2 library can be digested, but a larger initial reaction will prevent repeat digestions and subsequent quality control checks. Ensure that no more than 10% of the reaction is enzyme to prevent star activity or inactivation of restriction enzymes. The resulting fragment of 8,800 base pairs was gel purified and eluted in a smaller volume. Following recovery of both elements, either NT of EMC2 guide containing inserts were T4 ligated (ensure it is NEB #M0202M) with an insert to vector ratio of 1:2 for a 16 hours at 16C. Various vector:insert molar ratios were tested during piloting, with 1:2 resulting in the highest efficiency. A control ligation containing just the restriction digested CRISPRi-v2 library should be included.

To assess background, a small amount (0.5 ul of 20 ul) of the resulting ligations as well as the control were transformed into 10 ul of Stellar chemically competent cells (Takara #636763) using manufacturer guidelines. Various dilutions were plated ($1/10^{th}$, $1/100^{th}$, and $1/1000^{th}$) with the resulting colonies counted on both control plates and dual library plates. Successful digestion and ligation should result in <2% background colonies, with the concern that single guides may pack much better than dual guides into lenti-virus, and therefore be over-represented.

To permit electrophoresis into MegaX cells at scale (ThermoFisher #C640003) the rest of the dual T4 library ligation for either NT or EMC2 dual libraries is selected with SPRISelect beads (Beckman Coulter B23317) and eluted in 20 ul of water. The entire resulting elution were electroporated into MegaX cells using manufacturer guidelines. Electroporated cells are allowed to recover and set up in an overnight culture of 200 ml LB supplemented with 100 µg/mL carbenicillin for each library. A smaller proportion of the culture was taken ($1/1{,}000^{th}$ and $1/10{,}000^{th}$) and plated to allow for the estimate of resulting colonies and therefore guide coverage, with the expectation of maintaining 50X coverage for the 100,000 element CRISPRi-v2 library. Resulting NT and EMC2 dual libraries were amplified and barcoded by PCR using NEB Next Ultra ii Q5 MM (M0544L) and index primers and a unique reverse primer (CAAGCAGAAGACGGCATACGAGATggaatcatgggaaataggccctc) that binds in the hU6 region upstream of the fixed guide. The standard CRISPRi-v2 library was amplified in parallel to allow for the assessment of guide representation in dual libraries. (Horlbeck et al., 2016) SPRISelect beads (Beckman Coulter B23317) were used to purify the dual DNA libraries (349 bp), and purified DNA was sequenced using an Illumina HiSeq2500 with the same sequencing primer as the standard CRISPRi-v2 library (GTGTGTTTTGAGACTATAAGTATCCCTTGGAGAACCACCTTGTTG). The NT and EMC2 dual libraries are available on Addgene (Library 197348 and Library 197349, respectively).

*Cell culture and cell line construction*

| **Cells** | **Source** |
|---|---|
| K562 KRAB-BFP-dCAS9 | Gilbert et al., 2014 |
| HEK 293T/17 | ATCC CRL-11268 |
| K562 KRAB-BFP-dCAS9 ER GFP1-10 | Guna et al., 2022 |
| K562-CRISPRi-Tet-ON-((ER)-GFP1-10)-(tet-RFP-P2A-SEC61β-GFP11) | This study, available upon request |

K562 cells expressing KRAB-BFP-dCas9 (Gilbert et al., 2014) were cultured in RPMI-1640 with 25 mM HEPES, 2.0 g/L $NaHCO_3$, and 0.3 g/L L- glutamine supplemented with 10% FBS (or Tet System Approved FBS), 2 mM glutamine, 100 units/mL penicillin, and 100 μg/mL streptomycin. Cells were maintained between $0.25 \times 10^6$ –$1 \times 10^6$ cells/mL. HEK293T/17 (ATCC CRL-11268) cells were cultured in DMEM supplemented with 100 units/mL penicillin and 100 μg/mL streptomycin. K562 and HEK293T cells were grown at 37C.

Cell lines expressing GFP1-10 in the ER lumen were generated as previously described (Inglis et al., 2020; Guna et al., 2022). CRISPRi K562 cells were infected with lenti virus containing CalR(GFP1-10)-KDEL and sorted with a Sony Cell Sorter (SH800S) as single clones into 96-well plates. Clones were expanded and confirmed by complementation with a construct targeted to the ER appended to GFP11. To generate the SEC61β line used for screening line, lentivirus containing ER(GFP1-10) and RFP-P2A-SEC61β-GFP11 under an inducible promoter were co-infected at one copy per cell line in CRISPRi (expressing KRAB-BFP-dCas9) K562 Tet-ON cells (Gilbert et al., 2014). Cells were then single cell sorted, verified by induction with doxycycline (100 ng/ul), and confirmed to localize to the ER by microscopy. These cells are referred to as K562-CRISPRi-Tet-ON-((ER)-GFP1-10)-(tet-RFP-P2A-SEC61β-GFP11).

*Lentivirus production*

Lentivirus was generated using standard protocols. Briefly, HEK293T cells were co-transfected with two packaging plasmids (pCMV-VSV-G and delta8.9, Addgene #8454) and either a desired transfer plasmid, or the dual libraries, using Transit-IT-293 transfection reagent (Mirus) (Stewart et al., 2003). Approximately 48 hours after transfection, the supernatant was collected and flash frozen. Virus was rapidly thawed at 37C prior to transfection.

*Flow cytometry reporter CRISPRi screens*

CRISPRi screens were performed as previously described, with minor modifications (Gilbert et al., 2014; Horlbeck et al., 2016). Either the NT or EMC2 dual libraries were transduced in duplicate into 330 million K562-CRISPRi-Tet-ON-((ER)-GFP1-10)-(tet-RFP-P2A-Sec61b-GFP11) cells at a multiplicity of infection less than one. Throughout the screen, cells were maintained in 1L spinner flasks (Bellco, SKU: 1965-61010) at a volume of 1L. 48 hours after transfection, BFP positive cells were between 30-35%. At this point, cells began treatment with 1 μg/mL puromycin for three days to select for guide positive cells. Cells were given two days to recover after puromycin selection

and the reporter was induced with doxycycline (100 ng/mL) for 24 hours and sorted on a FACSAria Fusion Cell Sorter. Cells were daily diluted to $0.5 \times 10^6$ cells/mL to ensure that the culture was maintained at an average coverage of more than 1000 per sgRNA.

During sorting, cells were gated for BFP (to select only guide-positive cells) and RFP and GFP (indicating an expressing reporter). Cells were sorted based on the GFP:RFP ratio of the final gated population, and roughly 40 million cells with either the highest or lowest 30% GFP:RFP ratios were collected, pelleted, and flash-frozen. Genomic DNA of the cell pellets was extracted and purified using a Nucleospin Blood XL kit (Takara Bio, #740950.10). Guides were amplified and barcoded by PCR using NEB Next Ultra ii Q5 MM (M0544L) and index primers and a unique reverse primer (CAAGCAGAAGACGGCATACGAGATggaatcatgggaaataggccctc) that binds in the hU6 region upstream of the fixed guide. This ensures that only DNA containing both the v2 library and one of the fixed EMC2 or NT guides is amplified and sequenced. SPRISelect beads (Beckman Coulter B23317) were used to purify the DNA library (349 bp), and purified DNA was analyzed on an Agilent 2100 Bioanalyzer prior to sequencing using an Illumina HiSeq2500 using the same sequencing primer as the standard CRISPRi-v2 library (GTGTGTTTTGAGACTATAAGTATCCCTTGGAGAACCACCTTGTTG). Post-sequencing analysis was performed using the pipeline in https://github.com/mhorlbeck/ScreenProcessing (Horlbeck et al., 2016). Guides with fewer than 50 counts were excluded to ensure proper coverage. For each screen, the strongest 3 sgRNA phenotypes were used to calculate the phenotype score of each gene. The Mann-Whitney p-value was calculated using all 5 sgRNAs targeting the same gene compared to negative controls (Supplementary Table 2). Since screens were performed in biological duplicate, the sgRNA phenotypes were averaged. Discriminant scores were calculated as the product of the gene's phenotype score and the Mann-Whitney p-value. Discriminant ranks for each screen were determined by ranking the list of genes from lowest to highest discriminant values, with the lowest score the highest rank.

*CRISPRi growth screens*

To perform the growth screen, the same cells for flow cytometry screens infected with either NT or EMC2 dual libraries were harvested after recovery from puromycin selection as Day 0, and then again after 10 doublings on Day 18. 50 million cells from each biological duplicate and each library were harvested. Cells were maintained at an average coverage of more than 1000 per sgRNA during all points of the growth screen, and >99% BFP positive cells were confirmed at the time of

harvesting. Resulting libraries were extracted, amplified, purified and sequenced identically as for the flow-cytometry based screen samples as described above (Supplementary Table 3).

*Flow cytometry*

For all reporter assays, K562 CRISPRi cells containing ER(GFP1-10) were spinfected with lentivirus of indicated guides and knock down was allowed for 6 days. Cells were then spinfected with lentivirus containing the indicated reporters and analyzed by flow cytometry after 48-72 hours. All reporter experiments were performed in biological triplicate. All samples were either run on an NXT Flow Cytometer (ThermoFisher) or a MACSQuant VYB (Miltenyi Biotec). Flow cytometry data was analyzed either in FlowJo v10.8 Software (BD Life Sciences) or Python using the FlowCytometryTools package.

*Quantitative PCR*

Quantitative PCR was used to analyze RNA levels after knockdown with dual or single guides. K562 cells expressing the CRISPRi machinery were infected with guides and after 8 days of knockdown, RNA was extracted and treated with DNaseI using a Direct-zol RNA MiniPrep Plus kit (R2072, Zymo). Purified RNA was reverse transcribed using the SuperScript III First-Strand Synthesis SuperMix for qRT-PCR kit (11752050, Invitrogen). Reactions were run on a StepOnePlus Real-Time PCR system and knockdown efficiency was calculated using the housekeeping gene HPRT1. Samples were collected and analyzed in triplicate with the means and standard deviations plotted. The primers used were: EMC2 (fwd AGACAGTTCCCTGGCAGTCAC, rev TCCACATTTTTCCCCTGGGCT); GET2 (fwd CCGGATCATGGGCTTTCACA, rev CCTGCTGGTCAGTTGTTCCT).

**2.8 Availability of Data and Materials**

All materials necessary for dual-guide construction are available via Addgene, with accession numbers listed in the Materials and Methods. All programmed sgRNA, reporter constructs and cell lines are available from the corresponding author upon request. The sequencing datasets generated during the current study are available in the Caltech DATA repository and can be publicly accessed, https://doi.org/10.22002/3hvyj-yzq30

## 2.9 Acknowledgements

We thank K. Hickey, R. Saunders, K. Popova and A. Inglis for helpful discussions. We thank the Whitehead Institute Flow Cytometry Core access to FACS machines and flow cytometers, and the Millard and Muriel Jacobs Genetics and Genomics Laboratory at Caltech for sequencing of screening libraries.

*C h a p t e r 3*

# ROLE OF A HOLO-INSERTASE COMPLEX IN THE BIOGENESIS OF BIOPHYSICALLY DIVERSE ER MEMBRANE PROTEINS

The following chapter is adapted from a manuscript in revision, Page, Nguyen, Pleiner et al., 2023 and modified according to the Caltech Thesis format.

**Page, K.R.*,** Nguyen, V.,*, Pleiner, T.*, Tomaleri, G.P., Wang, M.L., Guna, A., Wang, T., Chou, T., and Voorhees, R.M. (2023) "Role of a holo-insertase complex in the biogenesis of biophysically diverse ER membrane proteins."

## 3.1 Summary

Mammalian membrane proteins perform essential physiologic functions that rely on their accurate insertion and folding at the endoplasmic reticulum (ER). Using forward and arrayed genetic screens, we systematically studied the biogenesis of a panel of membrane proteins, including several G-protein coupled receptors (GPCRs). We observed a central role for the insertase, the ER membrane protein complex (EMC), and developed a dual-guide approach to identify genetic modifiers of the EMC. We found that the back of sec61 (BOS) complex, a component of the 'multipass translocon,' was a physical and genetic interactor of the EMC. Functional and structural analysis of the EMC•BOS holocomplex showed that characteristics of a GPCR's soluble domain determine its biogenesis pathway. In contrast to prevailing models, no single insertase handles all substrates. We instead propose a unifying model for coordination between the EMC, multipass translocon, and Sec61 for biogenesis of diverse membrane proteins in human cells.

## 3.2 Introduction

Integral membrane proteins are essential across all biological systems, including in mammalian cells and their pathogens. Human membrane proteins mediate a range of processes from cell-to-cell signaling to metabolite transport (von Heijne, 2007). Similarly, many viruses encode membrane proteins that are critical for fusion with a host cell, organization of the replication machinery, and transport of ions and small molecules (by viroporins) that enhance infectivity and morbidity (Harrison, 2008; Lenard, 2008).

In order to carry-out these functions, both the transmembrane (TMDs) and soluble domains have evolved diverse charge, hydrophobicity, and length (von Heijne, 2007). The accurate insertion and folding of these topologically and biophysically diverse proteins therefore represents a major challenge in human cells. Despite the importance of this process, how cells regulate biogenesis of the full complexity of the mammalian and viral membrane proteome is not understood.

The majority of membrane proteins, destined for either the plasma membrane or secretory system, begin their biogenesis at the endoplasmic reticulum (ER) (Rapoport et al., 1996; Shao & Hegde, 2011). For multipass proteins, a nascent polypeptide is captured in the cytosol by the signal recognition particle (SRP) for delivery of the ribosome nascent chain complex to the membrane (Halic & Beckmann, 2005; Shan & Walter, 2005). Once at the ER, TMDs must be inserted into the lipid bilayer. Insertion requires two simultaneous processes: (i) transfer of the hydrophobic TMD from the aqueous cytosol to a membrane-spanning topology within the lipid bilayer, and (ii) translocation of an associated soluble domain across the membrane into the ER lumen. The latter of which is energetically unfavorable and therefore typically catalyzed by a membrane protein insertase in cells (Guna, Hazu, et al., 2023).

The textbook model posits that the Sec61 translocation channel is the major insertase for multipass membrane proteins. It was hypothesized that its unique clam-shell architecture could accommodate all aspects of membrane protein biogenesis: axial opening creates a pore in the membrane for translocation into the ER lumen, while lateral opening would permit partitioning of a TMD into the bilayer (Van den Berg et al., 2004; Voorhees & Hegde, 2016). Studies of signal sequences suggested that opening of Sec61 is triggered by binding of a sufficiently hydrophobic segment of the nascent chain at the lateral gate of the channel (Rapoport et al., 2017). However, many multipass membrane proteins contain poorly hydrophobic TMDs, which while required for function, cannot autonomously gate Sec61 (Enquist et al., 2009; Schorr et al., 2020). Therefore, a simple model in which each TMD is sequentially inserted into the lipid bilayer by Sec61 alone cannot explain the insertion or folding of most multipass membrane proteins.

Recently, it has instead been proposed that substrates, including the physiologically essential family of G-protein coupled receptors (GPCRs), are inserted by a ‘multipass translocon’ that uses Sec61 as a structural scaffold, but does not rely on its insertase activity (McGilvray et al., 2020; Smalinskaitė et al., 2022; Sundaram et al., 2022). The multipass translocon is a dynamic, hetero-oligomeric 8-subunit complex that includes the GET and EMC like (GEL), back of Sec61 (BOS), and PAT complexes (Figure S2.1A). The GEL complex—composed of the Oxa1 superfamily insertase, TMCO1 and its binding

partner, OPTI—serves as the dedicated insertase of the multipass translocon, and is postulated to integrate nascent TMDs as they emerge from the ribosome behind Sec61. The PAT complex, containing Asterix and CCDC47, has two proposed roles: Asterix chaperones hydrophilic TMDs within the lipid bilayer during synthesis of a multipass protein (Chitwood & Hegde, 2020; Meacock et al., 2002); while CCDC47 directly engages the ribosome and closes the lateral gate of Sec61, helping to guide TMDs to the multipass translocon (Smalinskaitė et al., 2022; Sundaram et al., 2022). Finally, the function of the BOS complex remains unknown, but is thought to act as scaffold for recruitment of the remaining multipass components (Smalinskaitė et al., 2022; Sundaram et al., 2022). Together, these factors create a protected lipid cavity behind Sec61 thought to facilitate multipass membrane protein insertion and folding. In support of this model, it was shown that biogenesis of the GPCR rhodopsin is unaffected by inhibitors that prevent access to the lateral gate of Sec61, and nascent TMDs crosslink to multipass translocon components as they emerge from the ribosome. However, whether the GEL complex, which is metazoan specific and not essential in humans (Karczewski et al., 2020), is responsible for the integration of all multipass TMDs remains unclear.

Indeed, earlier work established that an additional insertase, the ER membrane protein complex (EMC), was required for biogenesis of many multipass membrane proteins. In mammals, the EMC is an abundant, nine-subunit complex that functions as both an insertase and chaperone (Z. Chen et al., 2023; Guna et al., 2018; Shurtleff et al., 2018; Tian et al., 2019). In addition to post-translational insertion of a subset of tail anchored proteins, the EMC also co-translationally inserts the first TMD of many GPCRs and other multipass membrane proteins that position their N-terminus in the ER lumen or extracellular environment (i.e., adopt an $N_{exo}$ topology) (Chitwood et al., 2018). Indeed, expression of rhodopsin, which does not rely on the lateral gate of Sec61 for insertion, is EMC dependent (Satoh et al., 2015). However, the function of all nine of EMC's subunits, in particular those that form its large lumenal domain that is not directly involved in insertion, is not known.

Structures of the yeast and human EMC show that substrate TMDs are inserted into the bilayer via a positively charged hydrophilic groove, through which the substrate's soluble N-terminus must also translocate (Bai et al., 2020; Miller-Vedam et al., 2020; O'Donnell et al., 2020; Pleiner et al., 2020). The positioning of positively charged residues within the membrane is a conserved feature of the Oxa1 superfamily of insertases and is required for their activity (Borowska et al., 2015; Kumazaki et al., 2014; McDowell et al., 2020; Pleiner et al., 2020, 2023). In contrast to the prevailing model, it is likely that multipass substrates are therefore directly delivered by SRP to the EMC (Wu & Hegde, 2023), leaving the EMC to act upstream of Sec61 and the multipass translocon.

However, this model leaves several central unanswered questions for how human and viral membrane proteins are accommodated by the biogenesis and quality control machinery in the ER. First, whether or how the EMC coordinates with the multipass translocon during multipass biogenesis is not known. Second, if the EMC is responsible for insertion of the first $N_{exo}$ TMD of many membrane proteins (including GPCRS), how substrates are transferred between the EMC, Sec61, and the multipass translocon is not clear. Finally, a systematic analysis of the substrate specificity and cooperation of the suite of biogenesis factors in the ER to ensure insertion and folding of their diverse clients, has not been explored.

## 3.3 Results

### Systematic analysis of membrane protein biogenesis

With the goal of unbiasedly identifying factors required for biogenesis of diverse membrane proteins, we selected a panel of substrates with distinct topologies, biophysical properties, and number of TMDs (Figure 3.1 and S3.1B). In this initial panel we included the co-translationally targeted multipass proteins AGTR2 and the viral ORF3a and M from SARS-CoV-2, as well as the post-translationally targeted single spanning tail-anchored protein, Sec61b. AGTR2 is a seven-TMD G-protein coupled receptor (GPCR; Figure 3.1A). GPCRs represent the largest single family of membrane proteins encoded by the human genome and are responsible for physiologically important signaling processes throughout the human body (Heldin et al., 2016; O'Hayre et al., 2014). AGTR2, like most GPCRs, contains TMDs with varying hydrophobicities, including those predicted to insert autonomously and those likely to only insert in the context of the intact multispanning protein (Cvicek et al., 2016). Outside of the TMDs, AGTR2 contains small cytosolic and extracellular loops, and a neutrally charged N-terminal soluble domain that must be translocated across the ER membrane (i.e. adopts an $N_{exo}$ topology) during biogenesis.

Second, we chose two closely related viral proteins ORF3a and M from SARS-CoV-2 (Figure 3.1C,D). These multipass membrane proteins both adopt an identical three-TMD topology but their TMDs have distinct biophysical properties and insertion propensities (Dolan et al., 2022; Kern et al., 2021). Further, ORF3a and M have soluble N-termini of different length (42 vs 19 amino acids) and charge (0 vs -2), which we hypothesized could alter the suite of host factors required for their

biogenesis. The ability to query two topologically related proteins that are also innocuous upon overexpression was a unique advantage of using viral substrates.

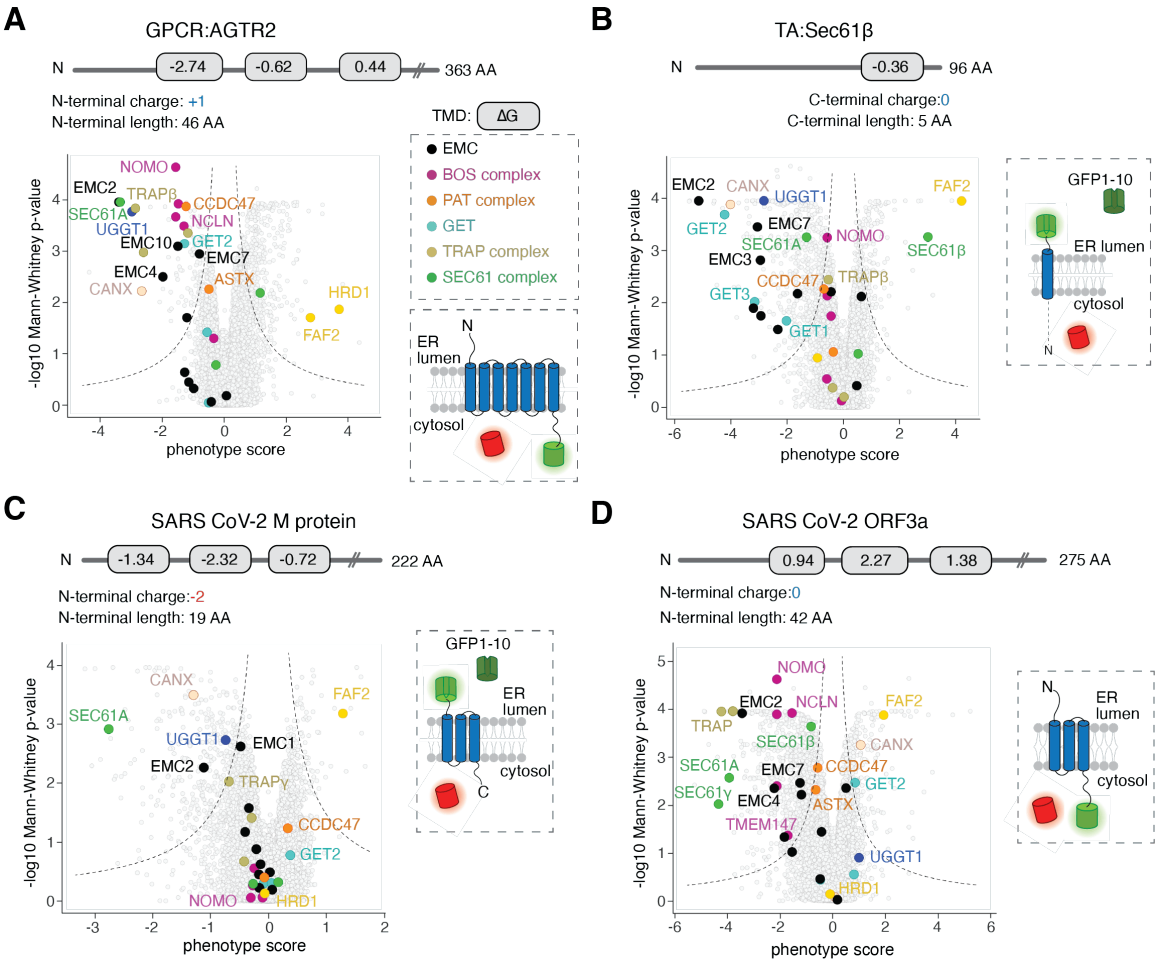



**Figure 3.1. Genome-wide CRISPRi screens to systematically query biogenesis factors for diverse ER substrates. (A)** (Top) Schematic of the GPCR reporter AGTR2. The gray rectangle indicates a transmembrane domain (TMD) with its corresponding ΔG value, calculated using the ΔG prediction server v1.0 (Hessa, Meindl-Beinker, et al., 2007). (Right) Topology of the AGTR2 fluorescent reporter used in the genome-wide CRISPR interference (CRISPRi) screen. Full-length GFP is fused to AGTR2 and RFP is expressed as a normalization marker from the same open reading frame, separated by a P2A sequence. (Bottom) Volcano plot of the GFP:RFP phenotype [log2(High GFP:RFP/low GFP:RFP)] for the AGTR2 reporter with the three strongest sgRNAs plotted with Mann-Whitney p-values from two independent replicates. Individual genes are displayed in grey, and specific factors that increase or decrease AGTR2 stability are highlighted

and labelled. Genes that fall outside the indicated dashed lines represent statistically significant hits. **(B)** As in (A), for the tail-anchored protein Sec61β (Guna, Page, et al., 2023). Here, The GFP11 sequence is appended to the C-terminal of Sec61β. Upon TA insertion into the ER, GFP11 complements with the GFP1-10 independently localized to the ER lumen, resulting in fluorescence. **(C)** As in (A) for the SARS-CoV-2 M protein. The GFP11 sequence is appended to the N-terminal of SARS-CoV-2 M with GFP1-10 expressed in the ER lumen. **(D)** As in (A) for the SARS-CoV-2 ORF3a protein. SARS-CoV-2 ORF3a is appended to full-length GFP. Abbreviations: AA, amino acids.

Finally, as a post-translational control we included the single spanning tail anchored (TA) protein Sec61b (Figure 3.1B). Tail anchored proteins contain a single TMD within ~35 amino acids of their C-terminus, and thus cannot access the co-translational biogenesis pathways typically utilized by multipass proteins (Guna, Hazu, et al., 2023; Kutay et al., 1993). The targeting and insertion of Sec61b has been extensively characterized biochemically, and therefore serves as a comparison for machinery required for biogenesis of multipass vs singlepass membrane proteins (Guna et al., 2018; Mateja et al., 2009; Stefanovic & Hegde, 2007).

Individual cell lines stably expressing these four substrates were generated in which the GFP-tagged membrane protein was expressed along with a translation normalization marker (RFP) in human K562 cells expressing the CRISPR inhibition (CRISPRi) machinery (Figure S3.1C) (Gilbert et al., 2014; Horlbeck et al., 2016). Previous experiments have established that when factors required for membrane protein targeting, insertion, or folding are depleted, substrates are robustly recognized and degraded by the ubiquitin proteasome pathway leading to a decrease in GFP fluorescence (Guna et al., 2018; Pleiner et al., 2020). Conversely, if membrane protein quality control is disrupted, substrates accumulate beyond normal levels, leading to an increase in GFP fluorescence. Therefore, following transduction with a genome-wide sgRNA library, cells that displayed altered substrate levels (i.e., GFP fluorescence) relative to the normalization control were sorted using florescence activated cell sorting (FACS; Figure S3.1C). Deep sequencing of the sgRNAs enriched in both the low and high GFP-fluorescing cells was used to identify putative biogenesis and quality control factors, respectively.

In addition to specific factors related to the unique physiologic function of each substrate (e.g., lysosomal and vesicular trafficking factors for ORF3a; Figure S3.1D), we identified a panel of

factors that differentially affect the biogenesis or degradation of these four model substrates (Figure 3.1). For example, amongst the identified quality control factors, loss of the ER-localized E3 ubiquitin ligase HRD1 was found to stabilize only AGTR2, while the more general ERAD component, FAF2 was identified in all four screens (Kikkert et al., 2004; Olzmann et al., 2013). On the biogenesis side, the GET pathway components had the most pronounced effect on the tail anchored control (Sec61b), consistent with its role in post-translational insertion (Mariappan et al., 2011; Schuldiner et al., 2005, 2008; Vilardi et al., 2011). Conversely, both the translocon associated protein (TRAP) complex and the members of the multipass translocon were only significant hits for biogenesis of AGTR2 and ORF3a, but not Sec61b or M. Interestingly however, the central insertase of the multipass translocon, the GEL complex, was not a significant hit in any screen, despite near-complete depletion under these conditions (Figure S3.1E). The only universally identified biogenesis factor were components of the EMC, which is consistent with its established role in TA and $N_{exo}$ TMD insertion (Chitwood et al., 2018; Guna et al., 2018).

**Distinct pathways for biogenesis and quality control of diverse substrates**

To better delineate how the factors identified in the screens affected the biogenesis of a broader range of substrates, we generated a panel of 13 membrane protein reporters that represented a range of topologies and biophysical properties (Figure 3.2A). We included substrates with varying numbers of TMDs of distinct lengths and hydrophobicity, as well as those that differ in the structure of the intervening soluble domains. In the panel were multipass proteins in which the N-terminus must be translocated across the ER membrane (i.e., $N_{exo}$ topology: several GPCRs, ORF3a, and M); multipass proteins in which the N-terminus will remain in the cytosol ($N_{cyt}$ topology: TRAM2, EAAT1, GET2, and YIPF1); and single spanning (Type II: ASGR1; Type I: TRAPα); and TA proteins (SQS, VAMP2, and Sec61b). In order to allow for direct comparison, all reporters contained a full length GFP with the exception of Sec61b, whose targeting is affected by fusion with a florescent protein and therefore required use of the split GFP approach (Figure 3.1A) (Guna et al., 2022; Guna, Page, et al., 2023; Inglis et al., 2020).

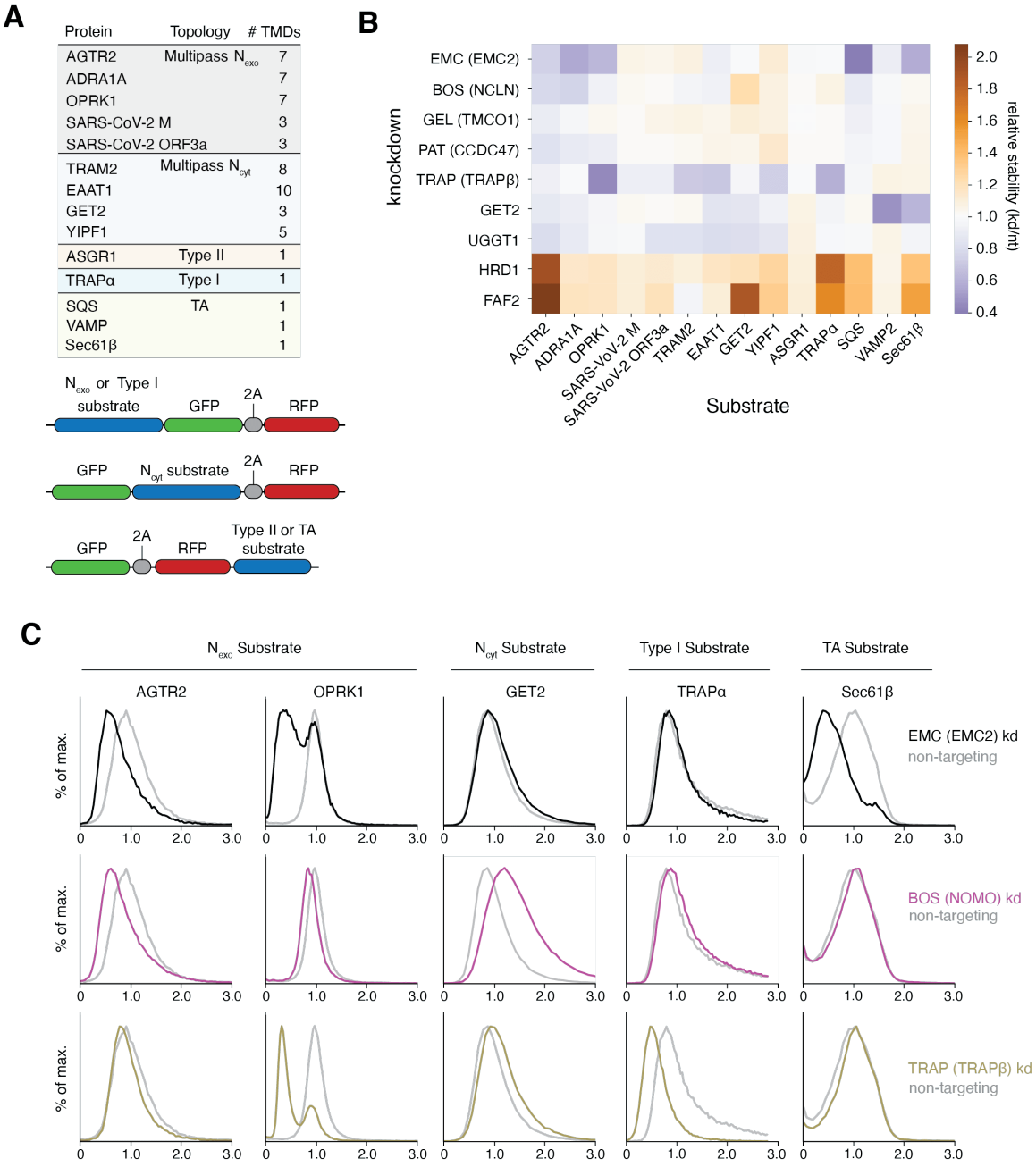



**Figure 3.2. Differential effects of ER factors on membrane protein biogenesis. (A)** (Top) Topology and number of TMDs for the panel of membrane protein reporters used to assess the effects of hits from genome-wide screens in an arrayed format. (Below) Schematic of fluorescent reporters. **(B)** Factors from forward CRISPRi screens in K562 cells (Figure 3.1) were systematically knocked down in RPE1 cells using CRISPRi and expression of reporters from (A) were assessed using flow cytometry. Results are shown in a heatmap indicating relative stability of

each reporter after gene knockdown compared to a non-targeting control. **(C)** Representative flow cytometry analysis of individual data points from the arrayed screen summarized in (B).

Using an arrayed screen, we tested the depletion of nine factors that represented each of the major biogenesis (EMC, BOS, GEL, PAT, TRAP, GET, UGGT1) and quality control (HRD1, FAF2) complexes identified in the screens (Figure 3.2B). Critically, in most cases knockdown of a single subunit is sufficient to deplete the entire complex (Figure S3.1E) (Colombo et al., 2016; Dettmer et al., 2010; Pleiner et al., 2021; Volkmar et al., 2019). For these experiments, we used the near-diploid human RPE1 cell-line because we postulated that redundancy and compensation between factors would be more pronounced in an aneuploid cell line. Comparison across substrates both validates the hits identified in the four genome-wide screens and suggests categories of dependency that correlate with substrate topology. As expected, the clearest delineation between TAs and other membrane protein substrates is dependence on the post-translational pathway for targeting and insertion into the ER membrane. Consistent with earlier biochemical data, we find that both our forward and arrayed screens show that TA biogenesis depends on the hydrophobicity of their TMDs (Guna & Hegde, 2018; Rao et al., 2016; Shao et al., 2017; F. Wang et al., 2010). TAs with sufficiently hydrophobic TMDs (VAMP2) rely on the GET pathway, low hydrophobicity TAs (SQS) rely on the EMC pathway, while intermediate hydrophobicity TAs (Sec61b) can utilize both (Guna, Page, et al., 2023). Unexpectedly, loss of the GET pathway insertase GET1/2 also appeared to have a small effect on biogenesis of several multipass substrates (e.g., the GPCR AGTR2 and EAAT1) in both the genome-wide and arrayed screens.

In contrast, several factors appear to be specific to co-translationally targeted substrates. For example, depletion of the Sec61 associated chaperone, TRAP affects multipass but no TA substrates (Figure 3.2C) (Gemmer et al., 2023; Shao & Hegde, 2011). Though an effect of TRAPb depletion was observed for the single spanning protein TRAPa, this may be due to an assembly rather than a biogenesis defect. While components of the multipass translocon—including the BOS, PAT, and GEL complexes—are required for substrates with multiple TMDs, they are not required for any of the single spanning membrane proteins. However, our data suggest that the function of the multipass translocon differs across cell-types, because GEL complex dependence for AGTR2 was only observed in RPE1 cells (Figure 3.2B) but not the K562s used in the screens, despite efficient knockdown (Figure S3.1E). It is possible this reflects cell-type specific changes in expression and partial redundancy and/or compensation of biogenesis factors in the ER. For

example, we consistently observe that depletion of the EMC leads to a compensatory increase in TMCO1 levels (Figure S3.1E-F).

Critically however, even in this relatively small panel, it is clear that these multipass-specific factors are not required for the biogenesis of all multispanning proteins. For example, we observe variability in dependence on the BOS, GEL, and PAT complex amongst the three related GPCRs tested in the arrayed screen. These data suggest that biophysical properties of the TMDs and surrounding regions are more important than topology in determining biogenesis pathway. This observation sets the stage for an in-depth study of the relationship between substrate properties and biogenesis requirements.

**Identification of genetic modifiers of the EMC genome wide**

One commonality across many substrates in our arrayed and genome wide screens was a dependence on the EMC for biogenesis. We therefore wondered how the EMC cooperates with other factors in the ER for insertion and folding of its multipass substrates like GPCRs. Indeed, immunoprecipitation of the human EMC from native membranes suggests it associates with a myriad of ER-resident chaperones (e.g., CNX) (Bergeron et al., 1994), biogenesis machinery (e.g., SRP receptor, glycosylation machinery, and components of the multipass translocon) (Kelleher & Gilmore, 2006; McGilvray et al., 2020; Shan & Walter, 2005), and quality control factors (e.g., the ATPase p97, responsible for extraction from the ER) (Meyer et al., 2012) (Figure 3.3A). Recruitment of factors required for the folding and surveillance of nascent proteins to the EMC would ensure that clients are immediately captured for maturation or degradation upon integration into the ER. These results establish the EMC as a central organizing factor for membrane protein biogenesis and quality control within the ER membrane.

To delineate which of these associated factors are phenotypically important, we developed a dual-guide CRISPRi approach to systematically identify genetic modifiers of the EMC genome-wide (Figure 3.3B) (Guna, Page, et al., 2023). Briefly, we generated a library that expresses two sgRNAs on a single plasmid: (1) a genetic anchor guide, targeting the core subunit, EMC2, which when depleted results in loss of the remaining EMC subunits (Pleiner et al., 2021; Volkmar et al., 2019b); and (2) a second randomized guide, targeting all open reading frames genome wide using the existing CRISPRi-v2 library (Horlbeck et al., 2016). Transduction of this dual library allows the

acute knockdown of both the genetic anchor and a second randomized gene simultaneously in each cell, and is compatible with a standard CRISPRi FACS-based screening and analysis pipeline.

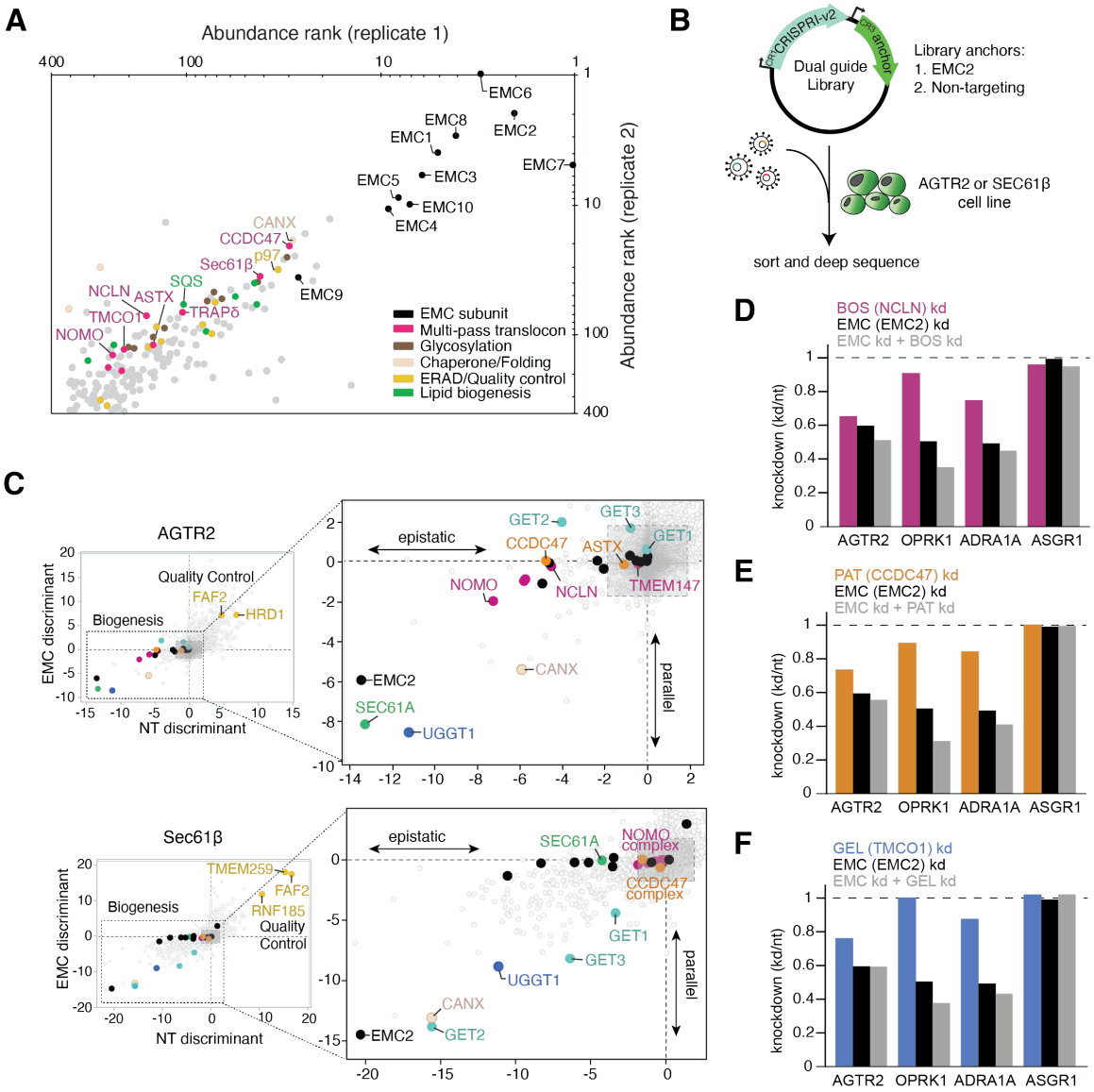



**Figure 3.3. Components of the multipass translocon are epistatic with the EMC. (A)** Scatter log-plot of abundance rank from co-immunoprecipitation of the EMC, expressed at endogenous levels, determined by mass spectrometry from two replicates. **(B)** Schematic of dual guide libraries used for genome-wide EMC genetic modifier screens in AGTR2 and Sec61β reporter cell lines. The CRISPRi-v2 sgRNA library and the genetic anchor guide, targeting either EMC2 or a non-targeting (NT) control, are expressed from a single vector (Guna, Page, et al., 2023). **(C)** Comparison of EMC2- and NT-dual library CRISPRi screens using the discriminant score as a

metric (encapsulating phenotype and -$\log_{10}$ Mann-Whitney p-value for each gene) for AGTR2 (above) and Sec61β (below). Biogenesis factors are boxed and displayed at right in greater detail. Factors off-diagonal are either epistatic or parallel to the EMC, as indicated in the plots, while factors along the diagonal are independent of the EMC. **(D)** Analysis of GPCR stability (AGTR2, OPRK1, and ADRA1A) using the GFP:RFP reporter system compared to a control protein (ASGR1) as analyzed by flow cytometry after knockdown with indicated guides.

Comparison of the hits identified in the EMC genetic anchor screen with those from a control screen performed with a non-targeting 'anchor' guide library results in three categories of factors (Figure 3.3C and S3.2). First, are those that have diminished phenotypes when combined with EMC depletion, indicative of an epistatic relationship and potentially a shared pathway with the EMC. Second, are factors that have enhanced phenotypes upon loss of the EMC, likely including factors that represent parallel or partially redundant pathways. Third, are factors that act independently of the EMC, and therefore show no change in phenotype with or without the EMC.

Interrogating both the EMC-dependent tail-anchored protein Sec61b and the GPCR AGTR2 allowed us to delineate EMC co-factors that function to support its post vs co-translational biogenesis roles (Figure 3.3C and S3.2). Validating this approach, all EMC subunits have a diminished effect in the EMC2 knockdown background for both TA and GPCR biogenesis. Conversely, the phenotype of known parallel pathways for TA insertion, including the GET components, are enhanced by EMC depletion, particularly in the TA screen (Figure S3.2C,D). Finally, several quality control factors, such as HRD1 and FAF2 exhibit EMC independent effects, suggesting their function may be agnostic to the insertion pathway.

Direct comparison of the biogenesis factors identified in the TA vs the GPCR genetic modifier screens suggest that many more factors are cooperating with the EMC in insertion and folding of multipass membrane proteins than of TAs. In particular, we identified many components of the multipass translocon, including subunits of the BOS and PAT complexes as epistatic with the EMC for AGTR2 biogenesis (Figure 3.3C-F and S3.2A,B). To test whether other GPCRs display a similar epistatic dependence, we performed an arrayed screen with dual guides targeting the BOS, PAT, or GEL complexes alone or in combination with a guide targeting the EMC. We included the GPCRs AGTR2, OPRK1 and ADRA1A and the type II membrane protein ASGR1 (Figure 3.3D-F). For the GPCRs but not for ASGR1, EMC displays an epistatic relationship with the BOS, PAT,

and GEL complexes, suggesting that this relationship may be a general feature of multipass membrane protein biogenesis. The BOS, PAT, and GEL complexes were further found to co-purify with the EMC under conditions where all components were expressed at endogenous levels (Figure 3.4A-B), suggesting both a genetic and physical interaction between these biogenesis machineries.

One potential trivial explanation for genetic epistasis between the EMC and BOS complex is that, following insertion of TMD1 of a GPRC by the EMC, the multipass translocon is responsible for inserting the remaining downstream TMDs. However, we found that addition of a signal sequence or signal anchor to the N-terminus of AGTR2 or the GPCR ADRA1A, which allows them to bypass the EMC and utilize Sec61 for insertion of its first TMD (Chitwood et al., 2018), markedly rescues its dependence on the BOS complex for biogenesis (Figure 3.4C and S3.3C-D). Indeed, the rescue of BOS complex dependence upon addition of a signal sequence is a similar magnitude to that observed for the EMC. We therefore concluded that there may be an additional role of the BOS complex at the EMC, beyond its previously reported function at as part of the multipass translocon. We therefore sought to determine whether the BOS complex may function as a co-factor of the EMC in biogenesis of multipass membrane proteins.

**The BOS complex is a direct physical interactor of the EMC**

Though the BOS and PAT complexes co-immunoprecipitated with the EMC, we first sought to confirm that this reflected a direct physical interaction. To do this we incubated intact cells under conditions where the EMC (Figure S3.3A) and the BOS complex (Figure S3.3B) components are present at endogenous levels with the amine-reactive chemical crosslinker DSP. DSP has a length of ~12Å, such that only factors within close proximity can be covalently crosslinked. The resulting crosslinked species were immunoprecipitated under denaturing conditions in SDS, where we found that subunits of the BOS (NOMO) and PAT (CCDC47) complexes specifically immunoprecipitated with EMC7 under conditions in which other EMC subunits are markedly depleted (Figure 3.4D). NOMO did not crosslink to EMC5, confirming that the limited crosslinking conditions, and thereby co-immunoprecipitation with EMC7, was specific.

To confirm that this result did not reflect a long-range interaction between the flexible IgG domains of NOMO and the EMC, we performed similar experiments using affinity tags on the other BOS complex subunits that are either more rigid (NCLN) or fully embedded in the bilayer (TMEM147). We found that immunoprecipitation of TMEM147 and NCLN after chemical crosslinking

specifically recovered EMC3,7, and 10, but not EMC2 (Figure 3.4E-F). We therefore concluded that the BOS complex is a direct physical interactor of the EMC in native membranes.

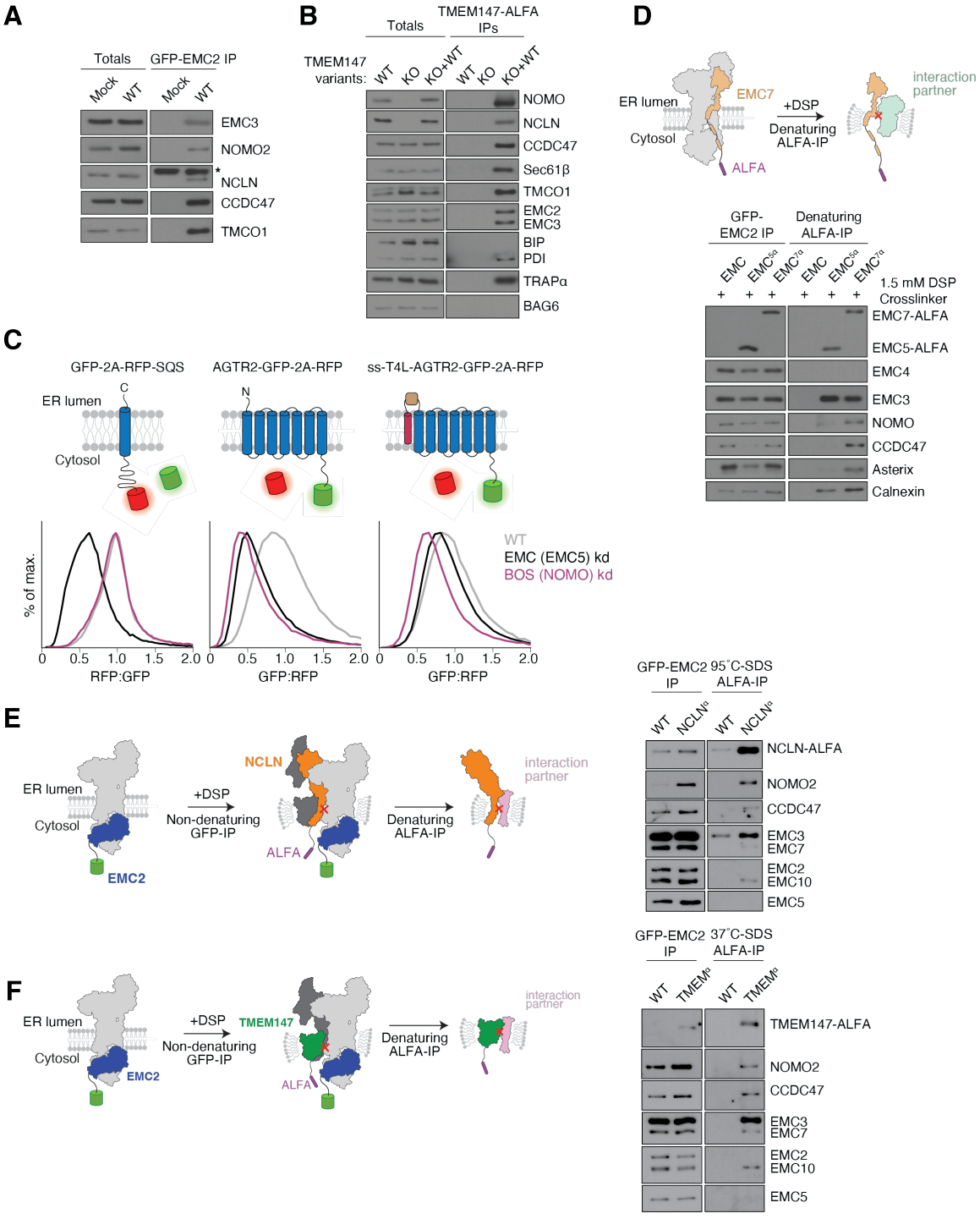



**Figure 3.4. The EMC and the BOS complex are direct physical interactors. (A)** Cell lines stably expressing GFP-EMC2 were solubilized in the detergent GDN, immunoprecipitated under

native conditions using anti-GFP nanobody, and analyzed by SDS-PAGE and western blotting compared to a mock control. An asterisk indicates a cross-reacting band. **(B)** As in (A) with anti-ALFA nanobody immunoprecipitation of TMEM147-ALFA expressed at endogenous levels in a TMEM147 KO cell line. **(C)** Flow cytometry analysis of the ratiometric SQS and AGTR2 protein reporters described Figure 3.2. The signal sequence (ss) of Pre-Prolactin followed by T4 Lysozyme (T4L) was appended to the N-terminal of AGTR2. RPE1 cells were treated with siRNAs targeting EMC5, NOMO, or a non-targeting control (WT) and transduced with the appropriate fluorescent reporter construct. Histograms of the RFP:GFP ratio (SEC61β) or GFP:RFP ratio (AGTR2 with and without the signal sequence) are shown. **(D)** Chemical crosslinking and immunoprecipitations from stable cell lines expressing GFP-EMC2 alone, GFP-EMC2 and EMC7-ALFA or GFP-EMC2 and EMC5-ALFA. The crosslinker DSP is thiol reducible, and the samples were analyzed using SDS-PAGE under reducing conditions and western blotting. **(E)** As in (D), but for cell lines stably expressing GFP-EMC2 and NCLN-ALFA. **(F)** As in (D) for cell lines expressing GFP-EMC2 and TMEM147-ALFA. Note that since TMEM147 almost quantitatively precipitates after boiling in SDS, denaturation at 37˚C was chosen to enable its immunoprecipitation.

To better understand the interaction between the EMC and the BOS complex, we sought to determine a structure of the 12-subunit holocomplex purified from human cells. Though we have shown that the BOS complex and the EMC interact without exogenous stabilization (Figure 3.4A-B), to increase their local concentration and thereby enable structural analysis, we introduced an ~50 amino acid linker between TMEM147 and EMC2. By using an extremely long and flexible linker, we avoid artificially stabilizing a non-physiologic interaction between the EMC and BOS complex. Indeed, modeling suggested that this >100 Å linker would not preclude interaction of the EMC and BOS complexes in any orientation or arrangement. Using single particle cryo-EM, we determined a modest resolution structure of the EMC•BOS holocomplex to an overall resolution of 5.2 Å in the detergent GDN (Figure 3.5A and S3.4A). From the EM density, it is clear that BOS and the EMC are within the same detergent micelle, and interact at a fixed orientation relative to each other.

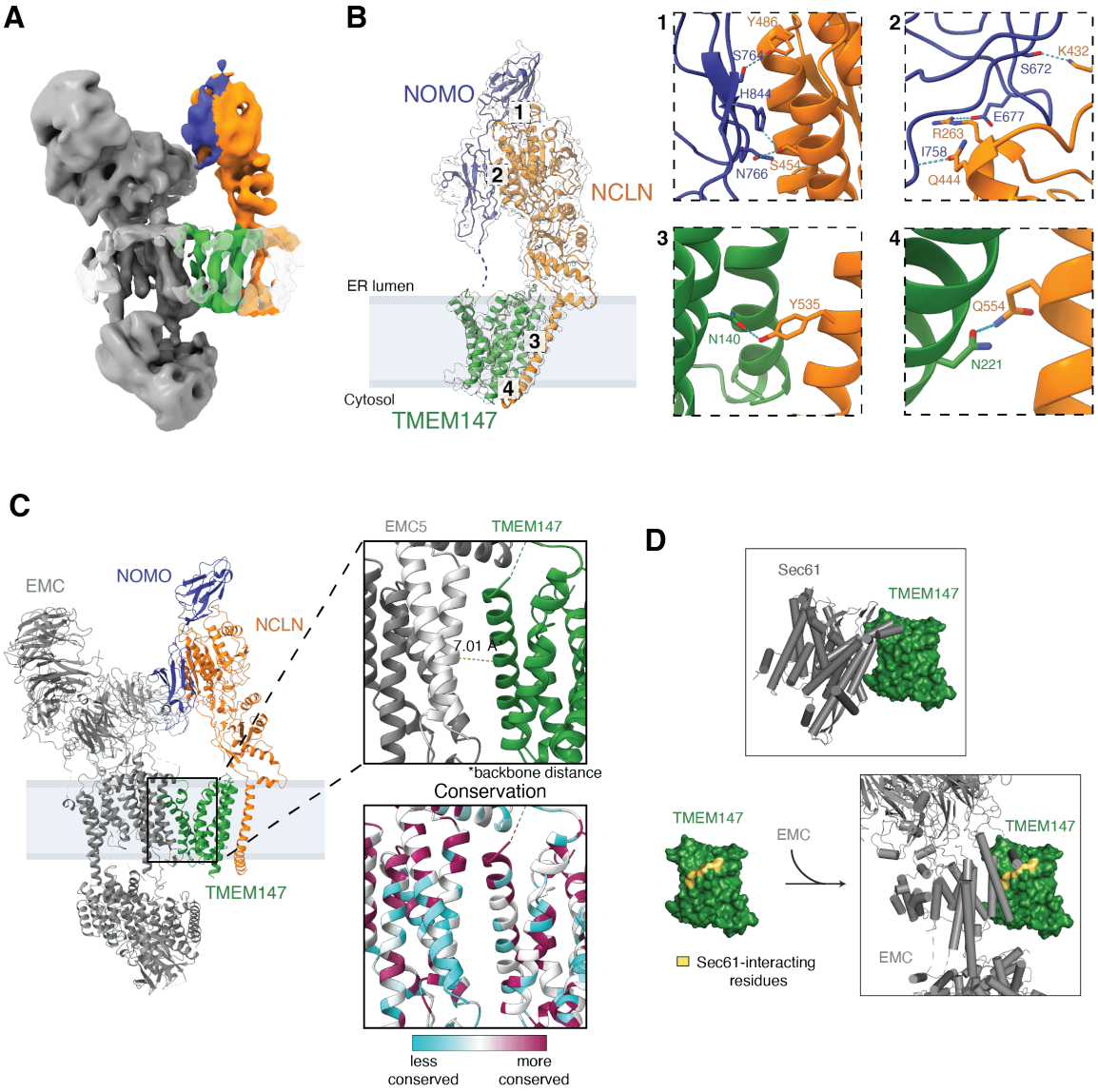



**Figure 3.5. Structural analysis of the EMC•BOS holocomplex. (A)** Coulomb potential map of the EMC•BOS complex. **(B)** In order to enable de novo modeling of the BOS complex, we isolated two different versions containing either full length NOMO with all 12 IgG repeats (fNOMO), or a truncated version including only the 3 c-terminal IgG domains (tNOMO). The molecular model of the tNOMO complex (BOS complex with truncated NOMO) in its EM map is shown. Insets 1-4 show zoomed-in regions where the subunits are in close contact with one another. Contacting residues are labeled accordingly. Insets 1 and 2 show contacts in the lumen, while 3 and 4 show contacts in the membrane. **(C)** Molecular model of the BOS (fNOMO) • EMC complex. The top right image shows a close-up view between EMC5 and TMEM147 to illustrate their proximity, with the backbone distance being 7.0 Å in one instance. The bottom right image shows the

conservation of the same region, with the less conserved residues colored in cyan and more conserved residues colored in dark magenta. **(D)** Comparison of the interacting surfaces with TMEM147 of Sec61 and EMC. Top: Cylinder cartoon representation of Sec61 interacting with TMEM147 shown in surface representation (PDB 6W6L) (McGilvray et al., 2020). Bottom left, the TMEM147 residues that are interacting with Sec61 are highlighted in yellow. These residues were determined using the *InterfaceResidues.py* script from the PyMOL wiki (https://pymolwiki.org/index.php/InterfaceResidues). Right: Cylinder cartoon representation of EMC interacting with TMEM147 with the Sec61-interacting residues being highlighted to show the overlap of interacting surfaces between Sec61 and EMC.

While we could unambiguously fit existing models of the isolated EMC into the holocomplex density map, the resolution was insufficient for de novo building of the BOS complex. Therefore, using an affinity tag on TMEM147 and more stringent conditions, we purified the isolated BOS complex and determined two structures: (1) BOS(fNOMO), using the full-length NOMO, including its 12 endogenous IgG domains (~8 Å resolution) (Figure S3.4B and S3.5A); and (2) BOS(tNOMO) in which we truncated all but the last 3 IgG repeats of NOMO (3.7 Å resolution) (Figure 3.5B and S3.4C). Notably, the truncated NOMO resulted in a higher purity protein sample and improved monodispersity upon freezing on an EM grid. In both structures, only the 2-3 terminal IgG domains of NOMO interact with the lumenal domain of NCLN, suggesting that the remaining IgG repeats are likely dynamic. Indeed, the previously reported structure of the multipass translocon complex (McGilvray et al., 2020; Smalinskaitė et al., 2022) could not unambiguously assign density for any of the IgG domains. Superposition of the density of the full-length NOMO and truncated NOMO in both complexes suggest that they are qualitatively identical, validating use of the truncated complex for high-resolution model building (Figure S3.5A).

We then used this model to unambiguously fit the EM density in the EMC•BOS holocomplex. The EMC and BOS interaction forms a buried surface area of ~855 $Å^2$. There are two primary interfaces: both an intramembrane interface, composed primarily of EMC5 and TMEM147, as well as a smaller lumenal interface between NCLN and EMC1. Notably, the intermembrane surface between the EMC and BOS complexes is composed primarily of conserved hydrophobic residues in EMC5 and TMEM147 (Figure 3.5C and S3.5C), suggesting that this interface may be the more important than that in the lumen. We hypothesize that the absence of crosslinks previously observed between EMC5 and NOMO (Figure 3.4D) can be explained by the lack of primary amines in the

membrane-embedded EMC5 subunit. Finally, the interaction surface with the BOS complex is distinct compared to that reported for the chaperone-binding mode of EMC (Figure S3.5B) (Z. Chen et al., 2023).

Comparison with structures of the multipass translocon bound at Sec61 suggest that BOS binding to the EMC and Sec61 is mutually exclusive (Figure 3.5D) (McGilvray et al., 2020). Conversely, binding of BOS to the GEL and PAT complexes would all be compatible with interaction at the EMC (Figure S3.5D). Indeed, we observed co-immunoprecipitation of both PAT and GEL complex subunits with the EMC by both quantitative proteomics and western blotting, and observed that their interaction with the EMC is independent of EMC-BOS interaction (Figure 3.3A,3.4A and S3.5E). Further, we verified using chemical crosslinking that CCDC47 is a direct physical interactor of the EMC and that its interaction is EMC7-dependent (Figure 3.4D, S3.5F). EMC7 is a peripheral subunit of the EMC, and as such, knockout of EMC7 does not destabilize core EMC subunits (Pleiner et al., 2023). This therefore suggests that the interaction between the EMC and CCDC47 is highly specific, as loss of a single peripheral subunit abolishes CCDC47 interaction with the EMC. Cumulatively, these data suggest that the multipass translocon, including the BOS, PAT, and GEL complexes, are also assembled at the EMC in a mutually exclusive manner to their binding to Sec61.

**Biophysical properties of substrate soluble domains dictate biogenesis pathway**

Having established both a genetic and physical interaction between the EMC and BOS complex, we sought to determine the function of the BOS complex in the biogenesis of EMC-dependent substrates. Analysis of the genome-wide and arrayed screens revealed patterns in the biophysical and topological features that confer dependence on the BOS complex during biogenesis. First, it was clear that the function of the BOS complex is specific to multipass membrane proteins, as it was not identified in the TA screen, and its depletion did not affect single spanning substrates in our arrayed panel (Figure 3.1,3.2). However, not all multipass substrates are equally dependent on the BOS complex. We therefore reasoned that if the EMC is specifically required for insertion of the first TMD of these multipass substrates, dependence on the BOS complex may be conferred by properties of this TMD and its surrounding sequence.

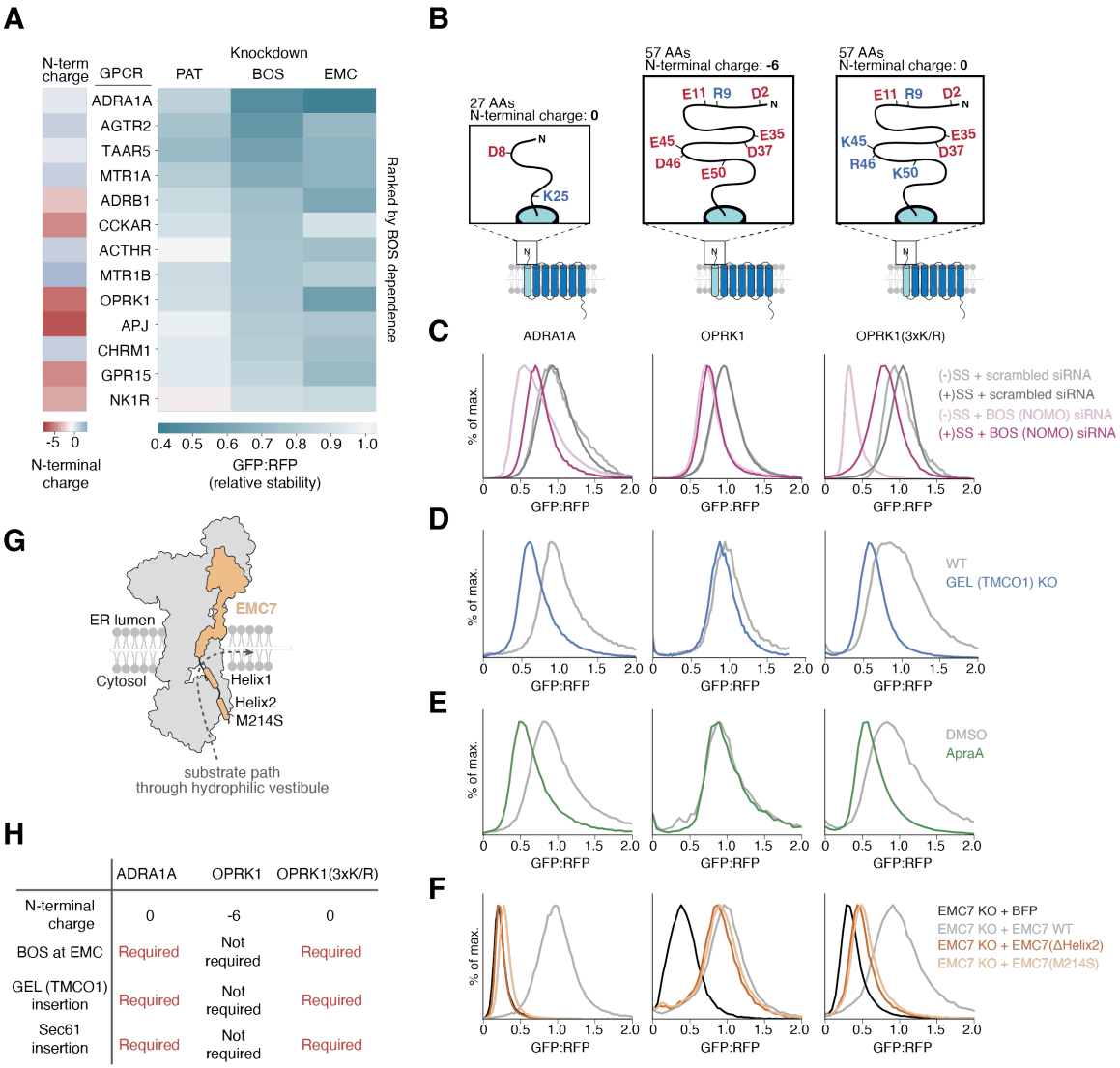



**Figure 3.6. Charge of the N-terminal domain of EMC substrates determines the requirement of the BOS complex at EMC**

**(A)** The indicated GPCRs were expressed in the GFP-2A-RFP cassette after siRNA knockdown of EMC5, NOMO, CCDC47, or a non-targeting control in RPE1 cells and analyzed by flow cytometry. The relative stability of each GPCR was determined as a ratio of GFP:RFP in the knockdown condition compared to the control. (Right) Shown is a heatmap of the relative stability of each GPCR, ranked by BOS dependence. (Left) A heatmap indicating the charge of the N-terminal domain of each GPCR. **(B)** Schematics of GPCRs ADRA1A, OPRK1 and a variant of OPRK1 with a net neutral N-terminal domain (3xKR). **(C)** Histograms from flow cytometry assays with ADRA1A, OPRK1, or OPRK1(3xKR) with or without the Pre-Prolactin signal sequence (SS) and T4L Lysozyme, as described in Figure 3.4C. Cells were treated with siRNAs targeting NOMO

or a scrambled control sequence. **(D)** As in (C), in WT or TMCO1 KO cells and with the signal anchored GPCR reporters. **(E)** Reporters in (D) were assayed for SEC61-dependent insertion using the inhibitor Apratoxin A. Cells were analyzed as in (D). **(F)** Reporters in (D) were assayed for EMC7 dependence using flow cytometry. EMC7 KO cells were transduced with either a BFP control, WT EMC7, EMC7 with a deletion of Helix2 (ΔHelix2), or EMC7 containing a single point mutation in Helix2 (M214S) before analysis as described in (D). **(G)** Cartoon of the EMC with EMC7 highlighted showing a substrate's path of insertion into the ER bilayer (Pleiner et al., 2023). **(H)** Summary of data in (C)-(F).

Comparison of the M and ORF3a screens suggested that one marked difference between these two topologically identical proteins, was the length and charge of the N-terminus of ORF3a (Figure 3.1C,D). Indeed, AGTR2 and ADRA1A, which also displayed a clear reliance on the BOS complex when utilizing the EMC (Figure 3.4C and S3.3C,D), contain an atypically positively charged soluble N-terminus in comparison to most GPCRs that lack signal sequences (Wallin & von Heijne, 1995). Previous experiments found that the EMC uses a positively charged hydrophilic vestibule for insertion, which limits integration of substrates containing positively charged soluble domains (Pleiner et al., 2023; Wu & Hegde, 2023). Consistent with this observation, most GPCRs and ER-targeted TA proteins contain neutral or negatively charged soluble domains that would be translocated through the hydrophilic vestibule of the EMC. We therefore hypothesized that when the first TMD of a multipass protein is inserted by the EMC, the charge of the N-terminal soluble domain confers dependence on the BOS complex.

To test this, we selected a suite of GPCRs and tested their dependence on EMC, BOS, and PAT complexes for biogenesis (Figure 3.6A and S3.6A,B). We found that in general, those substrates with more positively charged N-termini displayed a stronger dependence on the BOS complex (Figure 3.6A), but observed no connection between charge and EMC dependence (Figure S3.6A). However, this correlation was imperfect, suggesting that features beyond simply charge may play an important role.

To interrogate the relationship between N-terminal charge and BOS complex dependence directly, we chose two representative GPCRs, OPRK1 and ADRA1A. Both substrates display strong EMC dependence, but contain distinct N-terminal charge (-6 vs +0) (Figure 3.6B). As previously observed for AGTR2, ADRA1A only requires the BOS complex when utilizing the EMC for

insertion of its first TMD (Figure 3.4C, 3.6C and S3.6C,D). In contrast, while addition of a signal sequence to the N-terminus of OPRK1 rescues the effect of EMC depletion, it has minimal effect on the BOS complex dependence. However, an OPRK1 mutant with three additional positive charges within its N-terminus no longer relies on the BOS complex when inserted by Sec61. We therefore concluded that by studying the insertion of these three substrates we could precisely test the effects of the soluble N-terminus of a GPCR on its biogenesis.

Given earlier data suggesting that positively charged soluble domains are less efficiently translocated by the EMC during insertion, we wondered if these substrates might be more likely to rely on alternative insertase pathways. To do this, we tested whether GPCRs displayed differential dependence on the GEL complex (i.e., TMCO1) and the insertase activity of Sec61. To prevent pleotropic effects of Sec61 depletion and specifically query the role of its insertase activity, we used the inhibitor Apratoxin A, which prevents opening of the Sec61 lateral gate (Paatero et al., 2016; Thornburg et al., 2013). We found that substrates containing a positively charged N-terminus had increased dependence on TMCO1 and the lateral gate of Sec61. Importantly, while wild type OPRK1 displayed little or no dependence on TMCO1 or Sec61, the positively charged N-terminal OPRK1 mutant showed increased dependence on both alternative insertases (Figure 3.6D,E and S3.6D). Moreover, we observed an even greater dependence of the positively charged N-terminal OPRK1 mutant on TMCO1 and Sec61 when both insertases were impaired simultaneously, both in cells and in vitro (Figure S3.7). We therefore concluded that those multipass substrates that cannot be efficiently inserted by the EMC due to charge in their soluble domains instead rely on alternative, partially redundant, pathways for biogenesis.

Finally, we tested if these substrates might have increased dependence on the flexible, methionine-rich loops of the EMC that were previously shown to physically interact with substrates in the cytosol during their passage into the membrane (Pleiner et al., 2020, 2023). Indeed, GPCRs containing a positively charged N-terminus display increased dependence on the C-terminus of EMC7 that directly binds substrate TMDs below the hydrophilic vestibule of the EMC (Figure 3.6F,G and S3.6F). Comparison of the wild type and N-terminal positively charged mutant OPRK1 suggest that N-terminal charge is sufficient to increase the dependence on these flexible cytosolic domains. An increased reliance on these cytosolic domains of the EMC would be consistent with an increased dwell time at the EMC for positively charged substrates that are not efficiently translocated through the positively charged vestibule. This could also provide a potential

mechanism by which substrates are transferred to alternative insertases (i.e., TMCO1, Sec61) if not immediately inserted by the EMC.

## 3.4 Discussion

### Topology and biophysical properties dictate biogenesis pathway

For many decades it was postulated that the Sec61 translocation channel provided the major pathway into the ER for integral membrane proteins (Shao & Hegde, 2011). However, it is clear that Sec61 alone is insufficient for the insertion and folding of most ER-targeted substrates. By systematically studying the biogenesis of membrane proteins using a series of forward and arrayed genetic screens, we have begun to dissect the substrate specificity of the suite of biogenesis factors in the ER (Figure 3.1,3.2). Multiple insertases beyond Sec61, including the Oxa1 superfamily members GET1/2, EMC, and the GEL complex, are required for insertion of many single spanning and multipass membrane proteins. As expected, the post-translational insertase GET1/2 had the most pronounced role in TA biogenesis, but its depletion also affected some multipass substrates. This result could be indirect, resulting from increased flux of endogenous TAs through competing pathways in the absence of GET1/2. However, it is also possible that under some circumstances, GET1/2 can play a broader role in membrane protein biogenesis at the ER. In contrast, loss of the GEL complex, the insertase component of the multipass translocon, had no effect on single-spanning membrane proteins.

However even amongst proteins of identical topology, we observed distinct biogenesis requirements. For example, comparison of several GPCRs showed that the Sec61-associated chaperone TRAP, was required for OPRK1 biogenesis but not ADRA1A (Figure 3.2). Conversely, loss of the BOS complex had a more pronounced effect on ADRA1A and AGTR2 than OPRK1. Therefore, our data suggests that both topology and the biophysical properties of the substrate must determine the biogenesis pathway.

Earlier work has established that the hydrophobicity of a substrate TMD dictates biogenesis requirements, in some cases requiring unique insertion machinery (i.e. for TAs) or chaperones (i.e. Asterix) (Chitwood & Hegde, 2020; Guna, Hazu, et al., 2023; Meacock et al., 2002). Our data suggests that it is not only the TMD that can confer dependence on particular biogenesis pathways, but also the biophysical properties of the associated soluble domains. For example, while wild type

OPRK1 shows little to no dependence on the GEL complex or Sec61 for insertion of its first TMD, a mutant that contains three point mutations within the soluble N-terminus confers dependence on both alternative insertases. The immediate consequence of these observations is that sequences distant from the TMD, in some cases as far as 50 amino acids away, have a profound effect on both insertion efficiency and biogenesis pathway into the ER. It is therefore clear that not only are the biophysical properties of the TMD itself, but also its associated soluble domains an important consideration for biogenesis of diverse membrane proteins.

**The EMC is a central organizing hub for biogenesis and quality control**

Having delineated roles for individual factors in the ER, we used a series of functional and structural approaches to define how these components cooperate to achieve insertion and folding of diverse membrane proteins. Comparison of the forward and arrayed screens place the EMC at the center of the membrane protein biogenesis and quality control machinery in the ER. Biochemically, we observed tight coupling of the EMC with quality control machinery (e.g., p97), which ensures mislocalized or mis-inserted membrane proteins are efficiently triaged for extraction and degradation (Figure 3.3A). Conversely, co-localization of the EMC with ER resident chaperones (e.g., CANX) and post-translational modification machinery (e.g., glycosylation) guarantees that nascent substrates are immediately captured for modification and folding upon integration into the ER. This type of local recruitment of biogenesis and quality control machinery to the site of protein insertion is analogous to that observed for the translocation channel, Sec61 (Gemmer et al., 2023; Görlich et al., 1992; Hartmann et al., 1993; Jaskolowski et al., 2023; Kalies et al., 1998; Voigt et al., 1996). Further, direct recruitment of the multipass translocon components, including the BOS, GEL, and PAT complexes to the EMC facilitates organization of ‘biogenesis hubs’ in the ER membrane. The observation that at least some of these factors are stabilized by interaction with EMC7 (Figure S3.5F) provides one potential explanation for the conservation of the large soluble lumenal and cytoplasmic domains of the EMC: recruitment and retention of auxiliary factors to the site of membrane integration.

**A working model for multipass membrane protein insertion**

The organization of biosynthesis machinery into local and dynamic hubs within the ER—including the EMC, Sec61, and the multipass translocon components—provides a coherent working model for the insertion and folding of complex multipass membrane proteins. A ribosome nascent chain complex is delivered to the ER by the signal recognition particle (SRP), where substrates first probe

the hydrophilic groove of the EMC. Though the molecular details of handover from SRP to the EMC are not yet precisely defined, SRP receptor subunits were recovered in native co-immunoprecipitation of the EMC from ER membranes, suggesting one potential mechanism for recruitment to the EMC (Figure 3.3A). Models consistent with a 'first refusal' of substrates by the EMC, best explain data showing the EMC can enforce the correct folding of multipass substrates containing positively charged extracellular domains (Pleiner et al., 2023). Our observation that the PAT complex directly interacts with the EMC, including CCDC47, which is known to use its soluble coiled-coil domain to bind the ribosome, provides a putative mechanism for transiently stabilizing ribosome nascent chain complexes at the EMC (McGilvray et al., 2020; Smalinskaitė et al., 2022; Sundaram et al., 2022). Recruitment of the PAT complex in an EMC7-dependent manner also provides one potential explanation for the differential dependence of substrates on EMC7 (Figure 3.6F).

Substrates in which the first TMD can be efficiently inserted by the EMC, such as those with negatively charged and short N-terminal soluble domains, passage through its hydrophilic vestibule into the bilayer. Co-localization of the EMC with glycosylation machinery would then allow immediate post-translational modification of the soluble N-terminus. In contrast, substrates that are poorly inserted by the EMC (i.e., those that have increased N-terminal positive charge but also likely including those with longer or more structured soluble domains) have a longer dwell time at the cytosolic vestibule of the EMC. This is consistent with their increased reliance on the cytosolic C-terminus of EMC7 that contains several conserved hydrophobic residues, previously shown to directly interact with substrates in the cytosol (Pleiner et al., 2023). These data are consistent with the model that the rate limiting step for insertion is translocation of the N-terminal soluble domain through the hydrophilic vestibule of the EMC. The properties of the N-terminus therefore dictate the energetic barrier for translocation into the ER lumen.

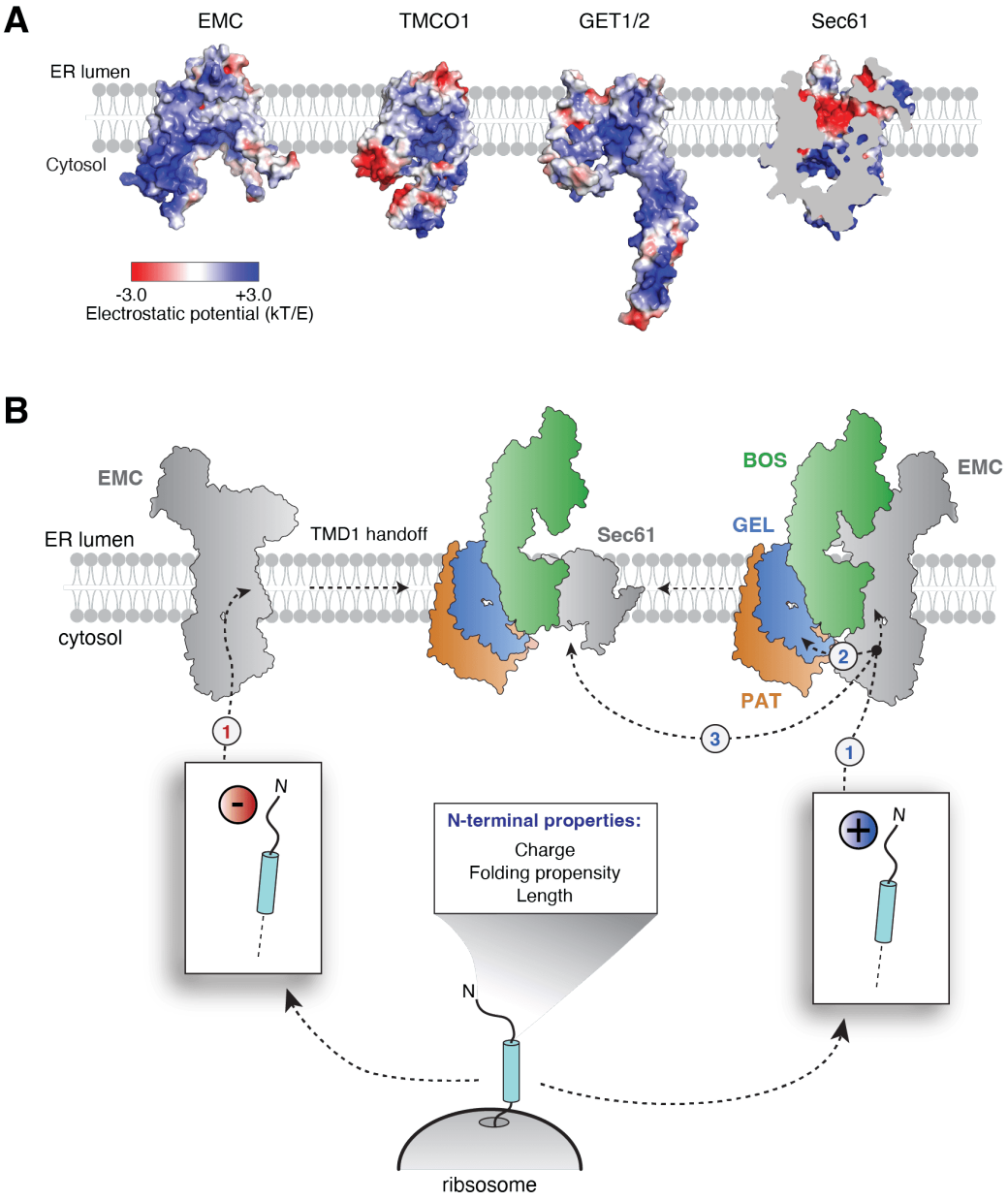



**Figure 3.7. Model of multipass membrane protein biogenesis at the ER**

**(A)** Comparison of the electrostatic surfaces of known human insertases at the ER membrane. Electrostatic potential was generated using APBS with range -3 to +3 kT/e and mapped onto surface representation of EMC3/6 (PDB ID 8S9S) (Pleiner et al., 2023), TMCO1 (PDB ID 6W6L) (McGilvray et al., 2020), GET1/2 (PDB ID 6SO5) (McDowell et al., 2020), and Sec61 (PDB ID 3JC2). (Voorhees & Hegde, 2016) **(B)** Model for membrane protein insertion at the ER. Multipass $N_{exo}$ proteins are co-translationally targeted to the EMC, where the properties of their N-terminal soluble domain alter the requirements for insertion of the first TMD. (Left) Proteins with a net

negative charge in their N-terminal domain are able to utilize the EMC for TMD1 insertion without additional co-factors. After insertion of TMD1, these proteins are then handed off to the multipass translocon at Sec61 for insertion of the remaining TMDs. (Right) Proteins that have net neutral or positive N-terminal domains require additional factors for insertion of the first TMD. The BOS complex physically associates with the EMC and is required for EMC-dependent insertion of the first TMD. Additionally, these substrates, are routed to TMCO1 or Sec61 for TMD1 insertion. After TMD1 is inserted, the multipass translocon at Sec61 inserts the remaining TMDs. Binding of the BOS, GEL, and PAT complex to the EMC may facilitate hand-over of substrates between the EMC and the multipass translocon during membrane protein biogenesis. This co-translational triaging of ribosome nascent chain complexes with exposed TMD1 based on N-terminal domain features is likely dictated by the biophysical properties and architecture of the EMC's hydrophilic vestibule. Charge repulsion between EMC's positively charged hydrophilic vestibule and its clients' N-terminal domains would increase dwell time and facilitates the engagement of multipass translocon factors and insertion via TMCO1 or Sec61.

TMDs that are not immediately inserted by the EMC are shuttled to alternative insertases including the GEL complex (i.e., TMCO1) and in some cases Sec61. We hypothesize that handover of those substrates with charged N-terminal domains between insertases is facilitated by recruitment of the BOS, GEL, and PAT complexes to the EMC. Indeed, these substrates appear to primarily require the activity of the multipass translocon when using the EMC for insertion of their first TMDs (Figure 3.6C).

Structures of the EMC•BOS holocomplex are most consistent with a model in which interactions between biogenesis machineries are dynamic. Rather than a rigid holocomplex that includes both the EMC and Sec61, co-localization may serve primarily to increase local concentration of insertases to accommodate diverse substrates as they passage into the bilayer. This would be consistent with the observation that recruitment of the BOS complex to the EMC is mutually exclusive to its binding to Sec61 (Figure 3.5D) (McGilvray et al., 2020; Smalinskaitė et al., 2022). Mutually exclusive binding of the multipass translocon components to the EMC may also encourage handover to Sec61. Certainly, recruitment of the multipass translocon components to the EMC provides a putative mechanism for how any substrate in which the first TMD is inserted by the EMC is transferred to the back of Sec61 to complete its insertion and folding.

**A central role for Oxa1 superfamily of insertases throughout all kingdoms of life**

Based on differences in their biophysical properties (e.g., charge of the hydrophilic vestibule, limitations on length/charge/structure to be translocated), it is likely that insertases display partial substrate preferences (Figure 3.7A). However, it is clear that Oxa1 superfamily insertases such as the EMC and GEL complex are partially redundant, and loss of any one results in compensation in human cells (Figure S3.1F). Remarkably, recruitment of the core insertase subunits EMC3/6 to the inner mitochondrial membrane was sufficient to rescue loss of Oxa1 in yeast (Güngör et al., 2022). Further, the ability of one Oxa1 superfamily insertase to compensate for loss of another explains genetic studies that suggest loss of function mutations in TMCO1 are not deleterious in the human population (Karczewski et al., 2020). In these cases, upregulation of the EMC or potentially even GET1/2 may be sufficient to ensure efficient insertion of all required membrane protein substrates.

This redundancy between Oxa1 superfamily insertases explains how membrane protein biogenesis is achieved across evolutionarily distinct kingdoms of life (Anghel et al., 2017; Kumazaki et al., 2014). Superficially, it appears that the multipass translocon components are a metazoan specific adaptation, without homologs in yeast and bacteria. For example, while the EMC is conserved in all eukaryotes, the GEL complex is metazoan specific (Anghel et al., 2017). However, it is clear that even in humans, while the EMC is essential, the presence of both the EMC and GEL complexes is not strictly required. Interestingly, in fungi, which may rely primarily on the EMC for multipass insertion, the EMC is missing one of its soluble subunits (i.e., EMC8/9) (Jonikas et al., 2009). The lack of EMC8/9 would allow the EMC to sit closer to the ribosome, potentially taking the place of the GEL complex in the multipass translocon. An analogous reliance on Oxa1 superfamily members in multipass membrane protein biogenesis has been observed in bacteria as well, where in some organisms YidC is recruited to the ribosome to cooperate with SecYEG in insertion (Sachelaru et al., 2013; Wickles et al., 2014).

These observations however raise questions as to why the biogenesis machinery has expanded from bacteria to mammals. Indeed, molecular dynamics simulations suggest that given sufficient time, even multipass membrane proteins can fold autonomously (Niesen et al., 2020). Though the increased size of the membrane proteome may contribute, it is more likely that a decrease in error tolerance in multicellular organisms is the driver for evolution of additional biogenesis components. Unlike a bacterium or yeast cell, mammals rely on post-mitotic cells, including neurons and cardiomyocytes, that must persist for the entire lifespan of the organism. In these cells, the risk of

cytotoxicity from misfolding or aggregation of nascent proteins required evolution of more stringent mechanisms to protect the proteostasis of the cell and thereby the health of the organism. As a result, the insertion, folding and assembly of membrane proteins must occur with even greater efficiency to avoid premature clearance by quality control machinery. These additional insertases fulfill this role, by increasing the efficiency of membrane protein biogenesis such that it can occur even in the face of robust competing degradation pathways.

### 3.5 Acknowledgements

We thank Songye Chen for technical assistance, T. Stevens and M. Hazu and all members of the Voorhees lab for thoughtful discussion. We thank the Caltech Flow Cytometry facility, the Caltech Cryo-EM facility, and the Caltech Millard and Muriel Jacobs Genetics and Genomics Laboratory. We thank Ville Paavilainen for Apratoxin A. Cryo-electron microscopy was performed in the Beckman Institute Center for TEM at Caltech, and data was processed using the Caltech High Performance Cluster, supported by a grant from the Gordon and Betty Moore Foundation. This work was supported by: the Heritage Medical Research Institute (RMV), the NIH's National Institute Of General Medical Sciences DP2GM137412 (RMV), the Burrough's Wellcome Fund (RMV), an NSF-CAREER award 2145029 (RMV), a Human Frontier Science Program Fellowship 2019L/LT000858 (AG), the Deutsche Forschungsgemeinschaft (TP), the Tianqiao and Chrissy Chen Institute (TP), a Rosen Family fellowship (KRP), and an Arie Jan Haagen-Smit Fellowship (KRP).

### 3.6 Declaration of interests

RMV and GPT are consultants for Gates Biosciences, and RMV is an equity holder. The authors have no additional competing financial interests.

## 3.7 Supplementary Material

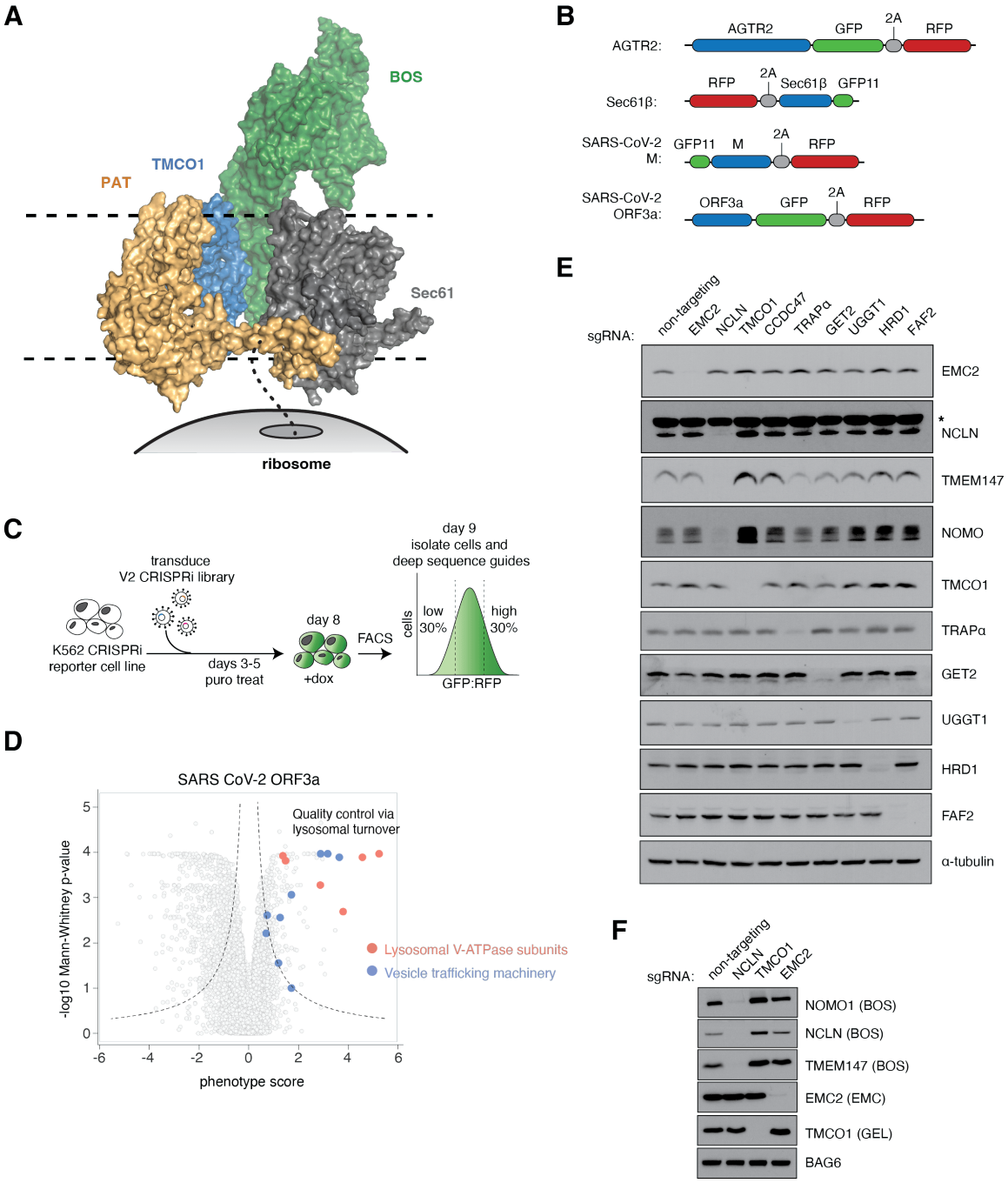



**Figure S3.1. Genome-wide analysis of membrane protein biogenesis factors.**

**(A)** Surface representation of the ER multipass translocon (PDB ID: 6W6L) (McGilvray et al., 2020) with respect to the ribosome. SEC61 is shown in grey, BOS complex in green, TMCO1 in

blue, and PAT complex in dark beige. The dotted line emerging from the ribosome represents the nascent chain of protein being inserted by the translocon. (**B**) Schematic of reporters used in CRISPRi screens. **(C)** Schematic of CRISPRi screens performed on the AGTR2, SEC61β, SARS-CoV-2 M, and SARS-CoV-2 ORF3a proteins. **(D)** Volcano plot of factors involved in the turnover of SARS-CoV-2 ORF3a protein. ORF3a is subject to degradation by the lysosome, and therefore knockdown of vesicle trafficking machinery and of the lysosomal V-ATPase subunits stabilizes the ORF3a reporter. **(E)** Analysis of gene knockdown in K562 cells by guides in CRISPRi screens. Samples were subject to SDS-PAGE and western blotting. **(F)** ER insertion machinery shows compensatory effects upon the loss of individual factors. Samples were generated and analyzed as in (E), in K562 cells. Note that loss of TMCO1 results in increased levels of NOMO, NCLN, and TMEM147, all components of the BOS complex. Additionally, loss of EMC2 results in increased levels of TMCO1, an evolutionarily related insertase within the OXA1 superfamily of insertases (Anghel et al., 2017).

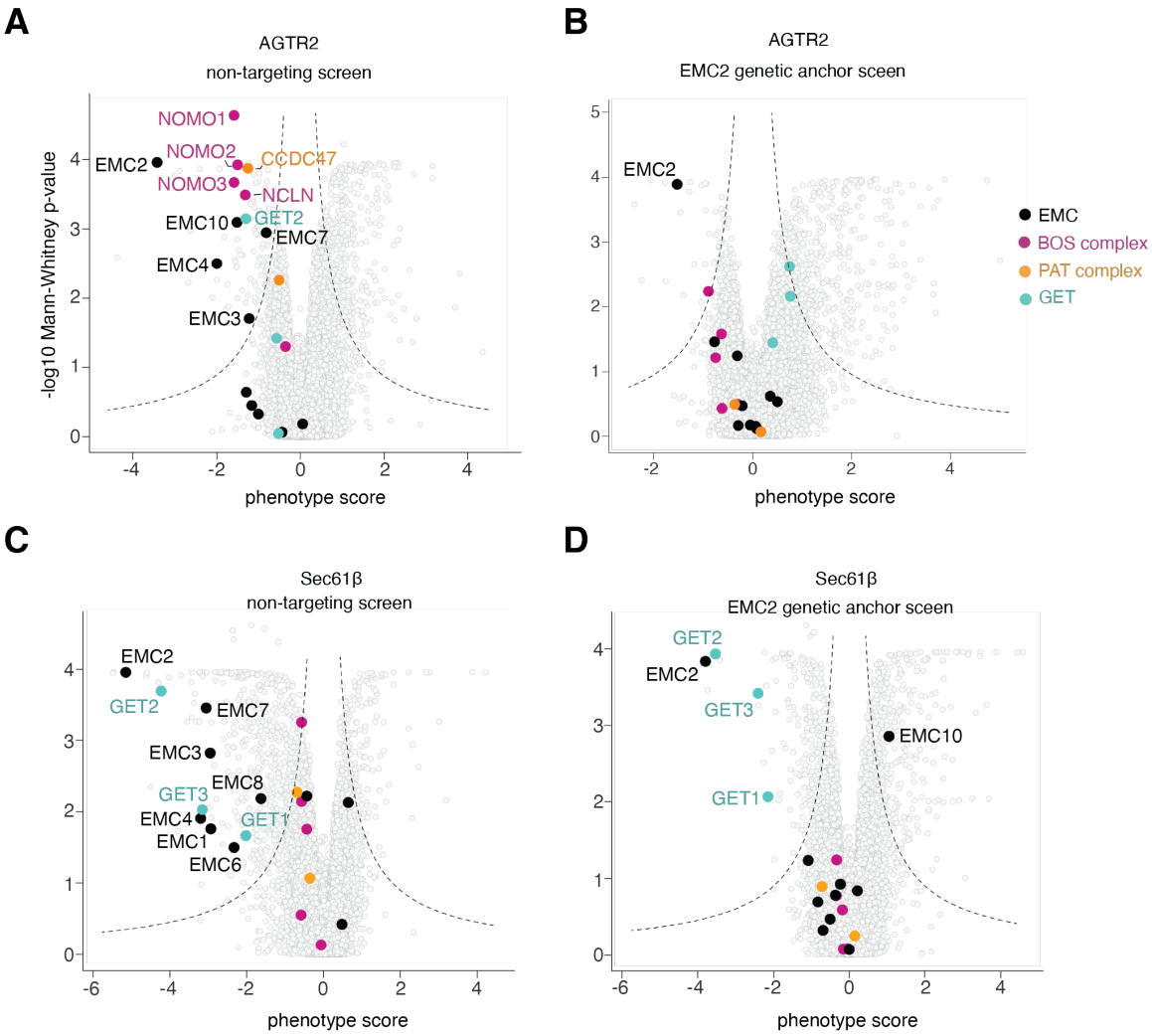



**Figure S3.2. Genetic modifier screens identify epistatic and parallel pathways with the EMC.** **(A)** Volcano plot of a dual guide CRISPRi screen of AGTR2 with the non-targeting (NT)-dual library. The EMC, multipass translocon components, and the GET pathway components are highlighted. Dotted lines indicate the significance of the gene's effect on the stability of the protein reporter, and only significant gene hits from the screen are labeled. **(B)** As in (A), but with the EMC2-dual library. Note that components in the same pathway as EMC no longer have a destabilizing effect on the AGTR2 reporter. **(C)** As in (A) for Sec61β with the NT-dual library (Guna, Page, et al., 2023). **(D)** As in (A) for Sec61β with the EMC2-dual library. Note that the EMC components are no longer significant but the GET pathway components, parallel to EMC, become more significant.

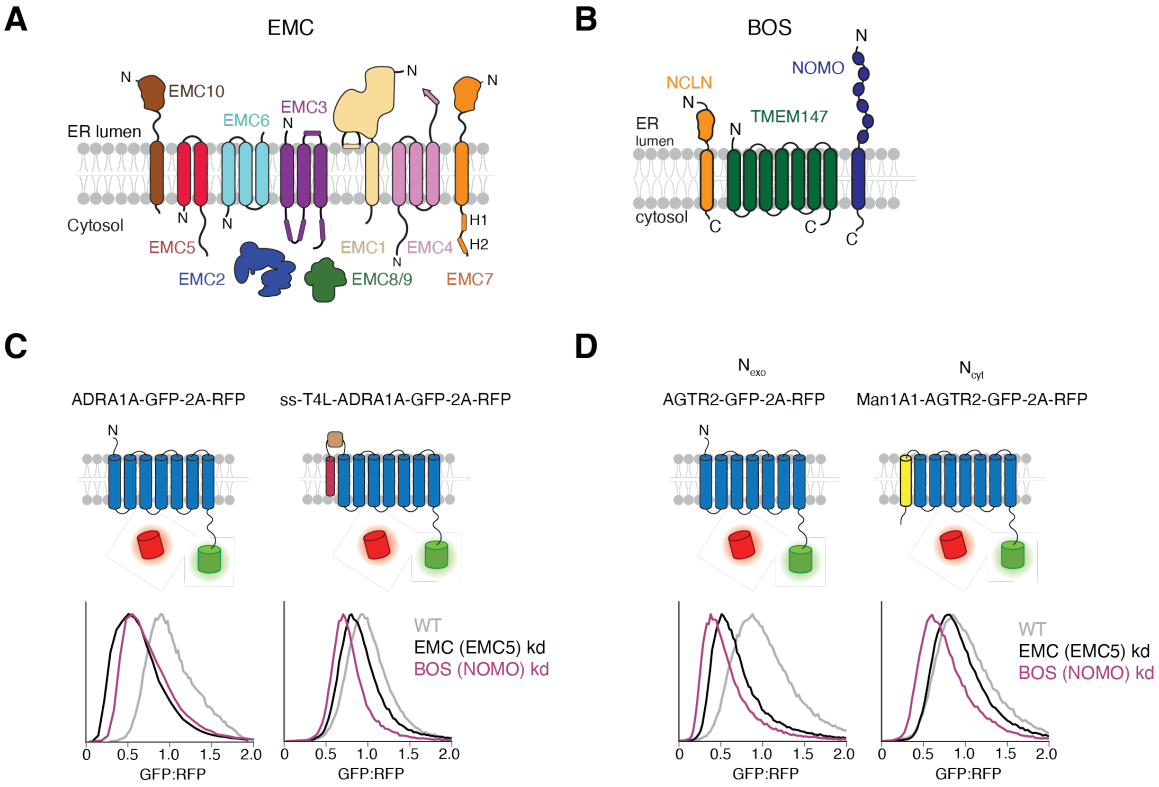



**Figure S3.3. The EMC and BOS complex topologies.**

**(A)** Topology of the nine-subunit human EMC. In mammals, the EMC is composed of seven membrane-spanning and two soluble subunits. The central insertase is composed of the two central subunits EMC3 (a homolog of YidC and member of the Oxa1 superfamily of insertases) and EMC6 (Bai et al., 2020; Miller-Vedam et al., 2020; Pleiner et al., 2020). In both yeast and humans, the EMC also contains a globular cytosolic domain (scaffolded around EMC2 and in mammals containing the redundant paralogs EMC8 or 9) and a large lumenal domain (composed of EMC1, 7, 10, and a single b-strand of EMC4). The function of the lumenal domain of the EMC is not known, but b -propellers, like those conserved in EMC1, are typically considered protein-protein interaction motifs. **(B)** Topology of the heterotrimeric BOS complex. NOMO contains 12 lumenal IgG repeats and a single C-terminal TMD. The single-spanning protein NCLN also has a large lumenal domain, and is homologous to nicastrin, a component of the g-secretase complex (Haffner et al., 2004). Finally, TMEM147 contains 7 TMDs and is homologous to APH-1 in the g-secretase complex (Dettmer et al., 2010). **(C)** Flow cytometry analysis of the ratiometric ADRA1A protein reporter with or without the N-terminal fusion of the signal sequence (ss) of Pre-Prolactin followed by T4 Lysozyme (T4L), as described in Figure 3.4C. **(D)** Flow cytometry analysis of the ratiometric

AGTR2 protein reporter with or without an N-terminal fusion of the first TMD of MAN1A1, a membrane protein of $N_{cyt}$ topology. Note that both the N-terminal fusion of MAN1A1 and of (ss)-T4 lysozyme (Figure 3.4C) behave similarly and rescue the NOMO kd and EMC5 kd phenotypes.

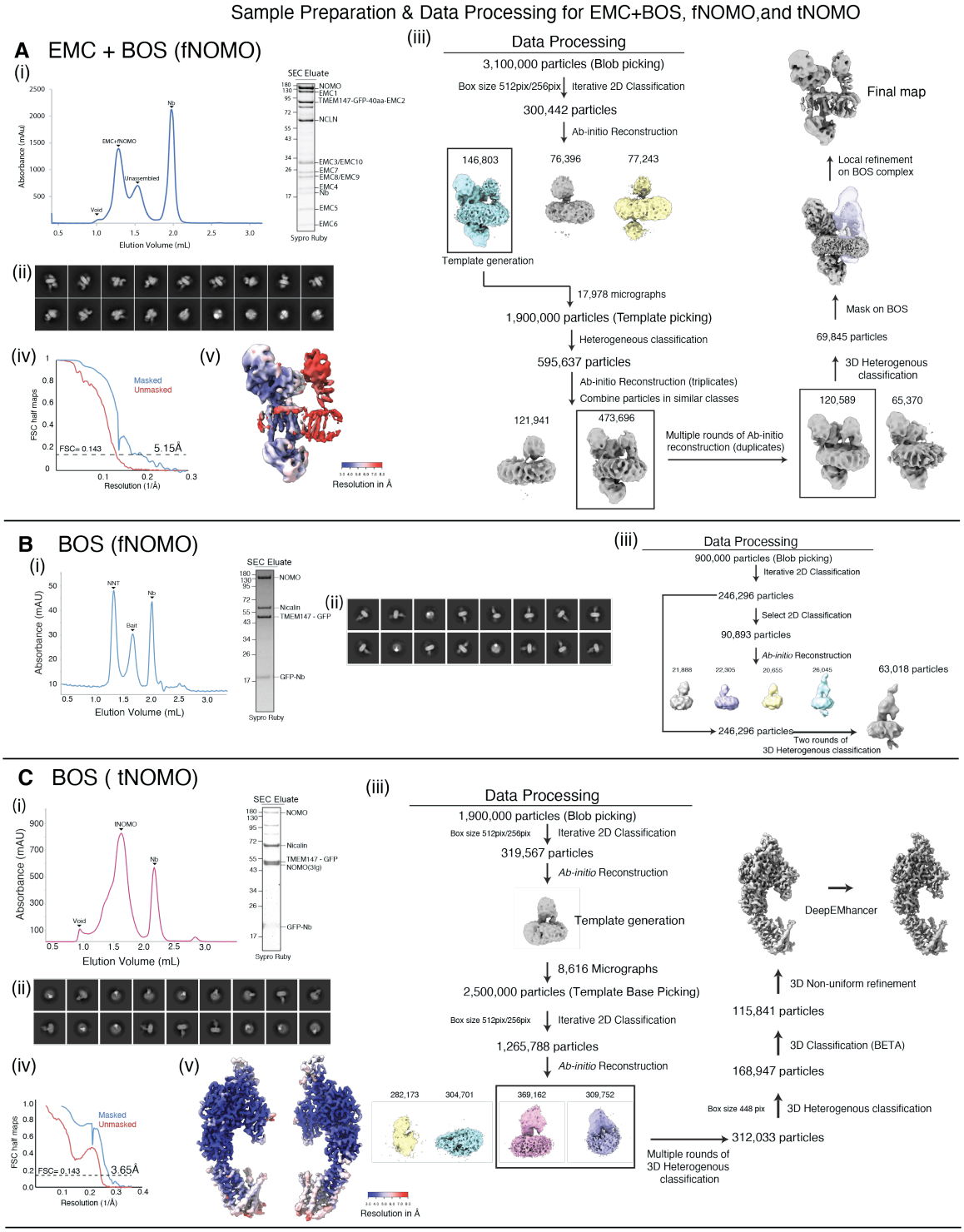



**Figure S3.4. Sample preparation and data processing of EMC • BOS (fNOMO), BOS (fNOMO), and BOS (tNOMO).**

**(A)** EMC • BOS (fNOMO). (**i**) Sample preparation of EMC • fNOMO. Left: Size exclusion chromatogram of EMC • BOS (fNOMO). The main peak corresponds to the EMC • BOS (fNOMO) super-complex, while a smaller side peak likely corresponds to non-interacting EMC and BOS (fNOMO) complexes. Right: SDS-PAGE gel of the size exclusion eluate that contained EMC • BOS (fNOMO), stained by Sypro Ruby. The subunits of EMC and NOMO complexes are labeled. (**ii**) Representative 2D classes of EMC • BOS (fNOMO) complex. (**iii**) CryoSPARC processing scheme used to identify particles corresponding to the EMC • BOS (fNOMO) complex. CryoSPARC processing scheme employed for identifying particles related to the EMC • BOS (fNOMO) complex. All processing was carried out using CryoSPARC; for details, refer to Methods (**iv**) GFSC curves for the masked and unmasked half maps of the overall EMC • BOS (fNOMO) complex map. Nominal resolution of 5.1 Å was determined using gold standard FSC cutoff = 0.143. (**v**) Resolution estimation of the EMC • BOS (fNOMO) complex, generated from cryoSPARC. **(B)** BOS (fNOMO). (**i**) Sample preparation of BOS (fNOMO). Left: Size exclusion chromatogram of BOS (fNOMO) eluant after GFP-immunoprecipitation. A280 of the run was monitored to identify peak fractions that contained BOS (fNOMO). Right: shows the SDS-PAGE gel of the purified BOS (fNOMO) complex, stained by Sypro Ruby. All the components of the BOS (fNOMO) complex were present (NOMO, NCLN, and TMEM147-GFP), along with excess anti-GFP nanobody. (**ii**) Representative 2D classes of the BOS (fNOMO) complex. (**iii**) Data processing scheme in cryoSPARC v3.2-4.2.1 to identify a set of particles that correspond to the BOS (fNOMO) complex. After a round of blob picking and iterative rounds of 2D classification, 246,296 particles were used to generate 2D classes for selection. Then, 90,893 particles were used for Ab-initio reconstruction to generate the initial models. Using the model that best resembles the BOS (fNOMO) complex, rounds of heterogeneous refinement were performed on the previously mentioned 246,296 particles to identify the final set of 63,018 particles corresponding to our final map of BOS (fNOMO). **(C)** BOS (tNOMO). (**i**) Sample preparation of BOS (tNOMO). Left: Size exclusion chromatogram of BOS (tNOMO) sample after GFP-immunoprecipitation and TEV cleavage. UV absorbance at 280nm was recorded to track protein fractions that contained BOS (tNOMO) complex. Right: SDS-PAGE gel of the peak fraction containing purified BOS (tNOMO) complex, stained by Sypro Ruby. The components of the BOS (tNOMO) complex were present (TMEM147, NCLN, and tNOMO). Since the bait used for purification was TMEM147-GFP, a portion of endogenous full-length NOMO was also detected. (**ii**) Representative 2D classes of BOS (tNOMO) complex. (**iii**) Processing scheme used in cryoSPARC to identify particles that represent the BOS (tNOMO)

complex. After an initial round of blob picking and iterative rounds of 2D classification, 319,567 particles were used to generate templates for template picking. Picked particles were put through iterative rounds of 2D classification and 1,265,788 particles were used to generate initial models to be used for multiple rounds of heterogeneous classification. 168,947 particles were subjected to 3D classification to identify the final set of 115,841 particles. The EM map corresponding to these particles was further refined using non-uniform refinement and sharpened with DeepEMhancer to yield the final map. (iv) Gold Standard Fourier Shell Correlation (GFSC) curves for the masked and unmasked half maps of the overall BOS (tNOMO) complex map. Nominal resolution of 3.6 Å was determined using gold standard FSC cutoff = 0.143. (v) Resolution estimation of the BOS (tNOMO) complex, generated from cryoSPARC.

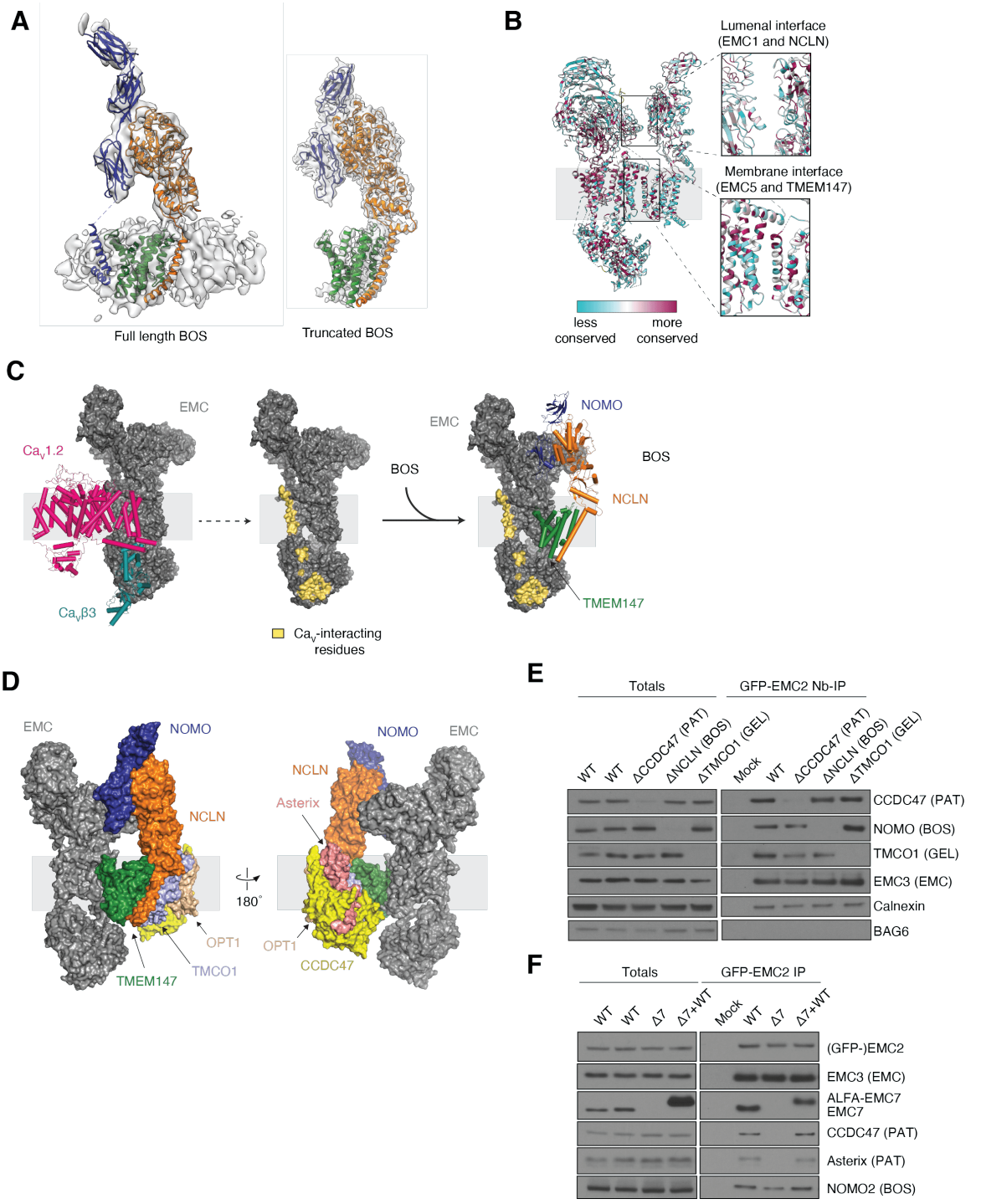



**Figure S3.5. Analysis of the BOS complex and EMC • BOS (fNOMO) complex.**

(**A**) Comparison between the full-length BOS complex and truncated BOS complex model docked in their respective EM map. From the maps, we observed that the truncated BOS complex retained the overall shape of the full-length complex, although only the last 2 IgG-like domains are visible

in the truncated BOS complex map. Here, NOMO is colored in blue, TMEM147 in green, and NCLN in orange. **(B)** Sequence conservation analysis of the EMC • BOS (fNOMO) complex. Left: Molecular model of EMC • BOS (fNOMO) with residues being colored by sequence conservation. Less conserved residues are in cyan and more conserved residues are in dark magenta. The sequence alignment contains sequences of EMC • BOS (fNOMO) from the following species: *Homo sapiens*, *Mus musculus*, *Xenopus tropicalis*, *Danio rerio*, *Drosophila melanogaster*, *Hydra vulgaris*. Right: Inlets showing conservation of the lumenal interface (between EMC1 and NCLN) and membrane interface (between EMC5 and TMEM147). We observed a conserved patch near the lumenal membrane between EMC5 and TMEM147. **(C)** Comparison of the binding surfaces of EMC with either ***Ca***V1.2/***Ca***Vβ3 complex or the BOS complex. The model was generated by aligning the EMC from its structure with ***Ca***V1.2/***Ca***Vβ3 complex (PDB ID 8EOI) (Z. Chen et al., 2023) or BOS complex (this study). Left: Cylinder cartoon representation of the ***Ca***V1.2/***Ca***Vβ3 complex (pink and dark cyan)(Z. Chen et al., 2023) and surface representation of the EMC (dark grey). Middle, the EMC residues that are interacting with ***Ca***V1.2/***Ca***Vβ3 are highlighted in yellow. These residues were determined using the *InterfaceResidues.py* script in PyMOL. Right: Cylinder cartoon representation of the BOS complex (TMEM147 in green, NCLN in orange, and NOMO in blue) and surface representation of the EMC (dark grey) with the ***Ca***V-interacting residues being highlighted to show that there is no overlap between the interacting interfaces with EMC by ***Ca***V and the BOS complex. **(D)** Superposition of EMC • BOS (fNOMO) complex with the other components of the 'ER translocon' at Sec61. The model was generated by aligning TMEM147 from EMC • fNOMO (this study) and the ER translocon (PDB ID 6W6L) (McGilvray et al., 2020). Sec61 and CCDC47's latch helices were removed from this model for clarity. EMC occupies the same position as Sec61 at the translocon. **(E)** HEK 293T cells (WT, ΔCCDC47, ΔNCLN, or ΔTMCO1) stably expressing GFP-EMC2 were subjected to anti-GFP nanobody purification under native conditions. Samples were analyzed by SDS-PAGE and western blotting. Note that the interaction of the EMC with individual multipass translocon components is independent of the EMC's interactions other multipass translocon components. **(F)** EMC7 is indispensable for interaction of the EMC with multipass translocon components. WT or EMC7 KO cells (Δ7) stably expressing GFP-EMC2 were transduced with WT EMC7 or a BFP control and subjected to anti-GFP nanobody purification under native conditions. Samples were analyzed by SDS-PAGE and western blotting.

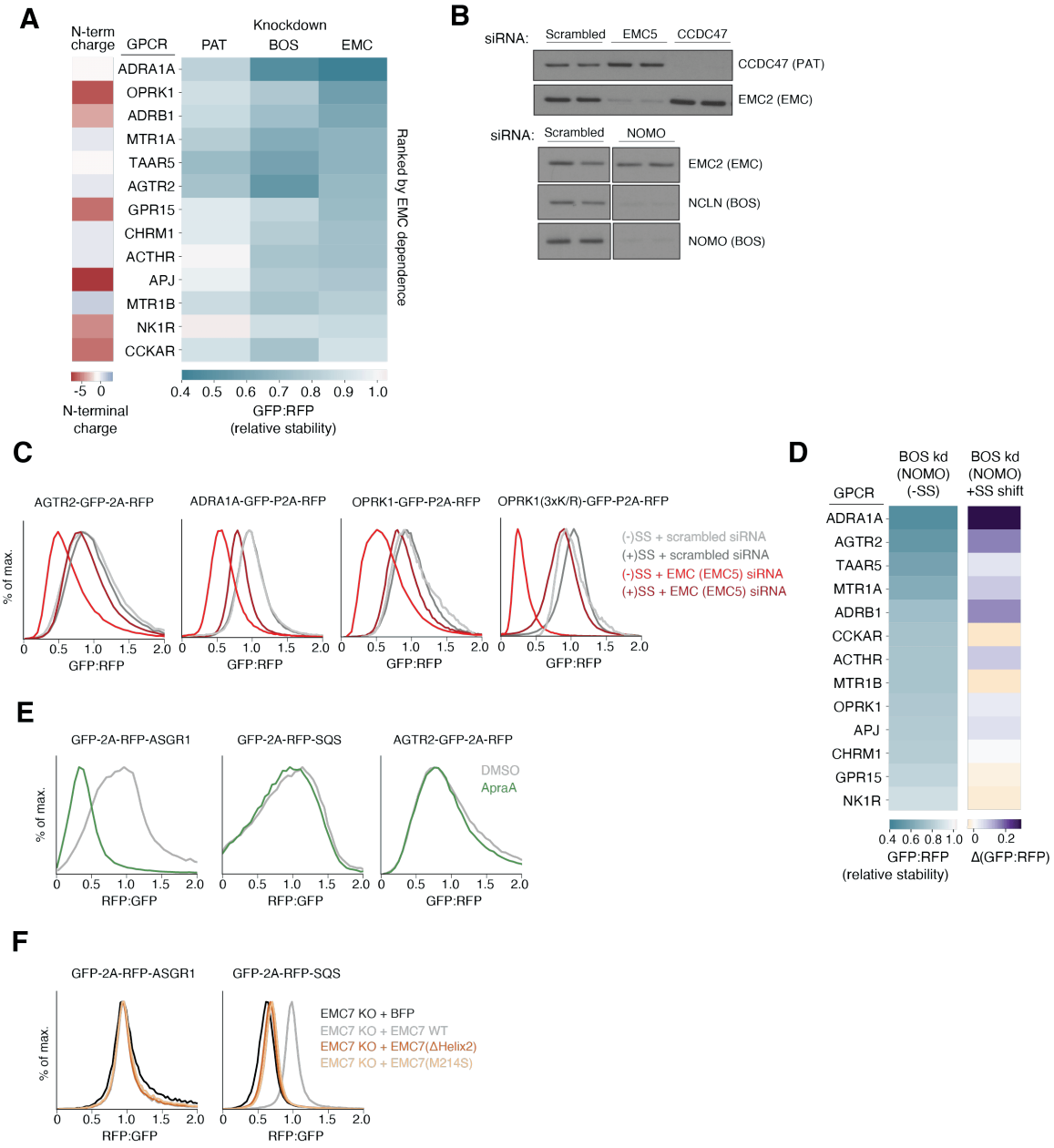



**Figure S3.6. Differential effects of biogenesis factors on diverse GPCR substrates.**

**(A)** A panel of GPCRs was assayed for EMC, BOS, or PAT dependence, as described in Figure 3.6. (Right) A heatmap displays the relative stability of each GPCR, ranked by EMC dependence. (Left) The net charge of the N-terminal soluble domain of each GPCR is indicated by the heatmap. **(B)** Validation of knockdown experiments in RPE1 cells using siRNAs targeting EMC5, CCDC47, NOMO used in Figures 3.4 and 3.6. Cells were harvested and analyzed by SDS-PAGE and western blotting. **(C)** Flow cytometry assay as in Figure 3.4C and 3.6C, with EMC5 knockdown. **(D)**

Bypassing the EMC for insertion of TMD1 of GPCRs reduces the dependence on the BOS complex. GPCR reporter constructs were generated using the GFP-2A-RFP cassette, with and without a signal sequence (SS) and T4 lysozyme fusion to the N-terminus of each GPCR. RPE1 cells were transduced with these reporters after siRNA knockdown of NOMO or a non-targeting control. (Left) A heatmap shows the relative stability of GPCRs without a signal sequence and is ranked by BOS dependence. (Right) A heatmap shows the change in stability between GPCRs containing the appended signal sequence and WT GPCRs. Darker purple indicates a positive shift in stability of the signal sequence-containing GPCR compared to the WT GPCR. **(E)** As in Figure 3.6E, for ASGR1, SQS, and AGTR2. **(F)** As in Figure 3.6F, for ASGR1 and SQS.

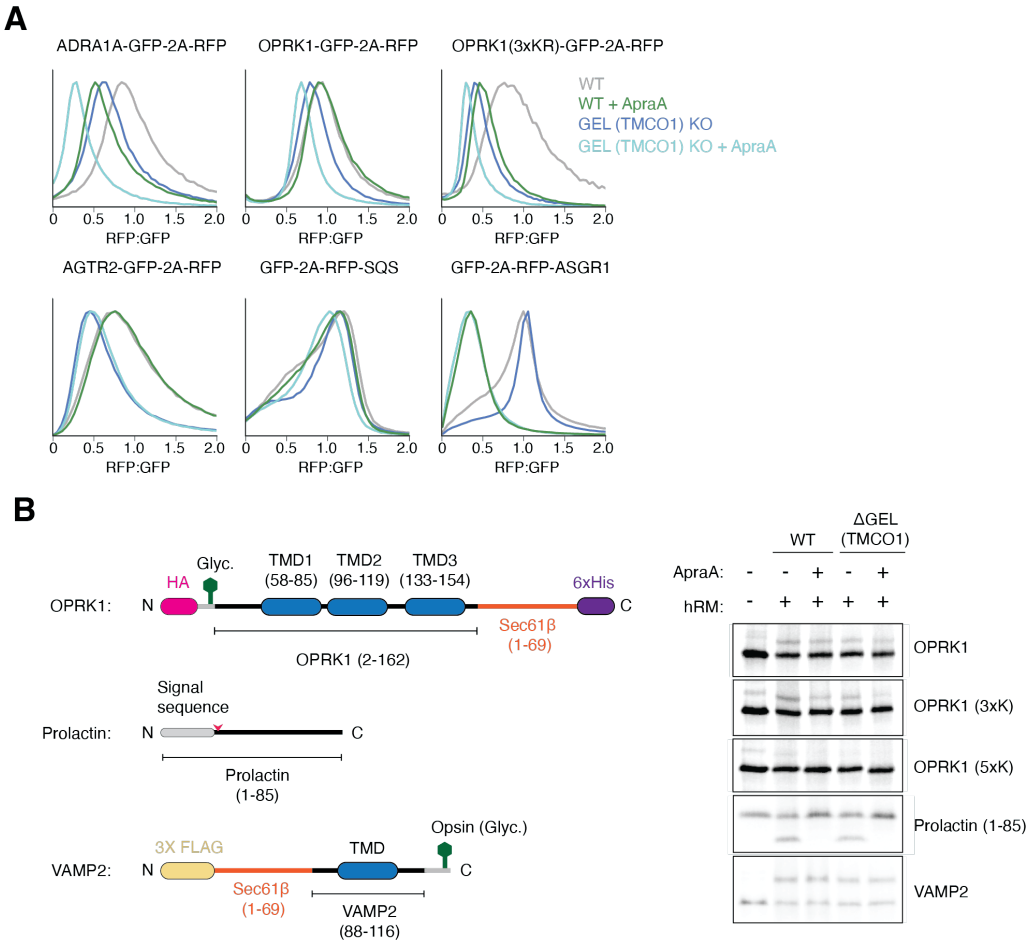



**Figure S3.7. Substrates are routed through TMCO1 and Sec61 when unable to access EMC for insertion.**

**(A)** Histograms as in Figure 3.6D,E, for WT and TMCO1 KO cells treated with Apratoxin A or a DMSO control. **(B)** *In vitro* translation of OPRK1 (wildtype, 3xK, or 5xK), Prolactin, or VAMP2. (Left) Schematic of constructs used in the translation reactions. The OPRK1 3xK and 5xK variants contain additional positive charges in the N-terminal soluble domain, as described in Figure 3.6B. (Right) Translations were performed in RRL in the presence of either 1 μM Apratoxin A in 0.5% DMSO or a 0.5% DMSO control and in hRMs derived from WT or TMCO1 KO cells. Insertion was determined by the fraction of glycosylated substrate (OPRK1 variants and VAMP2) or by signal sequence cleavage (Prolactin). In vitro insertion reactions tend to result in smaller phenotypes than those observed in cells, in part due to the absence of competing off-pathway reactions, including protein quality control and surveillance. Additionally, the time scales of an in vitro insertion reaction are beyond the physiologic time scale in cells, which gives substrates near-infinite time to be inserted into the ER microsomes.

**Table S3.1. CryoEM Data Collection, Refinement, and Validation Statistics**

| | Full-length BOS (EMDB-) | Truncated BOS (EMDB-) (PDB ) | BOS-EMC (EMDB-) (PDB ) |
|---|---|---|---|
| **Data Collection and Processing** | | | |
| Microcope | FEI Titan Krios | FEI Titan Krios | FEI Titan Krios |
| Voltage (kV) | 300 | 300 | 300 |
| Camera | Gatan K3 | Gatan K3 | Gatan K3 |
| Magnification (nominal) | 105,000 | 105,000 | 105,000 |
| Defocus range (μm) | -1.0 to -3.0 | -1.0 to -3.0 | -1.0 to -3.0 |
| Calibrated pixel size (Å/pix) | 0.416 | 0.416 | 0.416 |
| Electron exposure ($e^-/Å^2$/frame) | 60 | 60 | 60 |
| Number of frames per movie | 40 | 40 | 40 |
| Automation software | SerialEM | SerialEM | SerialEM |
| Number of micrographs | 11,870 | 15,929 | 17,978 |
| Initial particle images (no.) | 814,566 | 1,900,000 | 3,100,000 |
| Final particle images (no.) | 63,018 | 115,841 | 69,845 |
| Local resolution range (Å) | 7.8-12.6 | 3.1-8.8 | 3.2-9.3 |
| Map resolution (Å, FSC=0.143) | 8.85 | 3.65 | 5.15 |
| **Refinement** | | | |
| Software (*phenix.real_space_refine*) | | PHENIX 1.20.1-4487 | PHENIX 1.20.1-4487 |
| Initial model used (PDB code) | | AlphaFold | 8S9S+AlphaFold |
| Correlation coefficient ($CC_{mask}$) | | 0.83 | 0.64 |
| Map sharpening *B* factor ($Å^2$) | | -60 | |
| Model composition | | | |
| Non-hydrogen atoms | | 6848 | 24652 |
| Protein residues | | 896 | 3158 |
| Ligands | | NAG:1 | NAG:8 |
| *B* factors ($Å^2$) | | min/max/mean | |
| Protein | | 26.37/170.42/82.29 | 21.51/220.95/106.41 |
| Ligand | | 82.97/82.87/82.87 | 68.85/179.72/113.94 |
| R.M.S deviations | | | |
| Bond lengths (Å) (# > 4σ) | | 0.003 | 0.003 |
| Bond angles (°) (# > 4σ) | | 0.608 | 0.592 |
| **Validation** | | | |
| MolProbity score | | 2.10 | 2.78 |
| Clashscore | | 11.43 | 24.21 |
| Rotamer outliers (%) | | 0.29 | 5.83 |
| Cß outliers (%) | | 0.00 | 0.00 |
| CaBLAM outliers (%) | | 5.41 | 2.32 |
| Ramachandran plot | | | |
| Favored (%) | | 90.93 | 95.31 |
| Allowed (%) | | 9.07 | 4.69 |
| Disallowed (%) | | 0.00 | 0.00 |

## 3.8 Materials and Methods

*Plasmids and antibodies*

The sequences used in cell-based assays and structural analysis were derived from UniProtKB/Swiss-Prot. These include: SEC61β (SEC61B; NP_006799.1), squalene synthase isoform 1 (SQS/FDFT1; Q6IAX1), vesicle associated membrane protein 2 (VAMP2; P63027-1), type-2 angiotensin II receptor (AGTR2; P50052), SARS-CoV-2 Membrane protein (VME1_SARS2; P0DTC5), SARS-CoV-2 ORF3a (AP3A_SARS2; P0DTC3), kappa-type opioid receptor (OPRK1; P41145), alpha-1A adrenergic receptor (ADRA1A; P35348),
translocating chain-associated membrane protein 2 (TRAM2; Q15035), excitatory amino acid transporter 1 (SLC1A3/EAAT1; P43003), guided entry of tail-anchored proteins factor CAMLG (GET2/CAMLG; P49069), Yip1 domain family member 1 (YIPF1; Q9Y548), Asialoglycoprotein receptor 1 (ASGR1; P07306), trace amine-associated receptor 5 (TAAR5; O14804); melatonin receptor type 1A (MTR1A/ MTNR1A; P48039), beta-1 adrenergic receptor (ADRB1; P08588), cholecystokinin receptor type A (CCKAR; P32238), adrenocorticotropic hormone receptor (MC2R/ACTHR; Q01718), melatonin receptor type 1B (MTNR1B/MTR1B; P49286), apelin receptor (APLNR/APJ; P35414), muscarinic acetylcholine receptor M1 (CHRM1/ACM1; P11229), G-protein coupled receptor 15 (GPR15; P49685), tachykinin receptor 1 (TACR1/NK1R;), translocon-associated protein subunit alpha (SSR1/SSRA/TRAPA; P43307), ER membrane protein complex subunit 7 (EMC7; Q9NPA0), ER membrane protein complex subunit 5 (EMC5/MMGT1; Q8N4V1), ER membrane protein complex subunit 2 (EMC2; Q15006), Nicalin (NCLN; Q969V3), transmembrane protein 147 (TMEM147; Q9BVK8), Nodal modulator (NOMO2; Q5JPE7), and mannosyl-oligosaccharide 1,2-alpha-mannosidase IA (MAN1A1; P33908).

The 2nd generation lenti-viral packaging plasmid psPAX2 (Addgene plasmid #12260) was a gift from Didier Trono. The 2nd generation lenti-viral packaging plasmid pCMV-VSV-G was a gift from Bob Weinberg (Addgene plasmid # 8454). The pHAGE2 lenti-viral transfer plasmid was a gift from Magnus A. Hoffmann and Pamela Bjorkman. For inducible expression in K562 cells during CRISPRi screens, the SFFV-tet3G backbone was used (Jost et al., 2017). Though mCherry and EGFP variants were used throughout the study, they are referred to as RFP and GFP, respectively, for clarity. The GFP:RFP reporter system for reporter assays was used as previously described to assess substrate insertion (Guna et al., 2022; Inglis et al., 2020).

For expression in K562 cells during genome-wide CRISPRi screens, AGTR2 and SARS-CoV-2 ORF3a were cloned as N-terminal fusions to GFP, followed by a viral 2A skipping sequence, and RFP. For SARS-CoV-2 M, the reporter was designed using the split GFP system (Cabantous et al., 2005; Kamiyama, Sekine, Barsi-Rhyne, et al., 2016). Here, the GFP11 tag (RDHMVLHEYVNAAGIT) was inserted at the N-terminal separated by a 3X-GS linker to allow for complementation with GFP1-10. The M protein was designed as a split GFP reporter while AGTR2 and ORF3a were designed to contain full GFP fusions. The latter two substrates are unstable in cells with the additional length of GFP11 fused to the N-termini. Additionally, we note that in the arrayed screen in Figure 3.2, all substrates except for SEC61β contain full GFP or RFP fusions. The SEC61β reporter has been previously described (Guna, Page, et al., 2023). Briefly, the TMD and flanking regions were inserted downstream of the first 70 residues of the flexible cytosolic domain of SEC61β. At the C-terminal of Sec61β, the GFP11 tag (RDHMVLHEYVNAAGIT) was inserted, separated by a 2X-GS linker. To express GFP1-10 in the ER lumen, the human calreticulin signal sequence was appended to the N-terminal of GFP1-10-KDEL as previously described (Guna et al., 2022; Pleiner et al., 2020).

Programmed dual sgRNA guide plasmids were used in assays involving depletion of two genes (Replogle et al., 2020). The following sgRNA protospacer sequences were used to generate dual guide plasmids:

| **Gene** | **sgRNA protospacer sequence (5'-3')** |
|---|---|
| Non-targeting control | GACGACTAGTTAGGCGTGTA |
| EMC2 | GGAGTACGCGTCCGGGCCAA |
| GET2 | GATGTTGGCCGCCGCTGCGA |
| FAF2 | GGCTCCAGCGGCTCCACATC |
| NCLN | GAGCTGGGACGGGACGGCGG |
| TMCO1 | GTCGGTGGTGGCGGAGTCTA |
| CCDC47 | GGGTTCTGGCTGCGTCCTAG |
| TRAPB/SSRB | GTGGCTCCGGAAAGGCGTTT |
| TRAPA/SSRA | GTTCCGTCGGCTAAGGCTCT |
| UGGT1 | GGCGCGTGTCGGCCTCTCAC |
| SYVN1/HRD1 | GGGAGTGTTGTTAACCGGAG |

To generate knockout cell lines, the following sgRNAs were cloned into pX459 following a standard protocol: TMCO1 (GAAACAATAACAGAGTCAGC), CCDC47 (GTCGCATTCTACGTAACGGA), TMEM147 (ATGTCCCGGAATGCCGGCAA).

The following siRNAs were used in this study: negative control no. 2 siRNA (#4390846), EMC5 siRNA s41131, NOMO siRNA s49380, CCDC47 siRNA s32577 (all Silencer Select; ThermoFisher Scientific, USA).

Constructs for expression in rabbit reticulocyte lysate (RRL) were based on the SP64 vector (Promega). For *in vitro* translations, the ALFA epitope (PSRLEEELRRRLTEP) was appended to the C-terminal of SARS-CoV-2 ORF3a and M proteins,, separated by a flexible 3X-GS linker (Götzke et al., 2019).

All plasmids are available upon request.

The following antibodies were used in this study: EMC2 (25443-1-AP, Proteintech, USA); EMC3 (67205-1-Ig, Proteintech, USA); EMC4 (27708-1-AP, Proteintech, USA); EMC5 (A305-833, Bethyl Laboratories, USA); EMC7 (27550-1-AP, Proteintech, USA); EMC10 (ab180148, Abcam, UK); TMCO1 (27757-1-AP, Proteintech, USA); TMEM147 (PA5-95876, Invitrogen, USA); NCLN (A305-623A-M, Bethyl, USA); NOMO1 (PA5-47534, Invitrogen, USA); NOMO2 (PA5-95819, Invitrogen, USA); CCDC47 (A305-100A-M, Bethyl, USA); GET2 (Synaptic Systems, Germany); SYVN1 (13473-1-AP, Proteintech, USA); FAF2 (16251-1-AP, Proteintech, USA); UGGT1 (sc-374565, Santa Cruz Biotechnology, USA); Asterix (HPA06568S, Sigma Aldrich); Calnexin (10427-2-AP, Proteintech); BIP (#610979, BD Biosciences); PDI (ADI-SPA-891-D, Enzo, USA); alpha tubulin (T9026, Millipore Sigma, USA); Anti-HA-HRP (Millipore-Sigma, USA); Anti-FLAG-HRP (Millipore-Sigma, USA). The rabbit polyclonal antibodies against BAG6, TRAPA, and GFP were gifts from Ramanujan Hegde. Secondary antibodies used for Western blotting were goat anti-mouse- and anti-rabbit-HRP (#172-1011 and #170-6515, Bio-Rad, USA).

*Cell culture and cell line construction*

| **Cell Lines** | **Source** |
|---|---|
| K562 KRAB-BFP-dCAS9 | Gilbert et al., 2014 |
| HEK 293T/17 | ATCC CRL-11268 |
| K562-CRISPRi-Tet-ON-((ER)-GFP1-10)-(tet-RFP-P2A-SEC61β-GFP11) | Guna et al., 2023 |
| K562-CRISPRi-Tet-ON-(tet-AGTR2-GFP-P2A-RFP) | This study |
| K562-CRISPRi-Tet-ON-(tet-SARS-CoV-2_ORF3a-GFP-P2A-RFP) | This study |
| K562-CRISPRi-Tet-ON-(tet-SARS-CoV-2_ORF3a-GFP-P2A-RFP) | This study |
| HEK 293T TMCO1 KO | This study |
| HEK 293T TMEM147 KO | This study |

| | |
|---|---|
| HEK 293T CCDC47 KO | This study |
| 293T TREX Flp-In GFP-EMC2-P2A-RFP | Pleiner et al., 2020 |
| 293T TREX Flp-In GFP-EMC2-P2A-RFP + EMC7 KO | Pleiner et al., 2023 |
| 293T TREX Flp-In GFP-EMC2-P2A-RFP + hsEMC5-ALFA-2A-TagBFP-3xFLAG | This study |
| 293T TREX Flp-In GFP-EMC2-P2A-RFP + Prl(ss)-hsEMC7(24-end)-ALFA-P2A-TagBFP-3xFLAG | This study |
| 293T TREX Flp-In HEK 293T GFP-EMC2-P2A-RFP + hsNCLN-ALFA-P2A-TagBFP-3xFLAG | This study |
| 293T TREX Flp-In GFP-EMC2-P2A-RFP + hsTMEM147-ALFA-P2A-TagBFP-3xFLAG | This study |
| RPE1 KRAB-BFP-dCAS9 | Jost et al., 2017 |
| Expi293 TMEM147-TEV-GFP-P2A-RFP | This study |
| Expi293 TMEM147-GFP | This study |
| Expi293 NCLN-P2A-RFP | This study |
| Expi293 NOMO2-P2A-TagBFP | This study |
| Expi293 TMEM147-TEV-GFP-40aa-ALFA-EMC2 | This study |

K562 cells containing KRAB-BFP-dCas9 (Gilbert et al., 2014) were cultured in RPMI-1640 with 25 mM HEPES, 2.0 g/L NaHCO3, and 0.3 g/L L-glutamine supplemented with 10% Tet System Approved FBS, 2 mM glutamine, 100 units/mL penicillin, and 100 µg/mL streptomycin. K562 cells were maintained between $0.25 \times 10^6$ –$1 \times 10^6$ cells/mL. HEK 293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 100 units/mL penicillin, 100 µg/mL streptomycin, and 10% FBS. RPE1 cells containing the KRAB-BFP-dCas9 machinery (Jost et al., 2017) were cultured in DMEM F12 medium supplemented with 10% FBS and 2 mM glutamine. K562, HEK 293T, and RPE1 cells were grown at 37 °C. Expi293 cells were cultured in Expi293 Expression Medium (Gibco) supplemented with 10% FBS and 2 mM glutamine. Expi293 cells were maintained between $0.5 \times 10^6$ – $2 \times 10^6$ cells/mL and harvested at $6 \times 10^6$ cells/mL.

Clonal knockouts of TMCO1, CCDC47 and TMEM147 were obtained by transfecting HEK 293T cells with pX459 encoding the respective sgRNA using TransIT-293 transfection reagent (Mirus, USA). 72 h post transfection, single cells were sorted into 96-well plates using a SONY cell sorter (SH800S), and clones were selected following verification of protein depletion by Western blotting.

*Fluorescent reporter CRISPRi screens*

CRISPRi screens were performed as previously described, with minor modifications (Gilbert et al., 2014; Horlbeck et al., 2016). For AGTR2 and Sec61β, screens were performed using either the Non-targeting-dual guide library or the EMC2-dual library. For SARS-CoV-2 M and ORF3a,

screens were performed with the single CRISPRi-v2 library. We have previously demonstrated that the additional non-targeting guide in the dual guide cassette does not appreciably alter knockdown efficiency of the second guide in the cassette (Guna, Page, et al., 2023). CRISPRi libraries (single CRISPRi-v2 library, Non-targeting dual library [Addgene #197348], or EMC2 dual library [Addgene #197349]) were transduced at a multiplicity of infection less than one into 300-330 million K562-CRISPRi-Tet-ON cells containing the appropriate reporter. For the duration of the screens, cells were maintained in 1L spinner flasks (Bellco, SKU: 1965-61010) at a volume of 1L. 48 hours post-transfection, BFP positive cells were between 30-40%. After 48h, cells were treated with 1 µg/mL puromycin for three days to select for guide positive cells. Following approximately two days of recovery after puromycin selection, the reporter was induced with doxycycline (100-1000 ng/mL) for 24-48 hours and sorted on a FACSAria Fusion Cell Sorter. Cells were diluted to $0.5 \times 10^6$ cells/mL each day to ensure that the culture was maintained at an average coverage of more than 1000 per sgRNA.

For sorting, cells were gated for BFP (selecting guide-positive cells), RFP and GFP (selecting an expressing reporter). Cells were sorted based on the GFP:RFP ratio of the final gated population. Approximately 30 million cells with either the highest or lowest 30% GFP:RFP ratios were collected during sorting, pelleted, and flash-frozen. From cell pellets, genomic DNA was extracted and purified using a Nucleospin Blood XL kit (Takara Bio, #740950.10). The guides were amplified and barcoded by PCR using NEB Next Ultra ii Q5 MM (M0544L). For both single and dual guide CRISPRi screens, a unique forward index primer was used. For single guide CRISPRi screens, a reverse primer that binds downstream of the guide was used (5'-CAAGCAGAAGACGGCATACGAGATCGACTCGGTGCCACTTTTTC). For dual guide CRISPRi screens, a reverse primer that binds in the hU6 region upstream of the fixed guide was used (5'-CAAGCAGAAGACGGCATACGAGATGGAATCATGGGAAATAGGCCCTC), as previously described (Guna et al., 2023). SPRISelect beads (Beckman Coulter B23317) were used to purify the DNA library (279 or 349 bp), and purified DNA was analyzed on an Agilent 2100 Bioanalyzer prior to sequencing using an Illumina HiSeq2500 using the standard CRISPRi-v2 library sequencing primer (5'-GTGTGTTTTGAGACTATAAGTATCCCTTGGAGAACCACCTTGTTG). Analysis of the sequencing was performed using the pipeline in https://github.com/mhorlbeck/ScreenProcessing (Horlbeck et al., 2016). To ensure coverage, guides with fewer than 50 counts were excluded from analysis. The phenotype score for each gene was calculated from the strongest 3 sgRNA

phenotypes. The Mann-Whitney p-value was calculated using the 5 sgRNAs targeting the same gene compared to the negative controls. For screens that were performed in biological duplicate (SARS-CoV-2 M, SARS-CoV-2 ORF3a, and Sec61β), the sgRNA phenotypes were averaged. To calculate the discriminant scores used in Fig. 3, each gene's phenotype score was multiplied by its Mann-Whitney p-value.

*Lentiviral Transduction*

Lentivirus was generated via co-transfection of a transfer plasmid of interest along with packaging plasmid psPAX2 and envelope plasmid VSV.G, using TransIT-293 transfection reagent (Mirus). Lentivirus was harvested 48 hours after transfection, then aliquoted, flash-frozen, and stored for future usage.

Cell lines for structural analysis were generated in Expi293 cells. Suspension cells were transduced by mixing 10 million cells with 2.5 ml of harvested lentiviral supernatant in presence of 8 μg/ml polybrene in a final volume of 30 ml in a 125-ml vented flask. For BOS (tNOMO) cell line, 2.5 ml of lentiviral supernatant of each subunit TMEM147-GFP, NCLN-P2A-RFP, NOMO(Δ1-9Ig)-P2A-TagBFP were added during transduction. The cells were grown in a shaking incubator for ~16 hours before being pelleted and resuspended in 50 ml of fresh medium in order to remove lentiviral particles. Then the cells were continued to be grown for about a week until transduced cells expressing plasmid of interest were sorted with the Sony SH800S cell sorter (Sony Biotechnology).

K562 dCas9-BFP-KRAB cells were spinfected with lentivirus containing dual sgRNAs targeting two genes of interest or a non-targeting control. Briefly, 250,000 cells were mixed with 200 μl of lentivirus and RPMI medium in the presence of 8 μg/ml polybrene in a total volume of 2 ml in a 12-well plate. Plates containing K562 cells were spun at 1,000 g for 1.5-2 h at 30˚C, resuspended, and cultured in 12-well plates. Approximately 48 h after spinfection, 1 μg/ml puromycin was added for 5 consecutive days to select cells containing the dual guide cassette. To assess the percentage of guide-containing, BFP-positive cells, samples were analyzed using flow cytometry, as described below. After a total of 8 days of knockdown, cells were pelleted, flash frozen, and used in western blot analysis to assess knockdown of individual genes.

For reporter assays in adherent HEK293 or RPE1 cells, lenti-viral transduction of 50-300 μl lentiviral supernatant and 8 μg/ml polybrene (Millipore-Sigma, USA) were added to ˜70%

confluent cells in 2.5 ml culture medium in a 6-well. Lenti-viral reporter constructs of all GPCRs, TRAM2, SARS-CoV-2 M, SARS-CoV-2 ORF3a, EAAT1, GET2, YIPF1 for use in HEK 293T and RPE1 cells, contained an upstream CMV promoter followed by the protein fused to GFP, a 2A site, and RFP (Figures 3.2B,C , 3.3D-F, 3.4C, 3.6A, 3.6C-F, S3.6A, C-F, S3.3C-D, S3.7A). Versions of GPCRs with a signal sequence (Figures 3.4C, 3.6C, and S3.6C) contained N-terminal fusions of the pre-prolactin signal sequence (KGSSQKGSRLLLLLVVSNLLLCQGVV) followed by a T4 Lysozyme soluble domain. In parallel, the first TMD (residues 33-75) of MAN1A1, a membrane protein with $N_{cyt}$ topology, was fused to the N-terminus of GPCRs. Both signal sequence-T4 lysozyme fusions and MAN1A1 fusions behaved similarly (Figure 3.4C and S3.3D). Sec61β, SQS, VAMP2, and ASGR1 lenti-viral reporters for use in HEK 293T and RPE1 cells contained an upstream CMV promoter, followed by GFP, a 2A site and RFP, which was fused to the reporter. The TMD and flanking regions of Sec61β, SQS, and VAMP2 the were fused directly to RFP, as described before (Guna et al., 2018; Guna, Page, et al., 2023; Pleiner et al., 2020). A charge mutant of OPRK1 (E45K, D46R, E50K) (+0 variant) was used in RPE1 cells in the same GFP-2A-RFP cassette as described above for GPCRs and as previously described (Pleiner et al., 2023) (Figures 3.6C-F, S3.6C, S3.7A).

For CRISPRi knockdown experiments in RPE1 cells, cells were transduced with sgRNA dual guide lenti-viral vectors. After 6 days of knockdown, cells were transduced with fluorescent reporter lenti-viral vectors described above and analyzed ~48h post-transduction (8 days after transduction with guide).

For rescue assay experiments, 300,000 HEK 293T and HEK 293T EMC7 KO cells were seeded into each 6-well plate on Day 1. On Day 2, cells were transduced with 300 µl lentiviral supernatant of rescue construct(s) and 8 µg/ml final concentration of polybrene, marking the start of the 72-hour rescue lentivirus addition. The media was exchanged on Day 3 to remove excess polybrene, and the 48-hour reporter lentivirus addition started by transducing the cells with 150 µl lentiviral supernatant of reporter construct(s) in presence of 8 µg/ml final concentration of polybrene. On Day 4, the cells were split 1:2 into a different set of 6-well plates to be used for Western Blot. Lastly, on Day 5, the cells were harvested, washed and resuspended in 500 µl Dulbecco's Phosphate Buffered Saline (Gibco) to be analyzed by flow cytometry or frozen for analysis via Western Blot.

*Flow cytometry*

RPE1 and HEK 293T cells were trypsinized, washed with 1xPBS, and resuspended in 1xPBS for flow cytometry analysis. K562 cells were analyzed directly from 12-well or 6-well cultures. Cells were analyzed using 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) or by Python using the FlowCytometryTools package.

The Sec61 inhibitor Apratoxin A was used to analyze the effect of SEC61 inhibition on membrane protein insertion (Paatero et al., 2016; Thornburg et al., 2013). HEK 293T (WT or TMCO1 KO) cells were transduced with reporter lenti-virus, and 48h later, cells were treated with 31.3 nM Apratoxin A in 0.1% DMSO for 12h. Cells were analyzed immediately following treatment with inhibitor. Apratoxin A was a gift from Ville Paavilainen.

*Preparation of human ER microsomes*

Human derived rough ER microsomes were generated as previously described, with minor modifications (Chitwood et al., 2018). HEK293T cells (WT, NCLN KO, TMCO1 KO, or EMC6 KO) were harvested and washed in 1X PBS. Cells were resuspended in 4 times the pellet volume of sucrose buffer (10 mM HEPES, pH 7.5, 250 mM sucrose, 2 mM magnesium acetate, 1X cOmplete EDTA-free protease inhibitor cocktail [Roche]) and lysed by douncing at 4 °C. Lysed cells were diluted 2X in sucrose buffer and pelleted at 3214 xg for 35 min. at 4 °C. Supernatant was transferred to a new tube and pelleted again at 3214 xg for 35 min. at 4 °C. To isolate the microsomal fraction, samples were pelleted in an ultracentrifuge in an MLA80 rotor (Beckman-Coulter) at 75,000 xg for 1h at 4 °C. Supernatant was removed, and the microsomal pellet was resuspended to an A280 of 75 in microsome buffer (10 mM HEPES, pH 7.5, 250 mM sucrose, 1 mM magnesium acetate, 0.5 mM DTT). To remove contaminating RNAs, microsomes (hRMs) were nucleased. $CaCl_2$ (1 mM) and micrococcal nuclease (0.125 U/µL) were added to hRMs and mixed before incubating for 6 minutes at 25 °C. To quench the reaction, EGTA (2 mM) was added to the sample and the sample was immediately mixed and place on ice. Nucleased hRMs were flash frozen and stored at -80 °C prior to use in *in vitro* translations.

*Mammalian in vitro translation*

Translation extracts were prepared using nucleased rabbit reticulocyte lysate (RRL) supplemented with human derived rough ER microsomes, as previously described (Sharma et al., 2010; Walter

& Blobel, 1983). DNA templates for *in vitro* transcription were made by PCR from SP64-based plasmids or directly from double-stranded DNA gene fragments (IDT or Twist Biosciences) using primers within the SP6 promoter (5' end) and following a stop codon and short untranslated region (3' end). Run-off transcription reactions were made by combining 4.8 µL T1 mix (Sharma et al., 2010), 0.1 µL RNasin (Promega), 0.1 µL SP6 polymerase (New England Biolabs) and 50 ng PCR product. Reactions were incubated at 37 °C for 2 hours, and then used directly in translation reactions, which were incubated for 20-45 minutes at 32 °C. To label nascent proteins, radioactive $^{35}$S-methionine (Perkin Elmer) was included in translation reactions, unless otherwise indicated. Samples were then analyzed directly using SDS-PAGE and autoradiography.

For experiments in which the insertion of the first TMD was assessed, substrates were translated in the presence of hRMs derived from HEK 293T cells. The OPRK1 constructs (wildtype or variants with 3xK or 5xK mutations in the N-terminal soluble domain) contain an Asn-Gly-Thr (NGT) glycosylation site at the N-terminus, which allows monitoring of insertion. The VAMP2 control protein contains a C-terminal Opsin tag that gets glycosylated upon insertion of the TA substrate and allows monitoring of insertion. A construct containing the first 85 amino acids of preprolactin was used as a control for signal sequence cleavage. For assays in which Sec61's insertion capacity was assessed, the inhibitor Apratoxin A was used at 1 µM.

*Preparation of the ALFA nanobody conjugated to HRP for Western blotting*

The ALFA nanobody was coupled to HRP-maleimide through a single engineered C-terminal cysteine residue, as previously described (Pleiner et al., 2018).

*DSP crosslinking*

Suspension adapted T-REx-293 cells stably expressing either GFP-EMC2 only or GFP-EMC2 plus EMC5-ALFA, EMC7-ALFA, TMEM147-ALFA, or NCLN-ALFA were harvested, washed in PBS, pelleted, and resuspended in PBS containing 1.5 mM final concentration of dithiobis(succinimidyl propionate) (DSP; Thermo Scientific). The cell mixture was incubated at 4°C with head-over-tail rotation for 2 hours. After the incubation, the reaction was quenched by addition of 1M Tris/HCl, pH 7.5 to 20 mM final concentration and incubated for 15 min. Then, the cells were pelleted, weighed, and flash frozen for storage prior to immunoprecipitation or prepared for mass spectrometry, as described below.

LC-MS/MS analysis for the IP experiment was performed with an EASY-nLC 1200 (ThermoFisher Scientific, San Jose, CA) coupled to a Q Exactive HF hybrid quadrupole-Orbitrap mass spectrometer (ThermoFisher Scientific, San Jose, CA). Peptides were separated on an Aurora UHPLC Column (25 cm × 75 µm, 1.6 µm C18, AUR2-25075C18A, Ion Opticks) with a flow rate of 0.35 µL/min for a total duration of 75 min and ionized at 1.6 kV in the positive ion mode. The gradient was composed of 6% solvent B (3.5 min), 6-25% B (42 min), 25-40% B (14.5 min), and 40–98% B (15 min); solvent A: 2% ACN and 0.2% formic acid in water; solvent B: 80% ACN and 0.2% formic acid. MS1 scans were acquired at the resolution of 60,000 from 375 to 1500 m/z, AGC target 3e6, and maximum injection time 15 ms. The 12 most abundant ions in MS2 scans were acquired at a resolution of 30,000, AGC target 1e5, maximum injection time 60 ms, and normalized collision energy of 28. Dynamic exclusion was set to 30 s and ions with charge +1, +7, +8 and >+8 were excluded. The temperature of ion transfer tube was 275°C and the S-lens RF level was set to 60. MS2 fragmentation spectra were searched with Proteome Discoverer SEQUEST (version 2.5, Thermo Scientific) against *in silico* tryptic digested the UniProt Human proteome Swiss-Prot database (UP000005640). The maximum missed cleavages were set to 2. Dynamic modifications were set to oxidation on methionine (M, +15.995 Da), deamidation on asparagine and glutamine (N and Q, +0.984 Da) and protein N-terminal acetylation (+42.011 Da). Carbamidomethylation on cysteine residues (C, +57.021 Da) was set as a fixed modification. The maximum parental mass error was set to 10 ppm, and the MS2 mass tolerance was set to 0.03 Da. Intensity-based quantification (iBAQ) was performed using the IMP-apQuant PD node (Doblmann et al., 2019; Schwanhäusser et al., 2013). The maximum false peptide discovery rate was specified as 0.01 using the Percolator Node validated by q-value.

*Protein purification for structure determination*

2 L of Expi293 cells stably expressing the protein(s) of interest by lentiviral transduction were pelleted, washed with PBS and flash-frozen for storage. For BOS (fNOMO), a cell line was generated stably expressing TMEM147-GFP-2A-RFP. For BOS (tNOMO), a cell line stably expressing TMEM147-TEV-GFP, NCLN-RFP, NOMO(Δ1-9Ig)-BFP. For the BOS (fNOMO) • EMC holocomplex, we generated a cell line stably expressing TMEM147-5aa-TEV-GFP-40aa(ALFA)-EMC2. All protein complexes were purified using an anti-GFP nanobody as described previously (Pleiner et al., 2020; Stevens et al., 2023). Briefly, cell pellets were harvested, washed with 1xPBS, and resuspended in solubilization buffer (50 mM HEPES/KOH pH 7.5, 200 mM NaCl, 2 mM $MgAc_2$, 1x cOmplete™ EDTA-free Protease Inhibitor Cocktail [Roche], 1%

[w/v] glyco-diosgenin [GDN; Anatrace], 1 mM DTT) at a ratio 6.8 ml solubilization buffer per 1 g cell pellet. Following incubation for 1 hour at 4°C, lysate supernatant was isolated by centrifugation at 18,000 rpm using an SS-34 rotor in a Sorvall RC6+ Superspeed Centrifuge at 4°C for 45 min.

Simultaneously, biotinylated anti-GFP nanobody was immobilized onto streptavidin magnetic beads. Specifically, 80 µl resuspended Pierce™ Streptavidin magnetic beads per 1 g cell pellet were washed and equilibrated in wash buffer (50 mM HEPES pH 7.5, 200 mM NaCl, 2 mM $MgAc_2$, 0.0053% GDN, 1 mM DTT). Then His14-Avi-$SUMO^{Eu1}$-tagged anti-GFP nanobody (Addgene #149336) was immobilized onto the washed magnetic beads for 30 min with mixing at 4°C using a ratio of 27 µg for every 80 µl beads. This immobilization was followed by incubation of beads with 50 mM HEPES/KOH pH 7.5 containing 100 µM biotin for 5 min on ice to block unbound biotin binding sites on the magnetic streptavidin beads. Subsequently, the beads were washed with solubilization buffer and incubated with clarified cell lysate for 1 hour with head-over-tail rotation. After incubation, 4 washes with wash buffer (2 volumes of wash buffer:1 volume of beads) was performed to remove unspecific binding to the beads. To elute the bound proteins, wash buffer containing 500 nM $SUMO^{Eu1}$ protease (Addgene #149333) was added to the beads and left to incubate for 30 minutes with mixing at 4°C. The eluent was further purified using size exclusion chromatography with a 3.5 ml Superose 6 column (GE Life Sciences). For BOS (tNOMO) sample, TEV protease was added (1 mg TEV protease for every 30 mg of BOS (tNOMO) protein) and incubated overnight at 4°C without mixing to remove the GFP-tag before size exclusion chromatography. The fractions corresponding to the protein complexes were concentrated using a 500-µl 30K MWCO concentrator (Millipore-Sigma).

*Grid preparation and data collection*

For BOS (fNOMO) sample, 3 µL of purified, concentrated protein at 2.48 mg/ml was applied to UltrAuFoil® R 1.2/1.3 holey gold film grid (Ted Pella, Inc.) that had been grow discharged with the PELCO easiGlow™ (Ted Pella, Inc.) at 20 mA for 60 s. The grid was blotted at 6°C, 100% humidity, -4 blot force for 4 seconds and plunged frozen in liquid ethane using the FEI Vitrobot Mark IV (Thermo Fisher Scientific). Data were collected on a Titan Krios operating at 300 keV and equipped with a Gatan K3 direct detector and a 20 eV slit width energy filter. Images were acquired using an automated acquisition pipeline in SerialEM (Mastronarde, 2005) and recorded at 105k magnification with a defocus range of -3.0 to -1.0 µm and total exposure dose of 60 $e^-/Å^2$ in

super resolution mode with a pixel size of 0.418 Å/pixel. 11,870 micrographs were collected for this data set.

For BOS (tNOMO), the grid was prepared in a similar manner, except the protein sample was concentrated to 4 mg/ml and mixed with 0.005% 3-([3-Cholamidopropyl]dimethylammonio)-2-hydroxy-1-propanesulfonate (CHAPSO; Sigma Aldrich) immediately before vitrification. 15,929 micrographs were collected for this data set.

For the BOS (fNOMO) • EMC homocomplex sample, the protein concentration was at 2.53 mg/ml and 17,978 micrographs were collected for this data set.

*Structure image processing*

The workflows for data processing of BOS (fNOMO), BOS (tNOMO), and BOS (fNOMO) • EMC are summarized in Fig. S5 A, B, C, respectively. Data processing was carried out using cryoSPARC v3.2-4.2.1 (Punjani et al., 2017). For preprocessing, micrographs were motion-corrected, Fourier-cropped twofold to 0.832 Å/pixel using 'Patch Motion Correction'; then, they were subjected to patch-based contrast transfer function (CTF) estimation with 'Patch CTF Estimation'. Movies were selected based on CTF fit cut-off of 5.0 Å in 'Curate Exposure'. From here on out, details of data processing differ for each structure.

For BOS (fNOMO), 814,566 particles were picked using 'Blob Picker' and extracted with box size = 512 pixels from 7,174 selected movies. Iterative rounds of '2D classifications' performed to remove background and junk particles. 2D classes that resemble BOS (fNOMO) complex were used as template for 'Template Picker' with particle diameter = 190 Å, which resulted in 904,456 picked particles. A round of '2D Classification' was performed to remove background particles, resulting in 246,296 particles, which were subjected to 2 more rounds of 2D classification. Then, the resulting 90,893 particles were used to generate 4 3D classes with 'Ab-Initio Reconstruction'. Using these 4 volumes, we performed 2 rounds of 3D 'Heterogeneous Refinement' on the 246,296 particles from earlier to arrive at the final EM map of BOS (fNOMO), generated from a set of 63,018 particles.

For BOS (tNOMO), 1,900,000 particles were picked using 'Blob Picker' with particle diameter of 120-320 Å and extracted from micrographs using box size = 512 pixels from 13,196 selected

movies. After iterative rounds of '2D Classification', 319,567 particles were used to generate a 3D volume using 'Ab-Initio Reconstruction'. This volume was then used for template generation for template picking, resulting in 2,500,000 picked particles. After iterative rounds of 2D classification, 1,265,788 particles were subjected to 'Ab-Initio Reconstruction' into 4 volumes. The particles that correspond to the 2 volumes that best resembled our BOS (tNOMO) complex were put through multiple rounds of 3D 'Heterogeneous Refinement' to give us 312,033 particles which were re-extracted using box size of 448 pixels. These particles were used as input in more rounds of 3D 'Heterogeneous Refinement' and '3D Classification (BETA)' to arrive at an EM map generated from 115,841 particles. This map was put through 3D 'Non-uniform Refinement' and sharpened with 'DeepEMhancer' to give us our final map.

For BOS (fNOMO) • EMC, 3,100,000 particles were picked from 17,586 selected movies using 'Blob Picker' with particle diameter of 175-450 Å, which were then extracted from micrographs 2x binned and subjected to 4 rounds of '2D Classification' to remove background particles. 300,442 particles were used to construct 3 ab-initio models with 'Ab-Initio Reconstruction' (with default settings except for maximum resolution = 7, initial resolution = 9, and initial minibatch size = 300, final minibatch size = 1000). The map that best resembled EMC with BOS was used to generate 2D templates for template particle picking ('Template Picker') with particle diameter of 300 Å, resulting in 1,932,646 picked particles. After 1 round of '2D Classification' and 'Heterogeneous Refinement' using 3 junk classes that resemble background and 1 class that resembles EMC with BOS, 595,637 particles were subjected to a first round of 'Ab-initio Reconstruction' into 2 classes (default settings except: maximum resolution = 9, initial mini batch size = 400, final mini batch size = 1200). From the 3 replicate runs of the first round of ab-intio reconstruction, the particles associated with the better 3D volume that represented BOS (fNOMO) • EMC were combined and used as input in the second round of ab-initio reconstruction into 2 classes (same settings as previous round except: maximum resolution = 6, initial resolution = 12). This process was repeated for the third time with the second rounds' particles that correspond to the better 3D volume (similar settings except: maximum resolution = 5, initial resolution = 7). For the subsequent rounds of ab-initio reconstruction, the settings were similar, except for maximum resolution = 4, initial resolution = 6. The particles were further classified using 3D heterogeneous refinement to achieve a final class of 69,845 particles. To get our final map, local refinement was performed on this map using a mask on the BOS complex generated in Chimera (Pettersen et al., 2004) with soft padding = 12, using

pose/shift gaussian prior during alignment with standard deviation of prior over rotation = 3 degree and standard deviation of prior over shifts = 1 Å.

*Model building and refinement*

For the BOS (fNOMO) structure, initial models of each subunit (TMEM147, NCLN, NOMO) were generated using AlphaFold2-Multimer ColabFold (AlphaFold2_advanced.ipynb) (Mirdita et al., 2022). Since the map quality was not sufficient for accurate model building and refinement, the initial models of each subunit were only rigid body fitted into the EM density (Fig. 5 A) and combined in COOT (Casañal et al., 2020; Emsley et al., 2010).

For the BOS (tNOMO) structure, we used the previously generated models from AlphaFold2 to rigid body fit into the density. The models were combined and manually refined in COOT (Casañal et al., 2020; Emsley et al., 2010). The final model was iteratively subjected to *phenix.real_space_refinement* (Afonine et al., 2018; Liebschner et al., 2019) with rigid body and secondary structure restraints.

For BOS (fNOMO) • EMC structure, we used BOS (tNOMO) and EMC structure (PDB: 8S9S) as initial models, which were combined, rigid body fitted and refined manually in COOT (Casañal et al., 2019; Emsley et al., 2010). The model was iteratively subjected to *phenix.real_space_refinement* (Afonine et al., 2018; Liebschner et al., 2019) with rigid body and secondary structure restraints.

CryoEM data collection, refinement, and validation statistics are reported in Table S4. Final models were evaluated with MolProbity. All figures in this study were generated with PyMOL (www.pymol.org) and ChimeraX (Goddard et al., 2018; Pettersen et al., 2021).

*Chapter 4*

# DIFFERENTIAL MODES OF ORPHAN SUBUNIT RECOGNITION FOR THE WRB/CAML COMPLEX

The following chapter is adapted from Inglis et al., 2020 and modified according to the Caltech Thesis format.

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

## 4.1 Abstract

A large proportion of membrane proteins must be assembled into oligomeric complexes for function. How this process occurs is poorly understood, but it is clear that complex assembly must be tightly regulated to avoid accumulation of orphan subunits with potential cytotoxic effects. We interrogated assembly in mammalian cells using a model system of the WRB/CAML complex: an essential insertase for tail-anchored proteins in the endoplasmic reticulum (ER). Our data suggests that the stability of each subunit is differentially regulated. In WRB's absence, CAML folds incorrectly, causing aberrant exposure of a hydrophobic transmembrane domain to the ER lumen. When present, WRB can correct the topology of CAML both in vitro and in cells. In contrast, WRB can independently fold correctly, but is still degraded in the absence of CAML. We therefore propose at least two distinct regulatory pathways for the surveillance of orphan subunits during complex assembly in the mammalian ER.

## 4.2 Introduction

A large fraction of the proteome is organized into multi-subunit complexes that must be assembled at a defined stoichiometry (Huttlin et al., 2017; Marsh & Teichmann, 2015). In the cytosol, unassembled subunits expose thermodynamically unfavorable interfaces to the crowded cellular environment, which could lead to aggregation and cytotoxic effects (M. Sung et al., 2016; Yanagitani et al., 2017). As a result, assembly of these complexes is tightly regulated to ensure that

orphan subunits, which have been synthesized in excess or cannot be assembled, are rapidly degraded to maintain cellular homeostasis (Harper & Bennett, 2016; Shemorry et al., 2013; M. Sung et al., 2016; Xu et al., 2016; Yanagitani et al., 2017). Despite increasing interest in cytosolic complex assembly, how multi-subunit membrane protein assembly is regulated within the endoplasmic reticulum (ER) membrane remains poorly understood (Dephoure et al., 2014).

Most membrane proteins are synthesized at the ER where their hydrophobic transmembrane domains (TMDs) must be inserted into the lipid bilayer, most commonly via the Sec61 insertion channel (Rapoport, 2007). A large proportion of membrane proteins must be further assembled into oligomeric complexes for function. Several lines of evidence suggest that this assembly process is highly regulated within the ER. First, orphan subunits of oligomeric membrane protein complexes are unstable and rapidly degraded by the ubiquitin-proteasome pathway (Juszkiewicz & Hegde, 2018; Lippincott-schwartz et al., 1988). Second, many membrane protein subunits require charged or polar residues for function or oligomerization, which prior to assembly would be exposed and thereby disfavored in the lipid bilayer. Finally, many TMDs situated at subunit interfaces are suboptimal and not predicted to insert autonomously, raising the question of how their insertion is coordinated with subunit assembly. Therefore, the mechanisms regulating oligomeric assembly within the ER are likely to be as defined and stringent as those in the cytosol.

Recent work demonstrates that in the cytosol, many multi-subunit complexes assemble co-translationally (Shiber et al., 2018): interaction between subunits occurs upon emergence of nascent domains from the ribosome, resulting in the temporal integration of polypeptide folding and oligomeric assembly. However, unlike in the cytosol, the steric constraints of the two-dimensional lipid bilayer, combined with the fact that the Sec61 channel is surrounded by over twenty integral membrane proteins, severely limits the space available for simultaneous insertion and oligomerization. How membrane proteins overcome these additional challenges to coordinate the folding and assembly of multiprotein complexes within the ER remains unknown.

In order to better understand membrane protein assembly and quality control in the mammalian ER, we have chosen to study the regulation of the WRB/CAML complex. WRB and CAML (Get1/2 in yeast) together form an insertase for tail-anchored proteins at the ER (Vilardi et al., 2011, 2014; Yamamoto & Sakisaka, 2012). Previous work suggests that WRB and CAML stability is interdependent, consistent with it assembling into an obligate oligomeric complex (Colombo et al.,

2016; Rivera-Monroy, Musiol, Unthan-Fechner, Farkas, Clancy, Ströbel, et al., 2016). The interaction between the two subunits is thought to be mediated by the TMDs, suggesting that it is likely an intramembrane signal that initiates a degradative pathway in the absence of the subunits' cognate binding partner (Vilardi et al., 2014; F. Wang et al., 2014; Yamamoto & Sakisaka, 2012). Despite this, the stoichiometry of the WRB/CAML complex remains to be precisely determined, as earlier work suggests CAML is in five-fold excess of WRB *in vivo*; however, no isolated populations of CAML or WRB were detected by Blue Native-PAGE analysis of mammalian cells, suggesting CAML and WRB are always found in stable oligomeric complexes (Carvalho et al., 2019; Colombo et al., 2016).

Here we report data suggesting at least two distinct mechanisms for regulation of orphan membrane protein subunits as exemplified by the WRB/CAML complex: (i) WRB is representative of a larger class of membrane subunits that insert independently but remain subject to degradation in the absence of their binding partners; (ii) in contrast CAML inserts incorrectly in the absence of WRB, aberrantly exposing a hydrophobic TMD to the ER lumen, which acts as a flag for degradation. Upon co-expression with WRB, we observe a topological change to CAML, suggesting that WRB acts as a chaperone for folding and assembly of the WRB/CAML complex. These observations set the stage for future work studying the regulation of the diversity of membrane protein subunits that must assemble at the ER.

### 4.3 Results and Discussion

#### WRB and CAML are destabilized in the absence of their binding partner

Earlier work has established that WRB and CAML expression is interdependent, though previous reports suggest that this regulation may occur partially at the transcriptional level (Carvalho et al., 2019; Colombo et al., 2016; Rivera-Monroy, Musiol, Unthan-Fechner, Farkas, Clancy, Ströbel, et al., 2016; Shing et al., 2017). We reasoned that there may be an additional layer of regulation of WRB and CAML at the post-translational level, as has been observed for other multi-subunit complexes (Beguin et al., 1998; Bonifacino et al., 1990, 1991; Dephoure et al., 2014; Lippincott-schwartz et al., 1988; Minami et al., 1987; Volkmar et al., 2019). To measure WRB and CAML stability, we utilized a fluorescent reporter system in which a single open reading frame encodes a GFP fusion of WRB or CAML followed by an RFP, separated by a viral 2A sequence (Figure 4.1A). We first demonstrated that the introduction of these fluorescent tags does not affect WRB and CAML association in HEK293T mammalian cells (Figure S4.1A). Therefore, ratiometric

analysis of GFP:RFP fluorescence using flow cytometry can be used as a proxy for subunit stability at the protein level (Itakura et al., 2016).

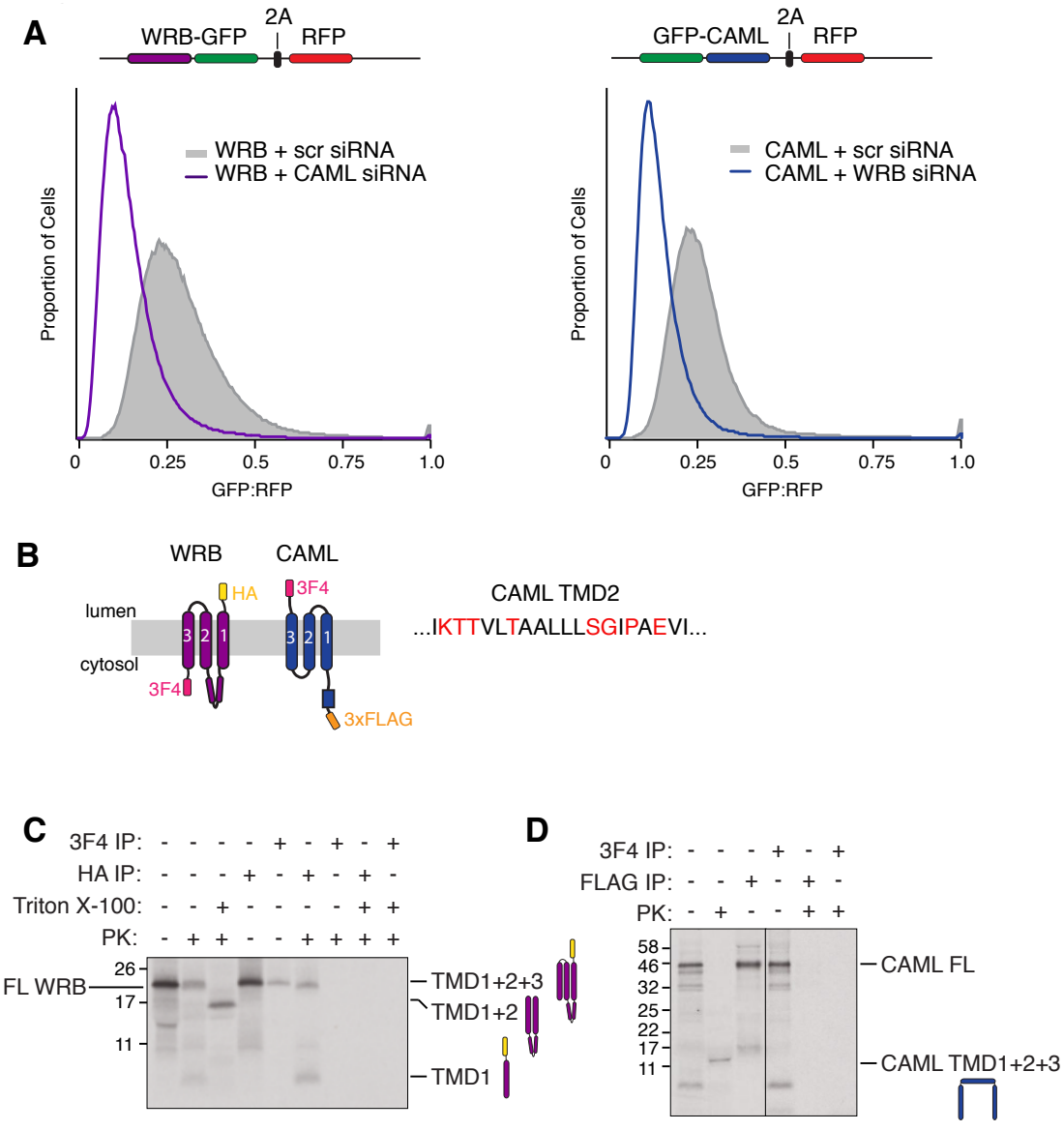



**Figure 4.1 Characterization of orphaned CAML and WRB**. (A) Histograms of CAML and WRB overexpression in HEK293T mammalian cells as determined by flow cytometry. siRNA knockdown of their respective binding partners results in a decrease in the GFP:RFP ratio for both CAML and WRB. The data shown are representative of three biological replicates. (B) Schematic depicting the expected correct topology of WRB and CAML, along with the epitope tags used for

*in vitro* translation. The sequence of the second TMD of CAML is shown, with polar, charged, and helix breaking residues highlighted. (C) $^{35}$S-methionine labeled HA-WRB-3F4 was translated in rabbit reticulocyte lysate (RRL) in the presence of canine-derived rough microsomes (cRMs). The total products were treated with proteinase K (PK) in the presence or absence of detergent and then analyzed directly or following immunoprecipitation via the 3F4 or HA tag. WRB adopts the expected topology, with the N- and C-termini in the lumen and cytosol, respectively. The coiled-coil domain between TMD1 and TMD2 partially protects the loop from cleavage by proteinase K, giving two major HA-tagged species in the absence of detergent. Upon the addition of detergent, the loop between TMD2 and TMD3 is cleaved, resulting in the loss of the HA tag. Replacing the C-terminal 3F4 tag with a larger 3xFLAG tag results in a larger change in molecular weight, consistent with the ~20 kDa band after proteinase K treatment, representing a fragment lacking the C-terminus (Figure S4.2A). See also Figure S4.2. (D) Similar to (C) but for the for FLAG-CAML-3F4 construct. There is an untagged protease protected fragment present that likely corresponds to TMDs1-3 of CAML. The lack of a protease protected 3F4 fragment demonstrates that the C-terminus of CAML remains aberrantly exposed to the cytosol. In each case, three biological replicates were performed.

Exogenous expression of either WRB or CAML individually results in rapid degradation of excess subunits, suggesting that each protein is independently unstable (approximately 65% of overexpressed WRB and 80% of overexpressed CAML that is synthesized is degraded [Figure S4.1B]). We observe a further decrease in the levels of both WRB and CAML upon siRNA knock down of their endogenous binding partner, indicating that orphaned WRB and CAML are destabilized (Figure 4.1A). Consistent with tight regulation of CAML and WRB levels by the cellular quality control machinery, we observe that overexpression of either subunit results in downregulation of the endogenous protein, and upregulation of its binding partner, as has been observed for other obligate hetero-oligomeric complexes (Figure S4.1C; (Guna et al., 2018; Juszkiewicz & Hegde, 2017).

**Two distinct mechanisms for recognition of orphan membrane subunits**

Unassembled subunits in the cytosol are recognized by quality control machinery due to the aberrant exposure of thermodynamically unfavorable subunit interfaces (Yanagitani et al., 2017). However, the biophysical properties of orphan membrane protein subunits that lead to their recognition and degradation are comparatively ill defined. We therefore tested the insertion and

topology of WRB and CAML to better understand how and why they are quality control substrates when unassembled.

We first demonstrated that our *in vitro* translation and insertion system, comprised of rabbit reticulocyte lysate (RRL) supplemented with canine ER microsomes (cRM), could recapitulate the stable assembly of WRB and CAML as observed in cells (Figure S4.1D). We then determined the topology of individually translated CAML and WRB using a protease protection assay (Figure 4.1B). WRB adopts the expected topology where all three TMDs are efficiently inserted, resulting in the positioning of the N- and C-termini in the lumen and cytosol, respectively (Figure 4.1C, Figure S4.2A).

If CAML also autonomously inserts correctly, we would expect to observe two protected fragments: an untagged fragment representing TMDs1-2 and a 3F4-tagged fragment representing TMD3. However, we do not detect any 3F4-tagged protease protected species, suggesting that the C-terminus of CAML is aberrantly localized to the cytosol (Figure 4.1D). As we do not detect a FLAG-tagged fragment, indicating that TMD1 is properly inserted with its N-terminus in the cytosol, these observations are consistent with two possible CAML topologies: (1) one where TMDs1-2 are properly inserted but TMD3 remains in the cytosol, or (2) one where TMD1 and 3 are inserted, but TMD2 is 'skipped' and remains in the ER lumen (Figure S4.2B). In order to differentiate between these two models, we initially sought to exploit a native glycosylation site in CAML's second loop, which would only be positioned in the ER lumen if TMD2 was skipped, as in model 2. However, further experiments suggest that this site is not accessible and thus cannot be used to infer the topology of CAML (data not shown). Instead, we exploited the fact that the native CAML sequence contains a single methionine residue in loop 1 (M225), which leads to incorporation of $^{35}$S-Methionine at this position (Figure S4.2B). If we make the conservative mutation M225C, the fragment remaining after protease digest would either be completely unlabeled in the case of model 1, or retain two radioactive methionine residues (in loop 2) in model 2 (Figure S4.2C). As the protected fragment of CAML M225C retains at least one $^{35}$S-methioinine and can therefore be clearly visualized, this experiment is most consistent with model 2, where the untagged protease protected fragment of CAML contains all three TMDs (Figure S4.2D). Furthermore, adding a methionine to either CAML TMD2 (S250M) or TMD3 (C284M) increases

the signal of the protected fragment, indicating both TMD2 and TMD3 are included within the protected fragment (Figure S4.2D).

Therefore, our protease protection experiments support a model where, when expressed alone, the first and third TMDs of CAML insert into the lipid bilayer, while the second TMD is aberrantly exposed to the ER lumen. TMD3 is therefore inserted in the incorrect orientation, with its C-terminus aberrantly localized to the cytosol. This is consistent with the predicted inability of the second TMD to autonomously insert due to the presence of several charged, polar, and helix breaking residues (Figure 4.1B, ΔG=1.879; (Hessa, Meindl-beinker, et al., 2007). Our biochemical evidence suggests the majority of the orphan CAML population is inserted in this manner, in contrast to previous reports in which both TMD2 and TMD3 are localized to the lumen (Carvalho et al., 2019).

We next tested whether the insertion of CAML was affected by the presence of WRB, using the appearance of the 3F4-tagged TMD3 after protease treatment as a proxy for CAML folding. Using a similar *in vitro* strategy, we observe that both co- and pre-expression of WRB resulted in increasing amounts of properly inserted CAML, as indicated by the appearance of a 3F4-tagged protease protected fragment (Figure 4.2A, Figure S4.3A). When WRB is translated prior to CAML rather than simply co-expressed, we consistently observe an increase in the levels of protected TMD3, suggesting that the timing of WRB recruitment is potentially important for CAML folding. Of note, detection of the protease protected 3F4-tagged TMD3 requires enrichment via immunoprecipitation, suggesting that at least *in vitro*, WRB does not correctly fold all of the exogenously expressed CAML.

To confirm that WRB-dependent insertion of CAML was not an artifact of the *in vitro* system, we exploited a split GFP system to determine CAML topology in cells (Figure 4.2B, Figure S4.3; (Hyun et al., 2015). We generated mammalian cell lines expressing the first ten β-strands of GFP in the ER lumen. Expression of constructs that position the eleventh β-strand of GFP in the lumen, but not in the cytosol, allow for complementation with $GFP_{1-10}$, and the resulting fluorescence can be measured by flow cytometry (Figure S4.3B and C). When $GFP_{11}$ is positioned at the C-terminus of CAML, a 5-fold increase in fluorescence is observed specifically in the presence of exogenous WRB but not another unrelated membrane protein (Figure 4.2C). This increase in GFP fluorescence upon co-expression of CAML and WRB at the ER is striking enough to be directly visualized by

fluorescence microscopy (Figure 4.2D). The low level of GFP complementation observed when CAML-GFP$_{11}$ is expressed individually may be due to partial insertion by endogenous WRB. The correct insertion of CAML's TMDs 2 and 3 is therefore dependent on association with WRB both *in vitro* and in cells. These data are consistent with recent findings that describe a WRB-dependent conformational change to CAML in cells (Carvalho et al., 2019).

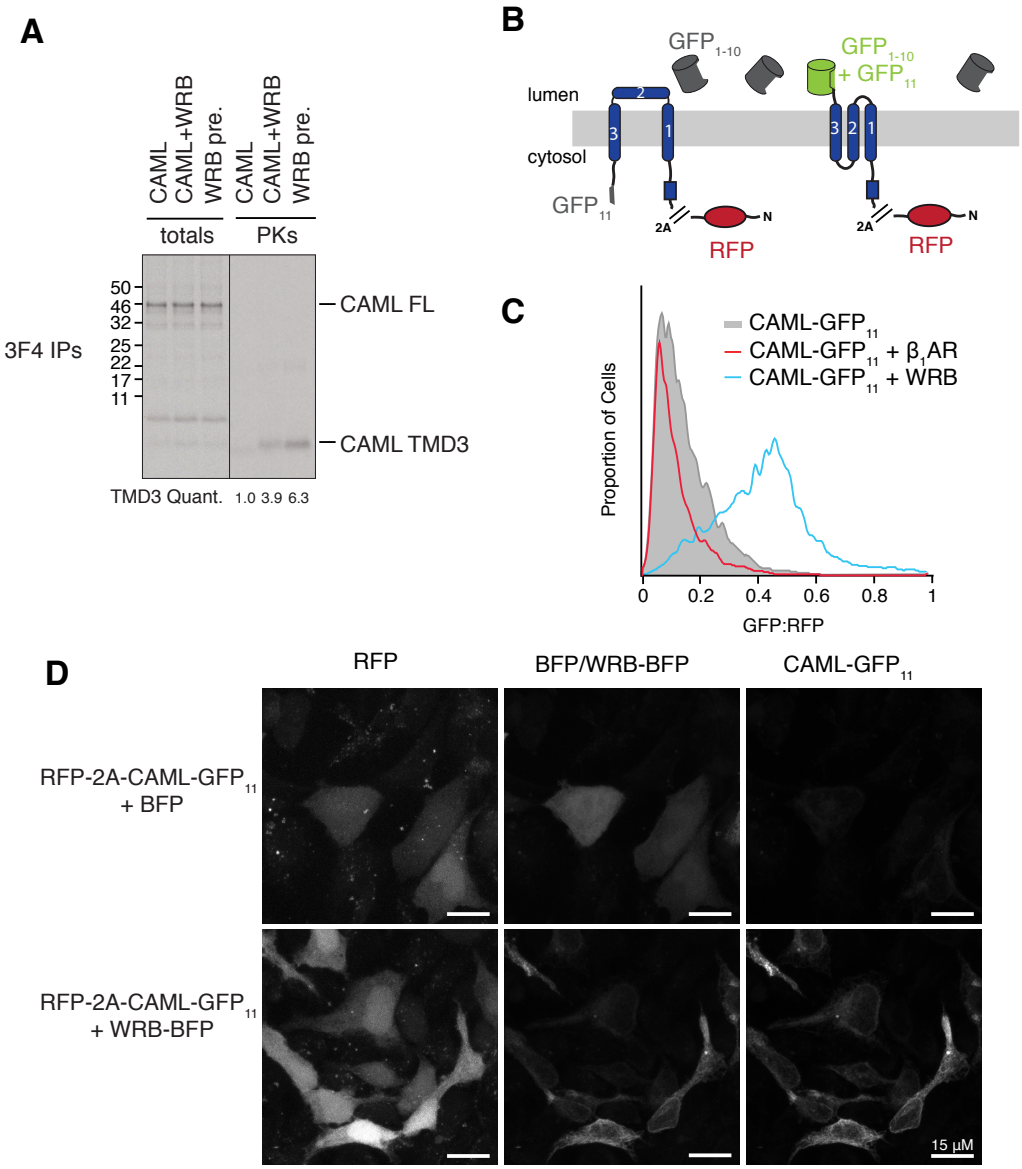


**Figure 4.2. CAML requires WRB for correct insertion.** (A) $^{35}$S-methionine labeled FLAG-CAML-3F4 was translated in RRL in the presence of cRMs either individually, alongside WRB, or with cRMs pre-loaded with WRB. Following digestion with proteinase K, total translations and digested reactions were immunoprecipitated via the 3F4 epitope tag. The positions of bands corresponding to full length (FL) CAML and CAML TMD3 are indicated. The amount of protected CAML TMD3 relative to total translated protein is indicated. Corresponding amounts of WRB present are shown in Figure S4.3A. (B) Schematic illustrating the split GFP system used to establish the topology of CAML in cells. CAML containing the eleventh β strand of GFP ($GFP_{11}$) at its C-terminus was transfected into cells stably expressing the remainder of GFP ($GFP_{1-10}$) in the ER lumen. The correct insertion of CAML TMD3 would localize $GFP_{11}$ to the ER lumen, resulting in complementation and GFP fluorescence. See also Figure S4.3. (C) Flow cytometry analysis of the system described in (B) for RFP-2A-CAML-$GFP_{11}$ expressed either alone or alongside an unrelated membrane protein ($\beta_1$AR-BFP) or WRB-BFP. (D) ER $GFP_{1-10}$ expressing cells were co-transfected with RFP-2A-CAML-$GFP_{11}$ and BFP or WRB-BFP. Fixed cells were then imaged by confocal microscopy. Three biological replicates were performed for all experiments.

Taken together, these observations suggest that there are at least two distinct mechanisms for recognition of orphan subunits at the ER. WRB, despite adopting the correct topology, is destabilized in the absence of CAML. This may be due to the presence of charged or polar residues within the TMDs that would generally be shielded at the subunit-interface with CAML. Exposure of such residues could lead to recognition of unassembled WRB by membrane-embedded quality control machinery. WRB is therefore representative of a larger class of membrane protein subunits that are properly inserted and folded, yet are degraded by the ubiquitin-proteasome pathway when unassembled (Bañó-polo et al., 2017; Lippincott-schwartz et al., 1988).

Conversely, the regulation of CAML in the absence of WRB is at least partly due to the incorrect insertion of its TMD2. Mutations that decrease the hydrophobicity of CAML TMD2 stabilize overexpressed CAML (Figure 4.3A). The effect is slight (approximately two-fold) which is consistent with mutant CAML remaining unassembled and misfolded, with TMD3 in the wrong orientation within the bilayer. Further, fusion of CAML TMD2 to an unrelated membrane protein results in its destabilization compared to fusion with a hydrophilic sequence of similar length (Figure 4.3B). Together, this suggests that exposure of CAML TMD2 to the ER lumen is both necessary and sufficient for destabilization of unassembled CAML. Aberrant exposure of this

hydrophobic segment serves as a flag for recognition, allowing CAML to recruit lumenal quality control machinery for its degradation (Feige & Hendershot, 2013).

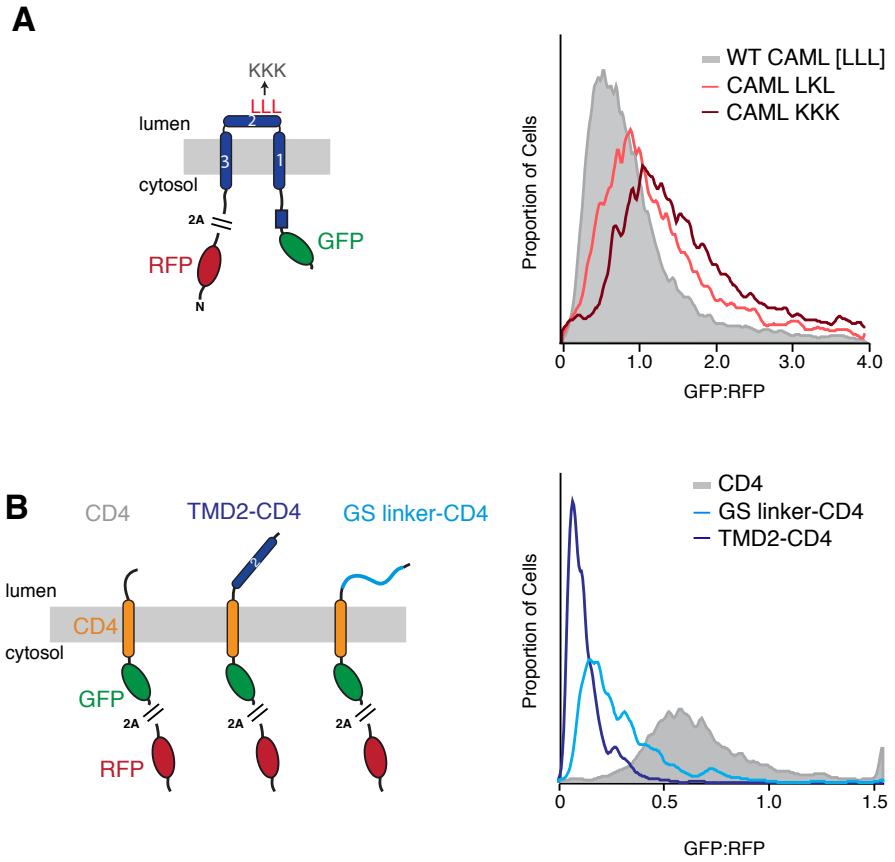



**Figure 4.3**. **Localization of CAML TMD2 to the ER lumen is both necessary and sufficient for degradation of orphan CAML**. (A) Orphan CAML degradation is contingent on the hydrophobicity of its TMD2, which is aberrantly exposed to the ER lumen. Mutation of either one (L248K) or three (L247K, L248K, and L249K) leucine residues within TMD2 has a stabilizing effect on overexpressed CAML, with the triple mutation resulting in approximately a two-fold stabilization over wild type. (B) Fusion constructs of the CD4 TMD-GFP with either CAML TMD2 or a length matched glycine-serine linker were targeted to the ER using the prolactin signal sequence. The stability of each construct was determined using flow cytometry as previously described. Two biological replicates were performed in each case.

**Reorientation of CAML TMDs 2-3 occurs after release from the ribosome**

Given the observation that WRB is required for CAML folding, the two most likely models are that the reorientation of TMDs 2 and 3 is happening (i) co-translationally during synthesis of CAML at the Sec61 translocation channel or (ii) post-translationally, after CAML has been released from the ribosome. To discriminate between these two possibilities, we first tested whether WRB can bind nascent CAML while it is still associated with the ribosome and Sec61 (Figure 4.4A). Consistent with a post-translational mechanism for insertion, we observe that CAML is able to immunoprecipitate significantly more WRB after release from the ribosome then when stalled immediately before the stop codon (Figure 4.4A). The observation that CAML cannot stably bind WRB prior to translation termination, when TMD3 is buried in the ribosomal exit tunnel, is consistent with the observation that TMDs1-2 of CAML are insufficient for stable recruitment of WRB, as truncation analysis demonstrates TMD3 (through residue 287) is necessary for the interaction (Figure 4.4B).

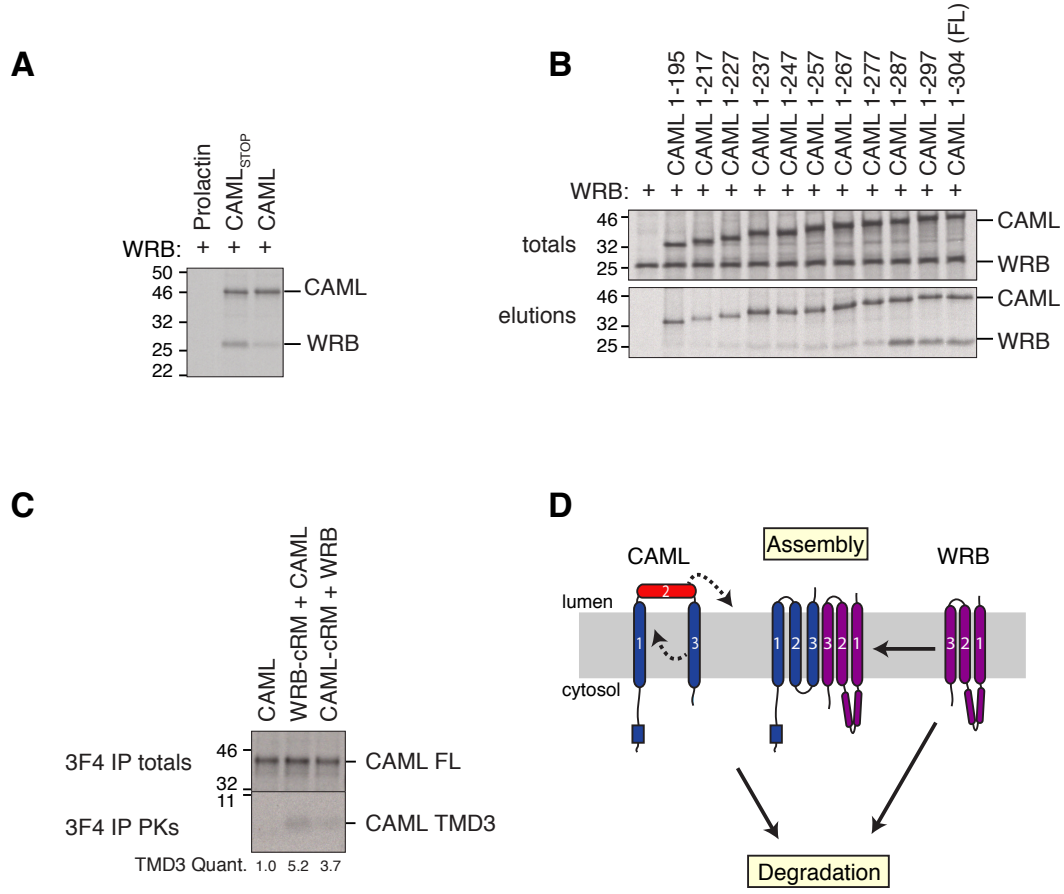


**Figure 4.4 WRB causes reorientation of CAML TMD2-3 following release from the ribosome.** (A) $^{35}$S-methionine labeled full-length FLAG-CAML with or without a stop codon (or an untagged control) was produced under conditions that maintain the peptidyl-tRNA linkage in the presence of cRMs preloaded with WRB. The membranes were solubilized, and complexes affinity purified via the FLAG tag of CAML. (B) $^{35}$S-methionine labeled CAML truncations were translated in the presence of WRB-preloaded cRMs, and the reactions treated with puromycin to release the truncated nascent chains from the ribosome. Solubilized complexes were affinity purified under native conditions via the FLAG tag of CAML. The minimal CAML truncation required to stably immunoprecipitate WRB is indicated with an asterisk. (C) cRMs were introduced during the translation of either i) no transcript, ii) CAML-3F4 or iii) WRB to produce i) empty, ii) $^{35}$S-methionine labeled WRB-preloaded membranes or iii) $^{35}$S-methionine labeled CAML-3F4-preloaded membranes. Membranes were purified before being used in a second round of translation to produce $^{35}$S-methionine labeled CAML-3F4 or WRB. Protection of CAML TMD3, as a proxy for CAML folding, was analyzed using a protease protection assay and immunoprecipitation via the 3F4 tag as described in Figure 4.2A. The amount of protected CAML TMD3 relative to total translated CAML is indicated. Three biological replicates were performed for all experiments. See also Figure S4.4. (D) A proposed model for the regulation of assembly of the WRB/CAML complex: upon initial synthesis CAML is misfolded, aberrantly localizing TMD2 to the ER lumen. The post-translational recruitment of WRB then allows CAML to insert and fold correctly. For simplicity we have depicted a single WRB/CAML heterodimeric interaction, but WRB may operate catalytically to fold multiple CAML subunits to account for the observed excess of CAML relative to WRB (Colombo et al., 2016). In the absence of WRB, TMD2 serves as a flag for degradation of orphaned CAML, which can exploit the lumenal quality control machinery for recognition and degradation. In contrast, WRB independently adopts the correct topology upon synthesis, yet is robustly degraded in the absence of CAML. Together WRB and CAML therefore represent two distinct mechanisms for stoichiometric regulation within the ER membrane.

To further explore the mechanism of CAML folding, we exploited our ability to pre-load membranes with either CAML or WRB to control the order of translation and insertion into the membrane (Figure S4.4). One would predict that if the folding of CAML must occur co-translationally, TMD3 insertion would be more efficient when WRB is translated first and thereby present throughout the synthesis of CAML. We observe a small but reproducible increase in the

amount of protected CAML TMD3 when WRB is expressed first (Figure 4.4C), consistent with the improved folding of CAML observed upon pre-loading vs. co-expression of WRB (Figure 4.2A).

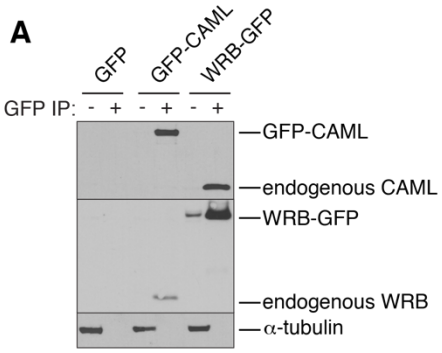
Together these experiments are more consistent with folding of CAML by WRB after translation termination and release from the ribosome, though suggest that successful reorientation of TMDs2-3 may depend on the timing of WRB recruitment to nascent CAML. However, we cannot exclude the possibility that: (i) WRB binding to CAML is initiated co-translationally, but that the interaction is too weak and/or transient to survive immunoprecipitation; or (ii) that CAML folding occurs after release from the ribosome, but in the context of the translocon, which could potentially reduce the energetic cost of reorientation of TMDs2 and 3 across the lipid bilayer.

Taken together we suggest a working model for the folding and assembly of the WRB/CAML complex (Figure 4.4D). Stable recruitment of WRB occurs after release from the ribosome, and is likely mediated by the first and third TMDs of CAML. Whether this partially folded version of CAML is stabilized by either an intramembrane and/or lumenal chaperone or remains associated with Sec61 prior to binding to WRB, remains to be determined. Similarly, unassembled WRB may also require stabilization by a membrane-embedded chaperone, to provide sufficient time for association with CAML. Upon binding, WRB is able to correctly reorient CAML into the ER membrane, thereby acting as an internal chaperone for the folding and assembly of the WRB/CAML complex. This strategy allows insertion of the poorly hydrophobic TMD2, which is not independently recognized by Sec61, suggesting at least one mechanism for inserting non-optimal TMDs that sit at the interface of two membrane protein subunits. The lack of certainty surrounding the complex stoichiometry means that we cannot conclude whether WRB is acting on a single CAML subunit as part of a stable complex or whether it is acting catalytically on multiple copies of CAML. Given that WRB/CAML is itself a membrane protein insertase, it is possible that this post-translational folding is a unique feature of assembly of this complex. However, evidence for other such post-translational topological changes in polytopic proteins suggests that this could be a more general mechanism utilized by multi-subunit complexes (Hegde & Lingappa, 1999; Lu et al., 2000; Serdiuk et al., 2016).

In the event that either CAML or WRB cannot assemble, their orphan forms are recognized and degraded by the ubiquitin-proteasome pathway. This recognition occurs via two distinct mechanisms: (i) improperly folded CAML aberrantly exposes its TMD2 to the ER lumen, which

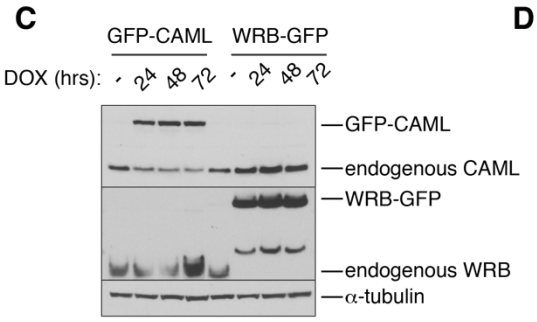
makes it a target for the lumenal quality control machinery while (ii) WRB, though folded correctly, must be recognized due to the aberrant exposure of its subunit interface within the lipid bilayer. As eukaryotic membrane protein subunits differ enormously in size, topology, and the biophysical properties of their exposed interfaces, interaction with such a diverse range of substrates would require a network of chaperones in the ER membrane that remain to be identified. This work sets the stage for future research to determine both the triage factors that target unassembled proteins towards either a biosynthetic or degradative fate, and how these pathways are coordinated to ensure the precise assembly of multi-subunit complexes at the ER.

## 4.4 Supplementary Material




B

| Cell line | Normalized GFP:RFP (%) |
|---|---|
| GFP | 100 ± 2.71 |
| GFP-CAML | 21.6 ± 0.77 |
| WRB-GFP | 30.4 ± 2.34 |



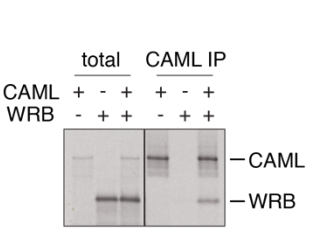




**Figure S4.1 WRB and CAML interact in vivo and in vitro (related to Fig. 1).** (A) Detergent solubilized lysates from stable HEK293 cell lines expressing GFP, GFP-CAML, or WRB-GFP were affinity purified under native conditions via the GFP epitope. Samples were analyzed for co-purification of endogenous CAML and WRB by Western blot. (B) Normalized GFP:RFP ratios from flow cytometry analysis of stable cell lines expressing GFP-2A-RFP, GFPCAML-2A-RFP, or WRB-GFP-2A-RFP. The GFP:RFP ratios were measured and normalized to the GFP-2A-RFP cell line. Displayed are the means and standard deviations (three biological replicates). (C) Exogenous expression of GFP-CAML or WRB-GFP was induced with doxycycline (DOX) for 24, 48 or 72 hours before analysis by Western blot with antibodies against CAML, WRB, GFP and α-tubulin. (D) 35S-methionine labeled FLAG-CAML or WRB were translated alone or in combination in rabbit reticulocyte lysate (RRL) in the presence of canine rough microsomes (cRMs). Following purification of the resulting microsomes and detergent solubilization, samples were affinity purified via the FLAG epitope for analysis. Three biological replicates were performed with consistent results.

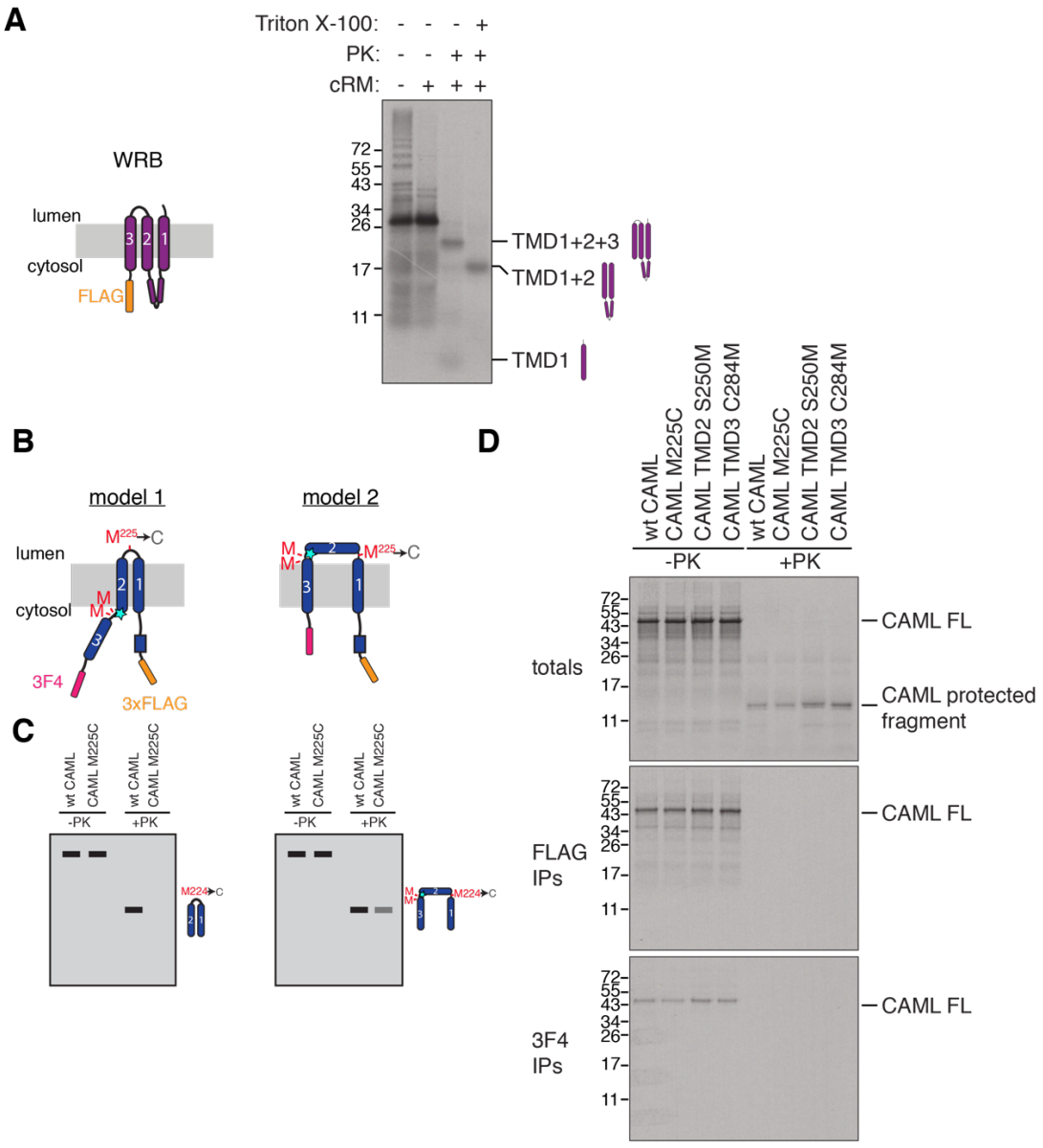



**Figure S4.2 While WRB independently adopts the correct topology, CAML aberrantly exposes TMD2 to the ER lumen.** (A) $^{35}$S-methionine labeled WRB-FLAG was translated in rabbit reticulocyte lysate (RRL) in the presence of canine-derived rough microsomes (cRMs). The total products were treated with proteinase K (PK) in the presence or absence of detergent and then analyzed directly. Schematics are shown alongside. After PK treatment, we detect two protected fragments as in Figure 4.1C. The full length WRB-FLAG shown here is larger than the construct used in Figure 4.1C (1xHA-WRB-3F4), however after PK treatment the protected fragments are the same size in both experiments. This is consistent with cleavage of the C-terminal tag (3F4 in

Figure 4.1C and 3xFLAG shown here). (B) Schematics of the two possible topologies of CAML that position both its N- and C-terminus in the ER lumen, consistent with the PK digests presented in Figure 4.1C. An endogenous glycosylation site is denoted with a star, and the position of $^{35}$S-labeled methionines are highlighted. (C) Mock autoradiograms of the expected protease protection results of wt and M225C CAML for each model. In model 1, the untagged protected fragment of CAML (TMDs1-2) would disappear in the M225C mutant, while in model 2 this fragment (TMDs1-3) would remain visible but be ~60% of the intensity, due to loss of a single radioactive methionine. (C) Protease protection of wt and M225C mutant CAML, as well as CAML mutants that position an addition $^{35}$S-methionine in either TMD2 (S250M) or TMD3 (C284M). The intensity of the observed, untagged, protease protected fragment of CAML is most consistent with model 2.

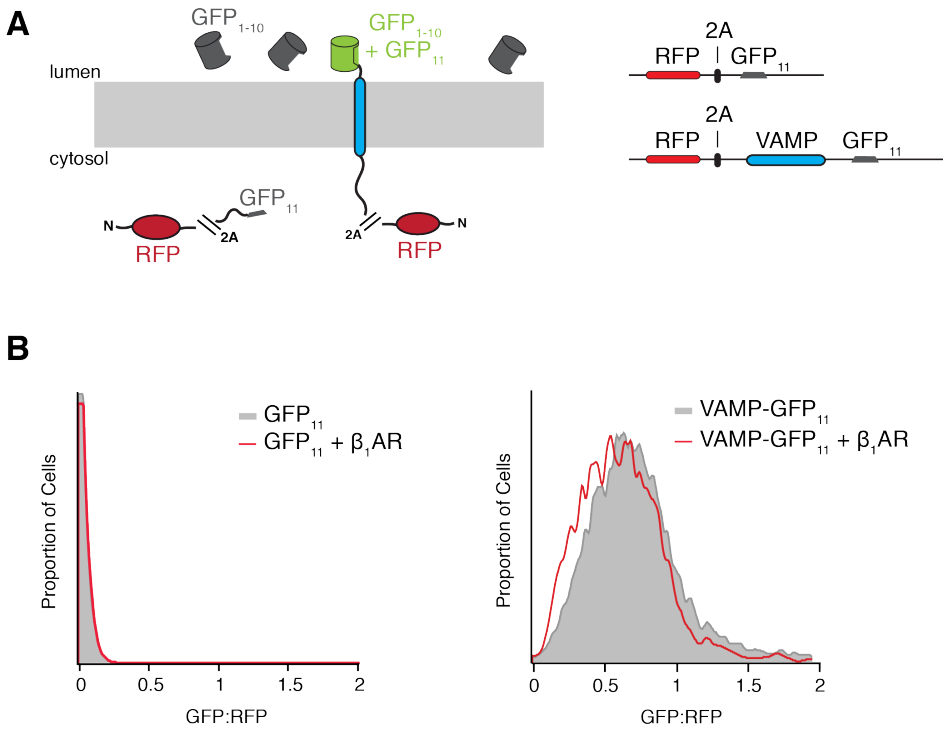



**Figure S4.3 WRB independently adopts the correct topology while CAML aberrantly exposes TMD2 to the ER lumen (related to Fig. 1).** (A) 35S-methionine labeled WRB-FLAG was translated in rabbit reticulocyte lysate (RRL) in the presence of canine-derived rough microsomes (cRMs). The total products were treated with proteinase K (PK) with or without detergent and then analyzed directly. Schematics of products are shown alongside. After PK treatment, we detect two protected fragments as in Figure 4.1C. The full length WRB-FLAG shown here is larger than the construct used in Figure 4.1C (1xHA-WRB-3F4), however after PK treatment the protected fragments are the same size in both experiments. This is consistent with cleavage of the C-terminal tag (3F4 in Figure 4.1C and 3xFLAG shown here). (B) Schematics of the two possible topologies of CAML that position both its N- and C-terminus in the ER lumen, consistent with the PK digests presented in Figure 4.1C. An endogenous glycosylation site is denoted with a star, and the position of 35S-labeled methionine residues are highlighted. (C) Mock autoradiograms of the expected protease protection results of wild type (wt) and M225C CAML for each model. In model 1, the untagged protected fragment of CAML (TMDs1-2) would disappear in the M225C mutant, while in model 2 this fragment (TMDs1-3) would remain visible but be ~60% of the intensity, due to loss of a single radioactive methionine. (D) Protease protection of wild type (wt) and M225C mutant CAML, as well as CAML mutants that position an addition 35S-methionine in either TMD2 (S250M) or TMD3 (C284M). The intensity of the observed, untagged, protease protected fragment

of CAML is most consistent with model 2. Experiments were repeated multiple times with consistent results.

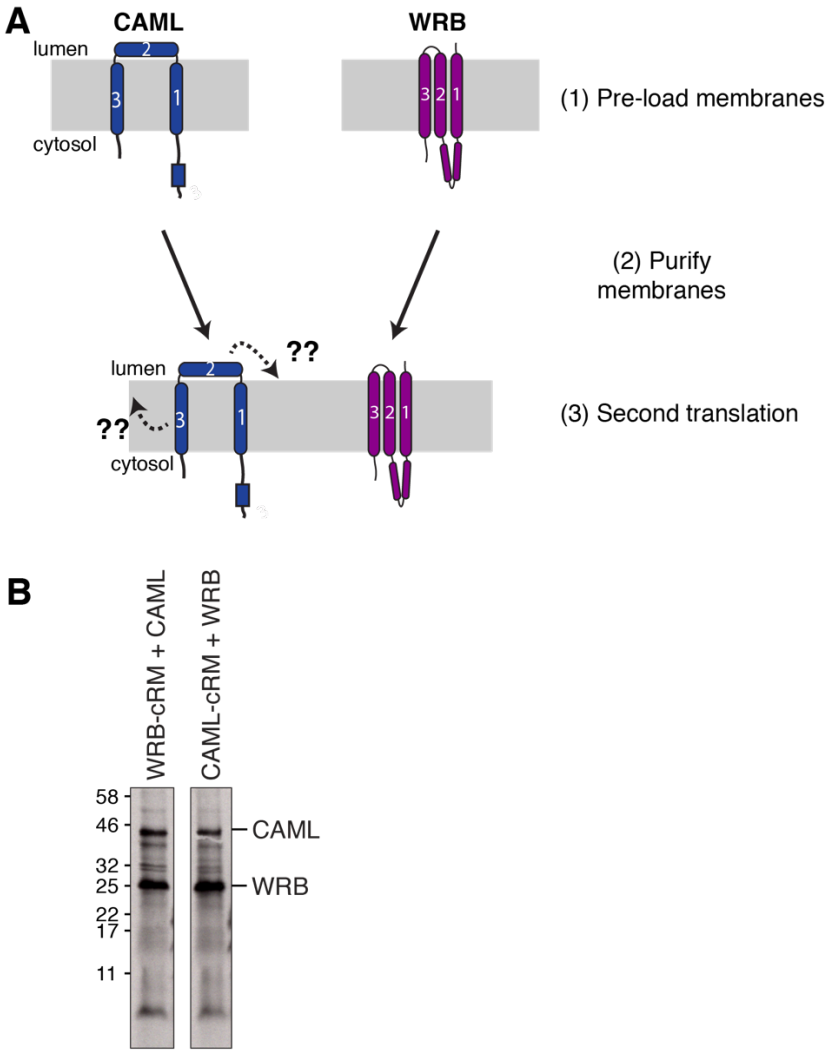



**Figure S4.4 Schematic showing cRM preloading to test the role of translation order on CAML folding. (A)** After an initial incubation to allow translation and insertion of each subunit, membranes, which now contain a radioactively labeled and inserted subunit, are re-purified. These membranes can then be used in a subsequent translation reaction, where the second subunit is introduced. (B) Total translation reactions for the experiment shown in Figure 4.4C to ensure equal amounts of WRB were translated in the two conditions. The positions of full length WRB and CAML are indicated.

## 4.5 Materials and Methods

Experimental Model and Subject Details

*Cell lines*

Flp-In 293 T-Rex cells (female) were maintained in Dulbecco's Modified Eagle's Medium (DMEM) with 10% fetal calf serum (FCS) in the presence of 15 μg/mL blasticidin and 100 μg/ml hygromycin. Cells were grown at 5% $CO_2$ and at 37 ˚C.

Method Details

*Constructs*

Constructs for expression in cultured mammalian cells were generated in either the pcDNA5/FRT/TO (Thermo Scientific) or pcDNA3.1 backbone. To create the fluorescent reporters described in Figure 4.1A, cDNA for human CAML [(cDNA)*CAMLG*] and WRB [(cDNA)*GET1*] was purchased from IDT and inserted into a pcDNA5 vector expressing GFP-2A-RFP resulting in an N- (CAML) or C-terminal (WRB) GFP fusion. In order to express the split $GFP_{1-10}$ in the ER lumen, a construct expressing the human calreticulin signal sequence preceding a $GFP_{1-10}$-KDEL was also generated in pcDNA5 (Cabantous et al., 2004; Kamiyama, Sekine, Barsi-rhyne, et al., 2016). WRB-BFP, the turkey $\beta_1$-adrenergic receptor, CAML-$GFP_{11}$ ($GFP_{11}$ tag: RDHMVLHEYVNAAGIT), cytosolic RFP-2A-$GFP_{11}$, and RFP-2A-VAMP-$GFP_{11}$ were inserted into pcDNA3.1 for transient mammalian expression. All experiments were performed in the Flp-In T-REx 293 cell line (Thermo Scientific). The mCherry and mEGFP versions of RFP and GFP are used throughout this manuscript, though are referred to as RFP and GFP for simplicity in the text and figures.

Constructs for expression in rabbit reticulocyte lysate (RRL) were based on the SP64 vector (Promega). For all protease protection assays (Figures 4.1C, 4.1D and 4.2A and 4.4C) CAML was expressed with an N-terminal 3xFLAG tag and a C-terminal 3F4-tag (Stefanovic & Hegde, 2007) while WRB was appended with an N-terminal 1xHA tag and C-terminal 3F4 tag (except in Figure S4.2A, where WRB is C-terminally 3xFLAG tagged). Tags were chosen to minimize interference with TMD insertion, with tags containing multiple charged or polar residues being placed on the cytosolic face.

*Cell culture*

Stable cell lines expressing GFP-CAML-2A-RFP, WRB-GFP-2A-RFP, or ER $GFP_{1-10}$ were generated using the Flp-In T-Rex 293 Cell Line (Thermo Scientific) according to the manufacturer's instructions. In brief, a 10 cm dish of cells was transfected with 9 μg of Flp-Recombinase (plasmid pOG44) and 1 μg of a specific pcDNA5/FRT plasmid using TransIT-293 transfection reagent (Mirus, MIR2705). 48 hours after transfection, cells were selected with 100 μg/mL hygromycin in DMEM media containing 10% fetal bovine serum and 15 μg/mL blasticidin. After 7-10 days the resulting isogenic cell population was expanded for maintenance and preservation.

For overexpression of GFP-tagged CAML and WRB (Figure S4.1C), cells were cultured in 6-well tissue culture plates, induced with 1 μg/mL doxycycline for 24 to 72 hours, and harvested in 5 mM EDTA pH 8.0 in 1X PBS. Cells were lysed with NETN lysis buffer (250 mM NaCl, 5 mM EDTA pH 8.0, 50 mM Tris-HCl pH 8.0, 0.5% IGEPAL CA-630, 1X protease inhibitors) for 1 hour at 4 °C. Cell lysates were used directly for analysis by Western blot. Samples were normalized by cell counting prior to lysis.

*Purification from cells*

Purification of GFP-tagged CAML and WRB from mammalian cells were performed using an anti-GFP nanobody (Kirchhofer et al., 2009; Pleiner et al., 2015). Briefly, cell lines of GFP-2A-RFP, WRB-GFP-2A-RFP, and GFP-CAML-2A-RFP were cultured in 10 cm dishes until 70% confluent, induced with 1 μg/mL doxycycline and harvested after 24 hours. Cells were lysed in Solubilization Buffer (50 mM HEPES pH 7.5, 200 mM KOAc, 2 mM $MgOAc_2$, 1% Digitonin, 1X protease inhibitors, 1 mM DTT) for 20 minutes at 4 °C. Pierce Streptavidin Magnetic Beads (Thermo Scientific, 88817) were equilibrated with 3.75 μg biotinylated anti-GFP nanobody in wash buffer (50 mM HEPES pH 7.5, 100 mM KOAc, 2 mM $MgOAc_2$, 0.25% Digitonin, 1 mM DTT). Cell lysates were incubated with anti-GFP nanobody immobilized on Streptavidin support for one hour at 4 °C. GFP-tagged proteins were eluted with 0.5 μM SUMOstar protease and used directly for Western blot analysis.

*Western blot analysis*

Antibodies were purchased against CAML (Synaptic Systems, 359 002), WRB (Synaptic Systems, 324 002), and α-tubulin (Sigma, T9026). The antibody against the 3F4 epitope was a gift from the

Hegde lab and has been previously described (Chakrabarti & Hegde, 2009). Secondary antibodies used were HRP-conjugated Goat Anti-Rabbit (BioRad, 170-6515) and Anti-Mouse (BioRad, 172-1011). Anti-FLAG (A2220) and HA resin (A2095) were obtained from Sigma (St. Louis, MO). Pre-designed Silencer Select siRNA from Thermo Fisher were obtained for CAML (s2370, s2371, s2372) and WRB (s14904, s14905).

*Flow Cytometry*

All siRNA experiments (Figure 4.1A) were performed in a 6-well tissue culture plate. Cells were transfected with 3 ng of siRNA per well using RNAiMAX lipofectamine (ThermoFisher, 13778150). After 48 hours, the integrated reporter gene was induced with 1 μg/mL doxycycline for 24 hours. Live cells were first incubated with trypsin before collection, pelleted, and resuspended in 300 μL of PBS containing 1 μM Sytox Blue Dead Cell Stain (ThermoFisher, S34857) and analyzed on a Miltenyi Biotech MACSQuant VYB Flow Cytometer. Data analysis for all flow cytometry experiments was performed using the FloJo software package.

*GFP complementation assays*

GFP complementation experiments by flow cytometry were performed in a 6-well tissue culture plate. Expression of the $GFP_{1-10}$ protein was induced for 72 hours with 100 ng/mL doxycycline before transfection of 0.17 μg of $GFP_{11}$ constructs, 0.17 μg of WRB-BFP or $\beta_1$AR-BFP, and 1.36 μg of pcDNA3.1 backbone with TransIT-293 transfection reagent. Cells were harvested and analyzed by flow cytometry 24 hours after transfection. For analysis by confocal microscopy, the cells were grown in a 24-well tissue culture plate containing 12 mm glass coverslips coated in poly-D-lysine. The induction and transfection conditions for imaged samples were identical as those subjected to flow cytometry, except cells were transfected with 30 ng of RFP-2A-CAML-$GFP_{11}$, 30 ng of BFP or WRB-BFP and 240 ng of pcDNA3.1 backbone. The cells were fixed for fluorescence microscopy according to standard protocol. In brief, the cells were washed with PBS before being incubated with 3.6% paraformaldehyde for 30 minutes. The cells were washed again, treated with Prolong Diamond Antifade Mountant (ThermoFisher, P36961) and sealed onto a slide. Imaging was performed using an LSM 800 confocal microscope (Zeiss).

*Mammalian in vitro translation*

Translation extracts were prepared using nucleased rabbit reticulocyte lysate (RRL) and canine derived pancreatic microsomes (cRMs) as previously described (Sharma et al., 2010; Walter &

Blobel, 1983). Briefly, templates for *in vitro* transcription were generated by PCR using primers that included the SP6 promoter at the 5' end and a stop codon followed by a short untranslated region at the 3' end. In the case of Figure 4.4A and 4.4B, primers were designed to anneal upstream of the stop codon in order to generate a truncated protein product in which the C-terminal residue is a valine, known to stabilize the peptidyl-tRNA product (Shao et al., 2013). Transcription reactions were incubated at 37 °C for 1 hour, and then used directly in a translation reaction, which was incubated for 35 minutes at 32 °C. Where stated, puromycin was added to a final concentration of 1 mM and samples were incubated at 32 ˚C for a further 10 minutes.

To generate pre-loaded membranes of either WRB or CAML, as used in Figures 4.2A, 4.4B and 4.4C, cRMs were included in an initial translation reaction for 12 minutes with the respective mRNA. Membranes were purified by pelleting for 20 minutes at 55,000 rpm in a TLA55 at 4 ˚C through a 20% sucrose cushion in physiological salt buffer (50 mM HEPES pH 7.5, 100 mM KOAc, 2 mM $MgOAc_2$). Pellets were resuspended in physiological salt buffer at a concentration of $A_{280}$ ~80. Membranes were either used directly in a second translation/insertion reaction or aliquoted and flash frozen for storage at -80 °C. We saw no reduction in translation and insertion efficiency after freezing.

*Proteinase K digestion*

Protease digestions were performed on ice by addition of 0.5 mg/mL proteinase K to translation reactions and incubated for an additional hour. The digestion was quenched by addition of 5 mM PMSF in DMSO, followed by transfer to boiling 1% SDS in 0.1 M Tris pH 8.0 (room temperature). Immunoprecipitation of protected fragments was performed in IP buffer (50 mM HEPES pH 7.5, 100 mM KOAc, 2 mM $MgOAc_2$, and 1% Triton X-100).

*Co-immunoprecipitation assays*

Co-immunoprecipitation experiments (Figures 4.4A, 4.4B and S4.1D) were performed by setting up translation reactions in the presence of cRMs, and then purifying the membranes via pelleting for 20 minutes at 55,000 rpm in a TLA55 at 4 ˚C through a 20% sucrose cushion in physiological salt buffer. The pellets were resuspended in physiological salt buffer before solubilization of the membranes in 1% digitonin. The samples were incubated on ice for 10 minutes, before being centrifuged for 15 minutes at 55,000 rpm in a TLA55 at 4 ˚C. The subsequent supernatants were then diluted four-fold and immunoprecipitated with anti-FLAG resin.

Quantification and Statistical Analysis

*Flow cytometry*

For Figure S4.1B, GFP:RFP ratios were calculated in triplicate, and normalized to the GFP-2A-RFP cell line (=1). The mean ± standard deviations are shown (n=3).

*Image quantification*

In Figures 4.2A and 4.4C, the amount of protected TMD3-3F4 was quantified in Image J by inverting the image, subtracting background, then normalizing the values to the total amount of CAML present.

Data and Code Availability

This study did not generate any datasets or codes.

**4.6 Acknowledgements**

We thank T. Pleiner for help with GFP affinity purification, and the Caltech Flow Cytometry facility for their help with FACS experiments. Confocal imaging was performed in the Caltech Biological Imaging Facility, with the support of the Caltech Beckman Institute and the Arnold and Mabel Beckman Foundation. The antibody against 3F4 was a kind gift from Ramanujan S. Hegde. This work was supported by the Heritage Medical Research Institute, the Kinship Foundation, the Pew-Stewart Foundation, and the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number DP2GM137412.

*Chapter 5*

# TXNDC15 IS A NOVEL ADAPTER OF THE MARCHF6 E3 UBIQUITIN LIGASE AND REGULATES THE CILIARY LANDSCAPE

## 5.1 Abstract

Roughly half of the proteome must assemble into multi-subunit complexes. How the cell regulates the assembly and quality control of these complexes is incompletely understood. Here, we study the quality control of GET1, subunit of the obligate heterooligomeric complex GET1/GET2 at the mammalian ER. Unbiased genome-wide CRISPR interference (CRISPRi) screens identified factors involved in the turnover of GET1, including the novel membrane protein TXNDC15. TXNDC15 localizes to the ER and interacts with the E3 ubiquitin ligase MARCHF6. Using genetic modifier screens, we show that TXNDC15 functions within the MARCHF6 ERAD pathway, and that this pathway functions in parallel to the HRD1 ERAD pathway. Additionally, we performed whole-cell proteomics to identify the endogenous substrates of the MARCHF6-TXNDC15 complex and discover that this complex is critical in regulating ciliary and cytoskeletal protein levels. Through mechanistic dissection of TXNDC15, we link its role in quality control to known pathophysiologic outcomes associated with defects of primary cilium development.

## 5.2 Introduction

Nearly half of the proteome is organized into multi-subunit complexes (Benschop et al., 2010; Gavin et al., 2006; Krogan et al., 2006). These multimeric complexes underpin a diverse range of essential cellular functions including transcription, ion homeostasis, and maintenance of cellular structure (Holmes et al., 1990; Koch et al., 2008; L. Lin et al., 2000; Nicholls et al., 2002). The cell is tasked with the challenge of orchestrating the assembly process to ensure that subunits are incorporated into functional oligomers with the correct stoichiometry. When these highly regulated processes are disrupted, a broad range of pathologies arising from gain- or loss-of-function can result (David et al., 2012; X. Wang et al., 2012). Moreover, any excess unassembled subunits must be properly recycled or degraded to avoid aggregating and causing cytotoxicity (Juszkiewicz & Hegde, 2018; Tye et al., 2019). To ensure fidelity of complex biosynthesis, the cell must carefully discriminate between orphaned subunits and fully formed complexes or subcomplexes. The

challenge of molecular discrimination is made significantly more complicated when considering the diversity of the proteome that the cell must constantly survey (Harper & Bennett, 2016).

For one class of oligomeric complexes, ER membrane proteins, the assembly process remains poorly understood. First, individual subunits translated by ribosomes in the aqueous cytosol must be properly integrated into the hydrophobic ER bilayer (Rapoport et al., 2017). Many of these subunits have charged or polar residues that lie at the interface with their binding partner, which would be unfavorably exposed to the hydrophobic environment prior to assembly (von Heijne, 2007). In this vein, we have previously demonstrated that there are different modes of orphan subunit recognition in the ER (Inglis et al., 2020). First, orphan membrane proteins may be incorrectly folded prior to assembly, exposing degrons to the cell environment, and thus recognized and degraded by quality control (QC) machinery within the ubiquitin proteasome pathway. The second class of subunits fold properly in the ER bilayer but are still recognized by QC machinery. It remains poorly understood how these properly folded subunits are recognized and degraded by QC machinery in the ER.

ER associated degradation (ERAD) factors within the ubiquitin proteasome system in the ER enable the turnover of misfolded or unassembled membrane proteins (Krshnan et al., 2022). This involves a cascade event in which ultimately protein substrates are tagged with ubiquitin, a process that involves E1-activating, E2-conjugating and E3-ligase enzymes (Komander & Rape, 2012). Dislocation of the ubiquitinated protein from the ER membrane is facilitated by the ATP-dependent p97 (Bar-Nun, 2005). The human cell contains upwards of 600 unique E3 ubiquitin ligases, many of which are membrane embedded and function at the ER (The UniProt Consortium, 2017). Additionally, there are many protein factors that work in combination with these E3 ubiquitin ligases to enable the specificity of protein targeting (Hegde & Ploegh, 2010). However, the suite of factors required to surveil the diversity of membrane proteins trafficked through the ER, including those proteins destined for assembly into multi-subunit complexes, has not been fully elucidated.

We have previously shown that the obligate heterooligomeric complex GET1/GET2 (formerly WRB/CAML) at the mammalian ER is regulated via two distinct mechanisms (Inglis et al., 2020). GET2 adopts an improper fold and exposes a TMD to the ER lumen that acts as a degron. However, while GET1 folds properly in the absence of GET2, it is unstable and degraded by QC machinery

(Colombo et al., 2016; Rivera-Monroy, Musiol, Unthan-Fechner, Farkas, Clancy, Coy-Vergara, et al., 2016). Here, we set out to identify factors in the mammalian ER that are responsible for turnover of orphan GET1 and to characterize the molecular mechanism of orphan subunit recognition.

## 5.2 Results

### Identification of the novel ERAD factor TXNDC15

GET1 and GET2 form a stable complex in the ER when co-expressed. However, exogenous expression of a single subunit on its own leads to degradation of the excess subunit (Figure 5.1A). To identify factors leading to degradation of GET1, we performed an unbiased genome-wide screen. We utilized a fluorescent reporter strategy in which GFP was appended to GET1 followed by a 2A site and an RFP normalization marker (Figure 5.1B) (Inglis et al., 2020; Itakura et al., 2016). A K562 cell line stably expressing this reporter and the CRISPR inhibition (CRISPRi) machinery was generated and transduced with the genome-wide CRISPRi-v2 sgRNA library (Gilbert et al., 2014; Horlbeck et al., 2016). Cells were collected based on the GFP:RFP ratio, with the highest and lowest 30% GFP:RFP sorted and deep sequenced to identify guides enriched in either population. Depletion of biogenesis factors destabilizes GET1, resulting in a decrease in GFP fluorescence. Depletion of QC factors causes GET1 to accumulate in the cell, resulting in an increase in GFP fluorescence. As expected, depletion of the binding partner GET2 destabilizes the GET1 reporter protein (Figure 5.1B). However, depletion of the E1 ubiquitin ligase UBA1, a known factor required in ERAD pathways (Schulman & Wade Harper, 2009) results in stabilization of GET1. We also identified TXNDC15 (thioredoxin-domain containing 15), a novel protein of unknown function. Depletion of TXNDC15 results in the greatest stabilization phenotype of GET1 in the genome-wide screen, and we therefore sought to characterize it further.

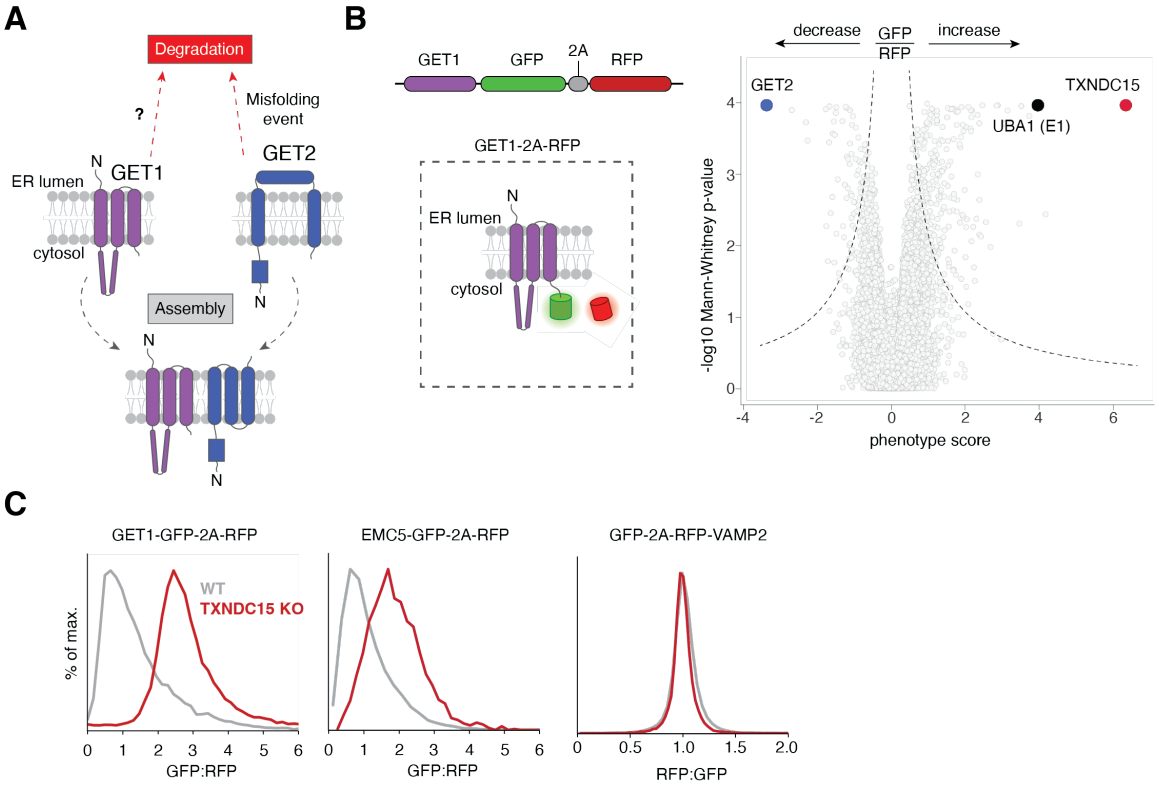



**Figure 5.1 Identification of a novel ERAD factor TXNDC15. (A)** Schematic of the assembly and quality control of the GET1/2 complex. **(B)** (Top) Schematic of the GET1 reporter mRNA used in a genome-wide CRISPR interference (CRISPRi) screen. (Bottom) Topology of the GET1-GFP reporter protein. RFP is expressed from the same open reading frame as GET1-GFP, separated by a P2A skipping sequence, and is used as a translational normalization marker. (Right) Volcano plot of the GFP:RFP phenotype for the GET1 reporter with the three strongest sgRNAs plotted with Mann-Whitney p-values. Individual genes are displayed in grey with specific factors highlighted and labelled. Genes that fall outside the indicated dashed lines represent statistically significant hits. **(C)** Flow cytometry assay of ER membrane protein reporters in WT or TXNDC15 KO cells. Histogram of the GFP:RFP ratio (GET1 and EMC5) or RFP:GFP ratio (VAMP2) is displayed.

TXNDC15, is an integral membrane protein that contains a signal sequence, a single TMD close to its C-terminus, and a thioredoxin domain in its N-terminal soluble domain. Thioredoxin domains contain catalytic cysteines that exhibit oxidoreductase activity and enable reversible disulfide formation of substrate proteins (Arnér & Holmgren, 2000; Martin, 1995). However, we note that the thioredoxin domain in TXNDC15 contains a serine in place of a catalytic cysteine, which suggests that it may not be active in disulfide bond formation (Figure S5.1A). Mutations in

TXNDC15 result in various ciliopathies, diseases that affect the development of the primary cilium, as well as physical deformations of the primary cilium (Alexiev et al., 2006; Breslow et al., 2018; Ridnõi et al., 2019). Since the only known information about TXNDC15 is that it is associated with ciliary development, we sought to characterize its localization in the cell. To avoid overexpression artifacts, we first engineered a HEK 293T cell line in which we endogenously tagged TXNDC15 with a GFP tag at its N-terminus (Figure S5.1B). We then used this cell line in immunofluorescence assays and observe ER localization of GFP-TXNDC15 (Figure S5.1C).

As a putative quality control factor of GET1, we set out to determine how general the effect of TXNDC15 is on membrane proteins at the ER. To do this, we tested whether depletion of TXNDC15 affects protein substrates beyond the GET1 substrate used in the screen. We generated TXNDC15 KO HEK 293T cell lines and assayed protein substrates in the GFP-2A-RFP reporter cassette. In addition to GET1, we included EMC5 and VAMP2. EMC5 is a multipass membrane protein that forms a complex with other ER membrane complex (EMC) subunits, including a subcomplex with EMC6 (Pleiner et al., 2020). VAMP2 is a tail-anchored membrane protein that is post-translationally inserted into the ER and represents a distinct class of proteins compared to GET1 and EMC5 (Guna, Hazu, et al., 2023). We observed that both GET1 and EMC5 are stabilized 2-3-fold in TXNDC15 KO compared to wildtype cells, while VAMP2's stability remained unchanged (Figure 5.1C). While VAMP2 is stable as a monomer, both GET1 and EMC5 are members of obligate heterooligomeric complexes in the ER and are therefore unstable and degraded in the absence of their partner subunit(s). This indicated to us that TXNDC15 may be involved in orphan subunit quality control at the ER. However, how TXNDC15 functions in quality control to clear unstable orphan proteins from the ER remained unclear.

**Genome-wide identification of genetic modifiers of TXNDC15**

Our observation that TXNDC15 depletion stabilizes a subset of membrane proteins, including orphan membrane proteins, suggests that it may be involved in specific ERAD pathways at the ER. We therefore used an unbiased screen-based strategy to identify genetic modifiers of TXNDC15. We used a dual guide CRISPRi screening approach that allows simultaneous depletion of two genes (Figure 5.2A) (Guna, Page, et al., 2023). We cloned two dual guide libraries in which there is a non-targeting (NT) guide or a guide targeting TXNDC15 in the anchor position followed by the CRISPRi-v2 sgRNA library. There are three class of genes identified from genetic modifier screens: 1) genes that have an enhanced phenotype in TXNDC15-dual compared to NT-dual that

represent factors in parallel or redundant pathways to TXNDC15, 2) genes that have a diminished phenotype in the TXNDC15-dual compared to NT-dual, indicating that they are epistatic or in the same pathway as TXNDC15, and 3) genes that have relatively unchanged phenotypes in the TXNDC15-dual compared to NT-dual and are therefore independent of the TXNDC15 pathway.

We performed the genetic modifier screens with the GET1 reporter as described in Figure 5.1 and identify several known QC factors that are significant in both the NT-dual and TXNDC15-dual screens, including FAF2, UBA1, UBE2J2, UBE2G2, and ATP13A1 (Li et al., 2007; McKenna et al., 2022; Olzmann et al., 2013; Tiwari & Weissman, 2001). Additionally, we identify many known ERAD components have little-to-no phenotype in the NT-dual screen but have strong phenotypes in the TXNDC15-dual screen, including the E3 ubiquitin ligase HRD1, the Derlin DERL2, and the E2 ubiquitin ligases UBE2J1 and UBE2V1 (Figure 5.1B) (Burr et al., 2011; Kikkert et al., 2004; Lilley & Ploegh, 2005). This suggests that these quality control factors act in partially redundant pathways to TXNDC15. Conversely, the ER-resident E3 ubiquitin ligase MARCHF6 (homolog of the yeast Doa10; (Christianson & Carvalho, 2022)) the E2 ligase UBE2Q1 have diminished phenotypes in the TXNDC15-dual screen compared to the NT-dual screen, suggesting that they may operate within the same pathway as TXNDX15.

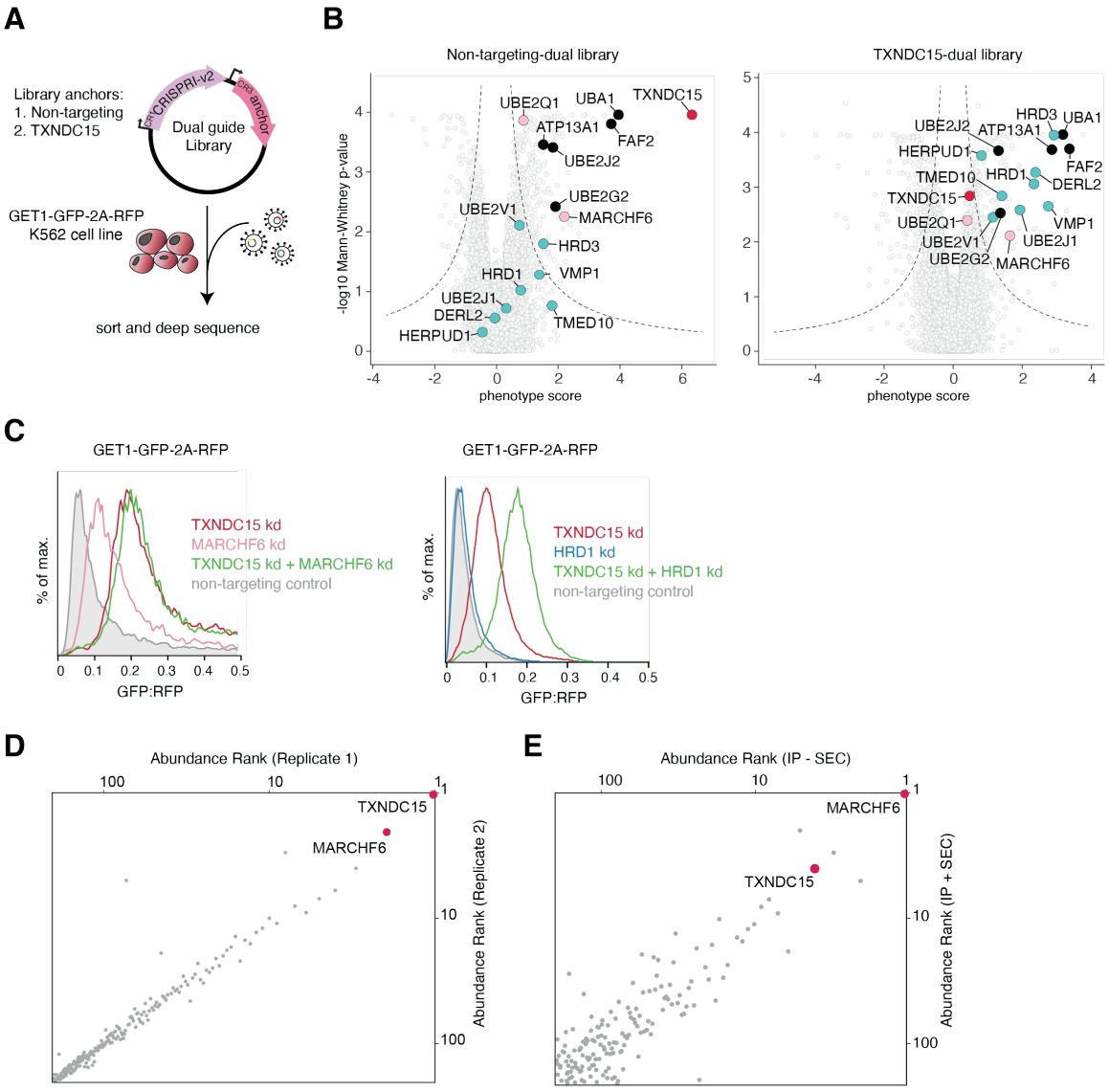



**Figure 5.2 TXNDC15 is an adapter of the E3 ligase MARCHF6. (A)** Schematic of genetic modifier screen of TXNDC15 in the GET1 reporter cell line. Dual guide libraries target the CRISPRi-v2 library and either the Non-targeting or TXNDC15 guides in the anchor position. **(B)** Volcano plots from the dual CRISPRi screens with Non-targeting (Left) or TXNDC15-dual libraries (Right). Quality control factors that are independent of TXNDC15 are highlighted in black. Factors parallel to the TXNDC15 pathway are colored in teal, and factors in the TXNDC15 pathway are colored in pink. **(C)** Analysis of GET1 stability using the GFP:RFP reporter system as analyzed by FACS after knockdown with indicated guides. **(D)** Scatter log-plot of abundance rank as determined by mass spectrometry of co-immunoprecipitation of TXNDC15 endogenously tagged with GFP. **(E)** As in (D) for exogenously expressed MARCHF6-GFP.

Degradation of integral membrane proteins by the ubiquitin proteasome pathway first requires transfer of a ubiquitin moiety to substrate proteins, which is typically performed by an ER localized E3 ubiquitin ligase (Komander & Rape, 2012). We therefore sought to definitively identify the E3 ligase functioning within the TXNDC15 pathway. Because HRD1 is significant in the turnover of GET1 only in the background of TXNDC15 depletion, we reasoned that HRD1 was likely operating in a parallel pathway to TXNDC15. However, the E3 ubiquitin ligase MARCHF6 is significant in both dual screens, decreasing in significance upon the loss of TXNDC15. We therefore set out to determine whether TXNDC15 may be operating within the MARCHF6 pathway and to verify that it was not acting within the HRD1 pathway. We generated programmed dual guides in which either TXNDC15 alone, MARCHF6 alone, or both genes were targeted, and assayed the effect of gene depletion on the GET1 substrate (Figure 5.2C, left). We observed that while depletion of MARCHF6 alone and TXNDC15 alone both have stabilizing effects on GET1, they do not result in an additive stabilizing effect when both genes are targeted simultaneously. We performed this assay with guides targeting HRD1 and TXNDC15, either alone or together, and observed that HRD1 and TXNDC15 depletion is additive (Figure 5.2C, right). Unlike MARCHF6, HRD1 depletion alone does not have an effect on GET1. However, when HRD1 depletion is combined with TXNDC15 depletion, GET1 is further stabilized, recapitulating the genetic modifier screen results. This altogether indicates that TXNDC15 acts within the MARCHF6 ERAD pathway and that HRD1 acts in parallel to this pathway.

We aimed to further characterize the MARCHF6 ERAD pathway to determine whether TXNDC15 is acting in concert with MARCHF6. Using the endogenous GFP-TXNDC15 cell line, we performed a non-denaturing immunoprecipitation using a GFP nanobody-based strategy (Stevens et al., 2023). Using mass spectrometry, we identified MARCHF6 as the strongest interactor of TXNDC15 (Figure 5.2D). In parallel, we exogenously expressed MARCHF6 tagged with a C-terminal ALFA tag and performed a non-denaturing immunoprecipitation using a similar nanobody-based strategy and analyzed the sample via mass spectrometry (Götzke et al., 2019). Similarly, we observed that TXNDC15 was one of the most significant interactors of MARCHF6, indicating that these two proteins physically associate in cells (Figure 5.2E). Moreover, we also determined that the orphan substrates GET1 and EMC5 are stabilized in a MARCHF6 KO cell line, while VAMP2 is unaffected, phenocopying that of the TXNDC15 KO cell line (Figure 5.1C and S5.2A).

**Interaction analysis of TXNDC15 and ER substrates**

Given that TXNDC15 and MARCHF6 physically associate and that their depletion leads to the stabilization of a set of membrane protein reporters, we wondered how the TXNDC15-MARCHF6 complex was recognizing its substrates. Both GET1 and EMC5 adopt the appropriate topology without their partner subunits (Inglis et al., 2020; Pleiner et al., 2020), however they are unstable and degraded by the TXNDC15-MARCHF6 complex. We set out to determine whether TXNDC15 was necessary for the recognition of these substrates. First, we expressed the GET1 substrate in a cell-free translation system (GET1-ALFA) or in cells (GET1-GFP) and performed an immunoprecipitation under conditions in which interactions with other factors would remain intact (Figure 5.3A). We reasoned that the cell-free system offered an advantage over expression in cells when trying to identify physical interactors of GET1. First, the degradation machinery is not active, so unstable nascent proteins accumulate and cannot be turned over by the proteasome. And second, nascent proteins cannot be trafficked, so transient interactions with biogenesis and quality control factors are stabilized. Indeed, we observed that MARCHF6 and TXNDC15 are major interactors of GET1 expressed from the cell-free translation system (Figure 5.3A). Similarly, TXNDC15 and MARCHF6 were identified when GET1 was expressed in cells. We also showed that when GET1-ALFA or the control membrane protein ORF3a from SARS-CoV-2 is translated in the presence of GFP-TXNDC15 in the cell-free translation system, GET1 but not ORF3a co-immunoprecipitates with GFP-TXNDC15 (Figure 5.3B). These data together indicate that the MARCHF6-TXNDC15 complex physically associates with GET1.

The GET1/GET2 complex functions as an insertase of tail-anchored membrane proteins into the ER (Schuldiner et al., 2008; Vilardi et al., 2011, 2014). To enable this function, both GET1 and GET2 contain hydrophilic and charged side chains within their respective TMDs. However, when orphaned, GET1 likely exposes these polar residues to the hydrophobic lipid environment, which is energetically unfavorable. We therefore hypothesized that these polar and charged residues may be mediating the interaction with the MARCHF6-TXNDC15 complex. We tested whether mutating the polar regions in the TMDs of GET1 to hydrophobic residues would reduce the interaction with TXNDC15. Using the cell-free translation system, we translated GET1 variants containing the hydrophobic amino acid leucine in place of polar side chains in TMDs 1-3 of GET1 and asked whether TXNDC15 could still associate with GET1 (Figure 5.3C,D). Surprisingly, we observed an increase in TXNDC15 association in these GET1 mutants compared to wildtype. Moreover, we observed that immunoprecipitation of TXNDC15 resulted in enrichment of ubiquitinated GET1

species, as compared to the total translation of GET1 that is not visibly ubiquitinated. This suggests that TXNDC15 is enriching for species of GET1 that are quality control substrates.

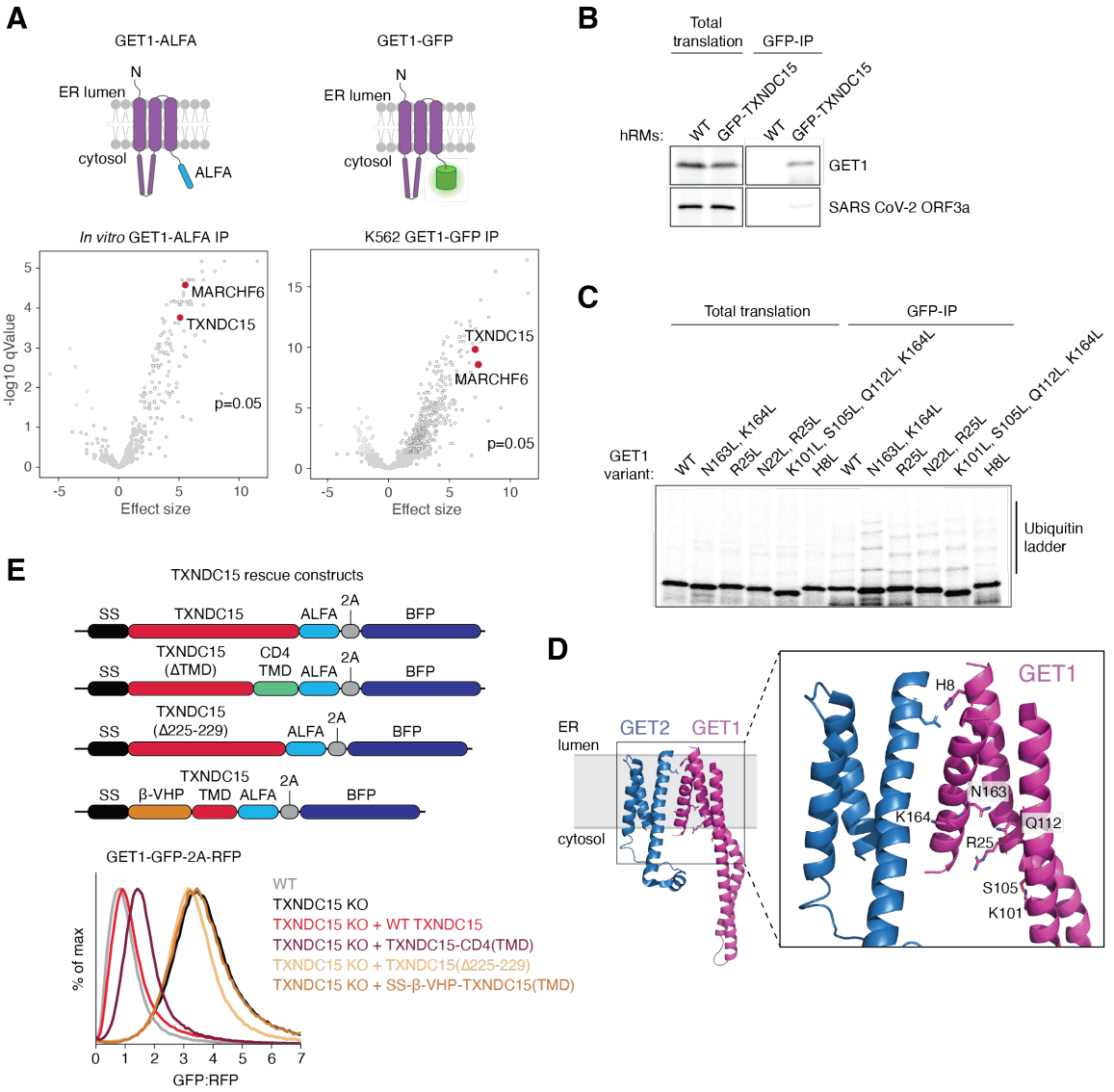



**Figure 5.3 Characterization of the interaction between GET1 and TXNDC15. (A)** (Top) Schematic of GET1-ALFA for *in vitro* expression (Left) or GET1-GFP cell expression (Right). (Bottom) Plot from mass spectrometry analysis of co-immunoprecipitation of the indicated GET1 reporter. **(B)** *In vitro* translation and GFP co-immunoprecipitation of GET1 or the control ORF3a viral protein in the presence of human-derived ER microsomes (hRMs) containing GFP-TXNDC15 at endogenous levels. Samples were analyzed by SDS-PAGE and autoradiography. **(C)** As in (B) for the indicated GET1 variants. Single-letter abbreviations for amino acids are as follows: N, asparagine; L, leucine; K, lysine; R, arginine; S, serine; Q, glutamine; H, histidine. **(D)** Locations

of individual mutations of GET1 in (C). PDB ID: 6SO5. **(E)** (Top) Schematic of TXNDC15 rescue constructs assayed with the GET1 reporter. (Bottom) TXNDC15 KO cells were transduced with constructs described above before transduction with the GET1 reporter.

To further characterize the interaction between TXNDC15 and GET1, we tested whether specific domains in TXNDC15 were required for the quality control of GET1 in cells. For this, we designed variants of TXNDC15 in which we mutated either the TMD or regions in the soluble N-terminal domain (Figure 5.3E). We observed that the TMD of TXNDC15 was dispensable for its function in quality control, as replacement of the TMD with that of the CD4 protein rescued the TXNDC15 KO phenotype on GET1. However, deletion of the N-terminal soluble domain of TXNDC15 or mutation of select residues (Δ225-229) completely disrupted TXNDC15's function in quality control. Interestingly, the variant containing a deletion of five residues within the thioredoxin fold (Δ225-229) is also known to cause a ciliopathy known as Meckel-Gruber Syndrome (Ridnõi et al., 2019). This mechanistically links TXNDC15's ability to perform quality control with a pathophysiological outcome in organisms.

**TXNDC15 and MARCHF6 are regulators of the ciliary landscape**

MARCHF6 and TXNDC15 physically associate and function together in the quality control of several orphan membrane proteins in the ER. Additionally, MARCHF6 is known to be integral for the regulation of lipid biosynthesis (Nguyen et al., 2022; Schultz et al., 2018; Scott et al., 2021) and immune regulation (H. Lin et al., 2019) by modulating the turnover of specific membrane-embedded and cytosolic proteins in these pathways (Stefanovic-Barrett et al., 2018). However, a systematic analysis of endogenous MARCHF6 and TXNDC15 substrates has not been performed. Therefore, we set out to characterize the proteome that is regulated by the MARCHF6-TXNDC15 complex. We knocked down MARCHF6, TXNDC15, or a non-targeting control in K562 cells and performed whole cell proteomics. We expected that proteins directly regulated by the MARCHF6-TXNDC15 complex would be stabilized upon knockdown. As expected, GET1 is differentially stabilized in both the MARCHF6 and TXNDC15 knockdown samples compared to the control sample (Figure 5.4A). Additionally, the known MARCHF6 substrate, SQLE, is stabilized in both the TXNDC15 and MARCHF6 knockdown samples (Tan et al., 2020). Interestingly, we also observed that TXNDC15 is downregulated in the MARCHF6 knockdown sample, indicating that it is unstable in the absence of MARCHF6. This is consistent with MARCHF6 being an obligate partner of TXNDC15 and is a common feature among multi-subunit complexes (Colombo et al.,

2016; Dettmer et al., 2010; Pleiner et al., 2021; Volkmar et al., 2019). Inversely, however, we do not observe a change in MARCHF6 levels upon TXNDC15 knockdown. MARCHF6 is therefore likely not dependent on TXNDC15 for its overall stability, and we cannot rule out the possibility that a fraction of MARCHF6 in cells exists without the TXNDC15 interaction partner.

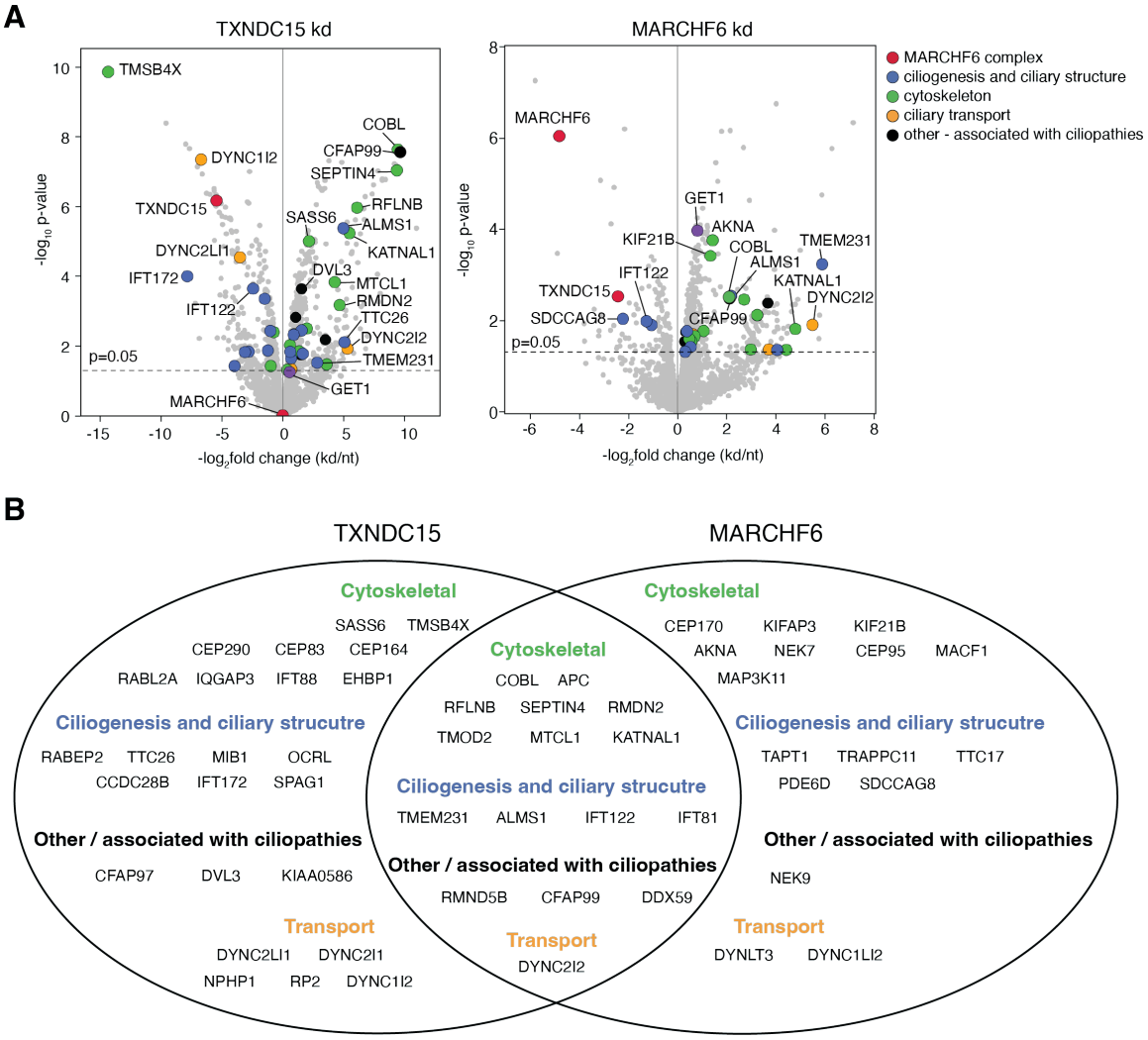



**Figure 5.4 TXNDC15 and MARCHF6 are regulators of the ciliary landscape. (A)** Whole-cell proteomics analysis of TXNDC15 kd (Left) or MARCHF6 kd (Right) compared to a non-targeting control. Individual factors are highlighted and labeled. **(B)** Venn-Diagram of proteins affecting the ciliary landscape differentially expressed in TXNDC15 kd and / or MARCHF6 kd from (A).

Strikingly, we observed that many of the proteins differentially expressed in the TXNDC15 or MARCHF6 knockdowns compared to the control are proteins involved in maintaining the cytoskeleton and in facilitating ciliogenesis, the development of the primary cilium (Figure 5.4A,B). This includes the protein TMEM231, a member of the Meckel syndrome complex (with TCTN1, TCTN2, and TMEM67) located at the base of the primary cilium (Lambacher et al., 2016). This complex at the transition zone, or base of the primary cilium, regulates membrane protein diffusion, and therefore localization, into the primary cilium (Chih et al., 2012; Roberson et al., 2015). Moreover, mutations in this gene cause Meckel-Gruber syndrome, the same ciliopathy associated with mutational variants of TXNDC15 (Shaheen et al., 2013). Altogether, this is consistent with the genetic connection of TXNDC15 to ciliopathies. In addition, many of these proteins are differentially expressed in both MARCHF6 and TXNDC15 samples (Figure 5.4B), suggesting that the MARCHF6-TXNDC15 complex is involved in the regulation of these substrates, and therefore critical in ciliary development. Indeed, these data suggest that TXNDC15's effect in ciliopathies is related to its role in quality control.

**5.4 Discussion**

How orphan subunits are regulated by quality control machinery in the cell is incompletely understood. Here, using genome-wide genetic screens, we have identified a novel quality control factor TXNDC15 that is involved in orphan subunit regulation at the ER. TXNDC15 is an integral membrane protein that localizes to the ER membrane and is required for turnover of the subunits GET1 and EMC5, but not of the monomeric tail-anchored protein VAMP2. To precisely place TXNDC15 within a specific ERAD pathway, we performed genetic modifier screens. Using this strategy, we show that TXNDC15 is epistatic with the E3 ubiquitin ligase MARCHF6, an ER-resident factor required for quality control of both misfolded membrane and cytosolic proteins (Christianson & Carvalho, 2022). In addition to the genetic interaction, we show that TXNDC15 and MARCH6 physically associate in cells and that MARCHF6 is an obligate partner of TXNDC15, consistent with TXNDC15 being a specific adapter to MARCHF6.

It is unlikely that a single factor or pathway can accommodate the wide diversity of proteins in the human membrane proteome. The cell likely requires a large number of quality control factors to both recognize and degrade proteins across the broad range of biophysical properties that exists in the proteome (Hegde & Ploegh, 2010). Consistent with this, the number of E3 ubiquitin ligases has greatly expanded, from approximately 30 in *S. cerevisiae* to over 600 in humans

(The UniProt Consortium, 2017). This expansion may also allow for redundancy, as many cell types in the human body have to be maintained throughout the lifespan, and the risks associated with protein accumulation are far greater (Zavodszky & Hegde, 2019). Indeed, in the genetic modifier screens, we show that the MARCHF6-TXNDC15 pathway functions in parallel to HRD1 and DERL2. We envision that this expansion of the repertoire of quality control factors is one strategy to maintain the proteostasis in the cell.

MARCHF6 and TXNDC15 are essential in humans, likely due to their role in proteostasis. Mutations within the TXNDC15 gene are linked to ciliopathies such as Meckel-Gruber and Joubert Syndrome, which display broad clinical phenotypes including neurodevelopmental delays, cystic kidneys, and heart defects (Alexiev et al., 2006; Saraiva & Baraitser, 1992). In the absence of TXNDC15, the primary cilium, a projection from the cell body that is essential in cell-cell signaling, is malformed (Breslow et al., 2018). However, it was not known how TXNDC15 caused these defects. Here, we show that TXNDC15 and MARCHF6 regulate the levels of proteins involved in ciliogenesis. Many proteins in cilia, including the membrane protein TMEM231, are upregulated in the absence of MARCHF6 and TXNDC15. This is consistent with a direct role of the MARCHF6-TXNDC15 complex in their quality control and regulation. This regulation of specific cellular processes by ERAD factors is not unprecedented, as MARCHF6 was previously shown to regulate proteins in lipid biosynthesis and ferroptosis (Nguyen et al., 2022; Scott et al., 2021). Additionally, many of the ciliary and cytoskeletal proteins dysregulated in the absence of MARCHF6 and TXNDC15 are members of multi-subunit complexes, consistent with our observation that the MARCHF6-TXNDC15 pathway regulates levels of orphan membrane proteins in the ER (Chih et al., 2012). Altogether, these data suggest that the MARCHF6-TXNDC15 pathway functions to regulate levels of membrane protein complex members and thereby plays a critical role in maintaining cellular structure, including the ciliary landscape.

## 5.5 Supplementary Material

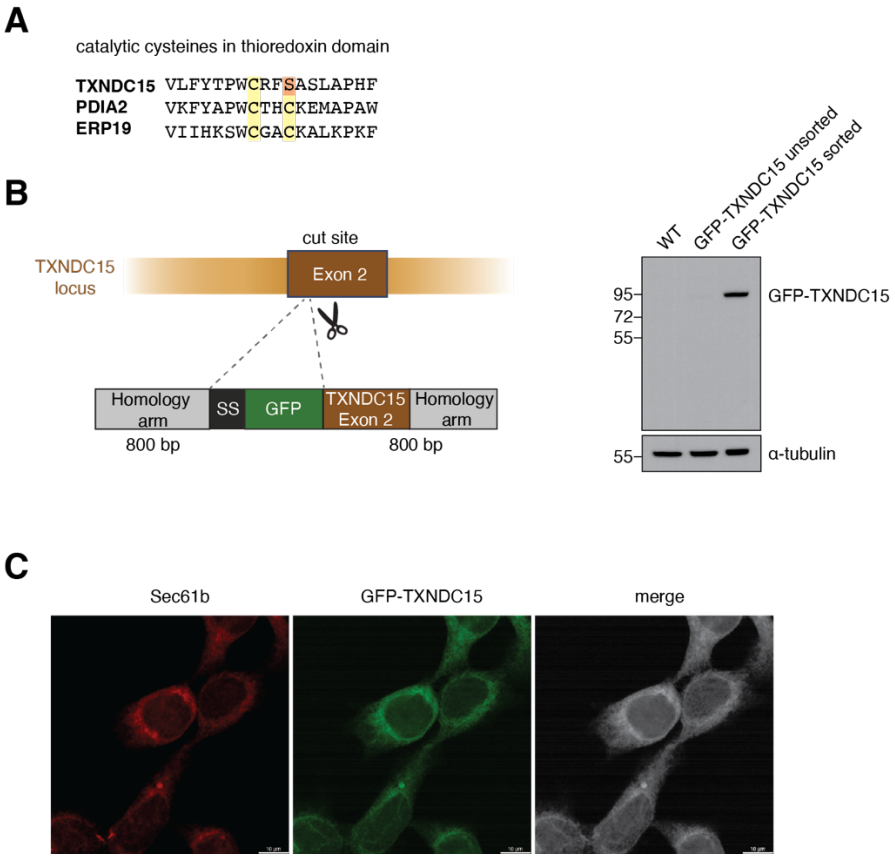



**Figure S5.1 Analysis and endogenous tagging of TXNDC15. (A)** Schematic of TXNDC15 and related proteins containing thioredoxin domains. The catalytic cysteines are highlighted. Note that TXNDC15 contains a serine in place of cysteine in the thioredoxin domain. **(B)** (Left) Schematic of endogenously tagging TXNDC15 with GFP at the N-terminus. (Right) SDS-PAGE and western blotting of cells expressing endogenous GFP-TXNDC15. **(C)** Confocal microscopy of cells endogenously expressing GFP-TXNDC15, displaying its ER localization. SEC61β was used as an ER localization marker.

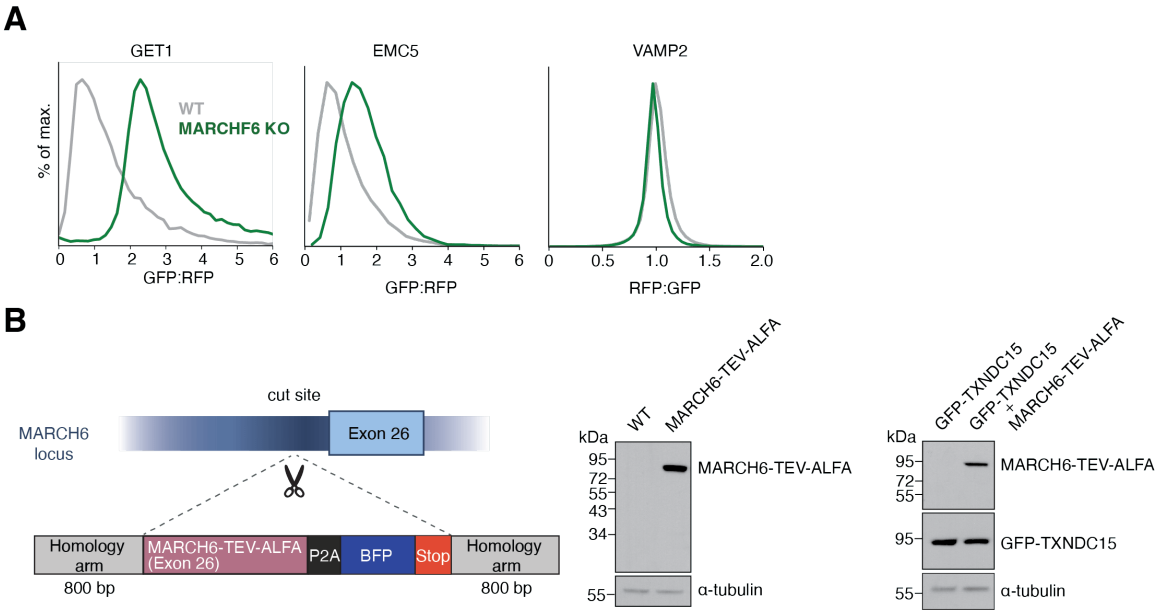



**Figure S5.2 Analysis and endogenous tagging of MARCHF6. (A)** As in Figure 5.1C for MARCHF6 KO cells. **(B)** (Left) Schematic of endogenously tagging MARCHF6 with ALFA-2A-BFP at the C-terminus. (Right) SDS-PAGE and western blotting of cells expressing endogenous MARCHF6-ALFA-2A-RFP alone or both GFP-TXNDC15 and MARCHF6-ALFA-2A-RFP.

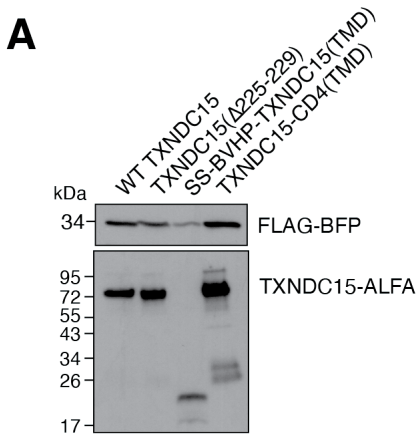



**Figure S5.3 Expression of TXNDC15 mutant constructs. (A)** Samples from Figure 5.3E were analyzed by SDS-PAGE and western blotting to show the relative levels of each TXNDC15 reporter construct. Antibodies targeted BFP or the ALFA epitope. Note that SS-β-VHP-TXNDC15(TMD) construct runs at a smaller molecular weight (between 17-26 kDa) compared to the other constructs.

## 5.6 Materials and Methods

*Plasmids and antibodies*

The sequences used in cell-based and *in vitro* experiments were derived from UniProtKB/Swiss-Prot. These include: guided entry of tail-anchored proteins factor 1 (GET1; O00258), vesicle associated membrane protein 2 (VAMP2; P63027-1), ER membrane protein complex subunit 5 (EMC5/MMGT1; Q8N4V1), thioredoxin domain-containing protein 15 (TXNDC15; Q96J42), E3 ubiquitin-protein ligase MARCHF6 (MARCHF6; O60337), SARS-CoV-2 ORF3a (AP3A_SARS2; P0DTC3), T-cell surface glycoprotein CD4 (CD4; P01730).

The 2nd generation lenti-viral packaging plasmid pCMV-VSV-G was a gift from Bob Weinberg (Addgene plasmid # 8454). The 2nd generation lenti-viral packaging plasmid psPAX2 was a gift from Didier Trono (Addgene plasmid #12260). The pHAGE2 lenti-viral transfer plasmid was a gift from Magnus A. Hoffmann and Pamela Bjorkman. The dual guide lenti-viral vector pJR103 was a gift from Jonathon Weissman (Addgene plasmid #187242). The SFFV-tet3G backbone was used for K562 cell expression during CRISPRi screens (Jost et al., 2017). RFP is used in the text and figures and refers to the mCherry variant that was used in this study. Similarly, GFP refers to the EGFP variant. The GFP:RFP reporter system was used as previously described (Chitwood et al., 2018; Guna et al., 2018).

GET1 was cloned with a C-terminal GFP fusion followed by a 2A and RFP normalization marker, as previously described, for expression in K562 cells during genome-wide CRISPRi screens (Inglis et al., 2020). For screens, expression of GET1-GFP-2A-RFP was driven from the doxycycline-inducible TET3G promoter. For expression in HEK 293T cells, GET1 and EMC5 were cloned with a C-terminal GFP fusion followed by a 2A and RFP in the pHAGE2 lenti-viral vector. The TMD and flanking residues of VAMP2 were cloned as a fusion to the C-terminal of RFP in the GFP-2A-RFP cassette in the pHAGE2 lenti-viral vector.

Endogenously tagging TXNDC15 and MARCHF6 was performed as previously described (Zhang et al., 2017). Briefly, to endogenously tag TXNDC15 with GFP at the N-terminus, a protospacer within exon two was cloned into pX459 (CCTGGCCCATCGGGTCATG). A vector encoding GFP and 800 bp of homology to the TXNDC15 locus on either side of the cut site was cloned. The GFP insertion site was designed such that the native signal sequence of TXNDC15 would remain intact. A similar strategy was used for tagging MARCHF6. The protospacer sequence was designed in

exon 26, near the stop codon and was cloned into pX459 (CCTTGTTTTTGCAACCTAGG). The sequence of TEV-ALFA-2A-BFP was cloned into an empty vector with 800 bp homology on either side of the cut site of the MARCHF6 locus. The ALFA epitope (PSRLEEELRRRLTEP) has been previously described (Götzke et al., 2019b).

Programmed single and dual guides were made to assay depletion of one or two genes (Replogle et al., 2020). The following sgRNA protospacer sequences were used to clone into pLG1 (single guides) or pJR103 (dual guides):

| **Gene** | **sgRNA protospacer sequence (5'-3')** |
|---|---|
| Non-targeting control | GACGACTAGTTAGGCGTGTA |
| TXNDC15 | GACGCGGGGCGGTCGTCGAC |
| MARCHF6 | GCCAGACAAGATGGACACCG |
| SYVN1/HRD1 | GGGAGTGTTGTTAACCGGAG |

To generate knockout cell lines, the following sgRNAs were cloned into pX459: TXNDC15 (CGGGCATTTCCAGCTCTTCA), MARCHF6 (CCATGGACTATCTGCTCTCT).

The following siRNAs were used in this study: negative control no. 2 siRNA (#4390846), (all Silencer Select; ThermoFisher Scientific, USA).

Constructs for expression in rabbit reticulocyte lysate (RRL) were based on the SP64 vector (Promega). For *in vitro* translations, the ALFA epitope (PSRLEEELRRRLTEP) and a flexible 3X GS linker was appended to the C-terminal of GET1 or the control protein ORF3a from SARS-CoV-2.

TXNDC15 constructs were cloned with a C-terminal ALFA fusion followed by a 2A and BFP in the pHAGE2 vector. Mutational variants of TXNDC15 were cloned such that either the native TMD was replaced with the TMD of CD4, the ER lumenal domain was replaced by a soluble β-VHP domain (LSDEDFKAVFGMTRSAFANLPLWKQQNLKKEKGLF), or the thioredoxin domain residues in the lumenal domain were deleted (Δ225-229). In all variants, the native signal sequence of TXNDC15 remained intact.

All plasmids are available upon request.

The following antibodies were used in this study: alpha tubulin (T9026, Millipore Sigma, USA); GFP (A11122, Invitrogen); Anti-FLAG-HRP (Millipore-Sigma, USA); Secondary antibodies used

for Western blotting were goat anti-mouse- and anti-rabbit-HRP (#172-1011 and #170-6515, Bio-Rad, USA).

*Cell culture and cell line construction*

| **Cell Lines** | **Source** |
|---|---|
| K562 KRAB-BFP-dCAS9 | Gilbert et al., 2014 |
| HEK 293T/17 | ATCC CRL-11268 |
| K562-CRISPRi-Tet-ON-(tet-GET1-GFP-P2A-RFP) | This study |
| HEK 293T TXNDC15 KO | This study |
| HEK 293T MARCHF6 KO | This study |
| HEK 293T SS-GFP-TXNDC15 | This study |
| HEK 293T MARCHF6-TEV-ALFA-P2A-BFP | This study |
| RPE1 KRAB-BFP-dCAS9 | Jost et al., 2017 |

HEK 293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 100 units/mL penicillin and 100 µg/mL streptomycin. K562 cells containing KRAB-BFP-dCas9 (Gilbert et al., 2014) were cultured in RPMI-1640 with 25 mM HEPES, 2.0 g/L NaHCO3, and 0.3 g/L L-glutamine supplemented with 10% Tet System Approved FBS, 2 mM glutamine, 100 units/mL penicillin, and 100 µg/mL streptomycin. Cells were maintained between $0.25 \times 10^6$ $–1 \times 10^6$ cells/mL. K562 and HEK293T cells were grown at 37 °C.

Clonal knockouts of TXNDC15 and MARCHF6 were obtained by transfecting HEK 293T cells with pX459 encoding the respective sgRNA using TransIT-293 transfection reagent (Mirus, USA). Single cells were sorted into 96-well plates using a SONY cell sorter (SH800S) 72 h post transfection. Clones were selected based on the stabilizing phenotype with the GET1-GFP-2A-RFP reporter.

*CRISPRi screens*

Fluorescent reporter CRISPRi screens were performed as previously described, with minor modifications (Gilbert et al., 2014; Horlbeck et al., 2016). CRISPRi dual libraries (Non-targeting dual library, or TXNDC15 dual library) were transduced at a multiplicity of infection less than one into 330 million K562-CRISPRi-Tet-ON cells containing the GET1-GFP-2A-RFP reporter. Cells were maintained in 1L spinner flasks (Bellco, SKU: 1965-61010) at a volume of 1L during the screens. BFP positive cells were between 25-30% 48 hours post-transfection. Cells were then treated with 1 µg/mL puromycin for three consecutive days to select for cells containing guides.

Following approximately two days of recovery after puromycin selection, the reporter was induced with doxycycline (100-1000 ng/mL) for 36 hours and sorted on a FACSAria Fusion Cell Sorter (BD Biosciences). To ensure that the culture was maintained at an average coverage of more than 1000 per sgRNA, cells were diluted to $0.5 \times 10^6$ cells/mL each day during the screen.

During the sorts, cells containing guides were gated using the BFP marker. Gates for RFP and GFP were used to select cells expressing the reporter. Cells were sorted based on the GFP:RFP ratio of the final gated population, where approximately 25 million cells with either the highest or lowest 30% GFP:RFP ratios were collected, pelleted, and flash-frozen. Extraction and purification of genomic DNA was performed using a Nucleospin Blood XL kit (Takara Bio, #740950.10) following the manufacturer's protocol. Guides were amplified and barcoded by PCR using NEB Next Ultra ii Q5 MM (M0544L). A unique forward index primer was used and a reverse primer that binds in the hU6 region upstream of the fixed guide in the dual guide vector was used (5'-CAAGCAGAAGACGGCATACGAGATGGAATCATGGGAAATAGGCCCTC), as previously described (Guna, Page, et al., 2023). The DNA library (349 bp) was purified using SPRISelect beads (Beckman Coulter B23317), and purified DNA was analyzed on an Agilent 2100 Bioanalyzer prior to sequencing using an Illumina HiSeq2500 using the standard CRISPRi-v2 library sequencing primer (5'-GTGTGTTTTGAGACTATAAGTATCCCTTGGAGAACCACCTTGTTG). Analysis of the screen sequencing results was performed using the pipeline in https://github.com/mhorlbeck/ScreenProcessing (Horlbeck et al., 2016). Coverage of the guide libraries was ensured by excluding from analysis guides with fewer than 50 counts. The strongest 3 sgRNA phenotypes were used to calculate the phenotype score for each gene. The Mann-Whitney p-value was calculated using the 5 sgRNAs targeting the same gene compared to the negative controls.

*Lentiviral Transduction*

Lentivirus was made by co-transfection of packaging plasmid psPAX2 and envelope plasmid VSV.G, along with a transfer plasmid of interest, using TransIT-293 transfection reagent (Mirus). Lentivirus was harvested 48 hours after transfection and stored at -80 °C for future use.

Spinfection with lenti-viral vectors containing programmed guides was performed in K562 dCas9-BFP-KRAB cells. Briefly, 250,000 cells were mixed with 200 µl of lentivirus and RPMI medium

in the presence of 8 µg/ml polybrene in a total volume of 1 ml. K562 cells in 12-well plates were spun at 1,000 xg for 1.5 h at 30˚C, resuspended, and cultured in 12-well or 6-well plates. Approximately 48 h after spinfection, 1 µg/ml puromycin was added for 3 consecutive days to select cells containing the guides. Cells recovered from puromycin treatment for 1.5 days and then were transduced with 50-100 µl of reporter lentivirus. 48h after reporter transfection, the cells were analyzed using flow cytometry, as described below.

For assays in adherent HEK293 cells, lenti-viral transduction of 100-250 µl lentiviral supernatant and 8 µg/ml polybrene (Millipore-Sigma, USA) were added to ~70% confluent cells in 2.5 ml culture medium in a 6-well. Lenti-viral reporter constructs contained an upstream CMV promoter in the pHAGE2 vector.

For rescue assays, $0.3x10^5$ WT HEK 293T or TXNDC15 KO HEK 293T cells were seeded into each 6-well plate. 24h later, cells were transduced with 300 µl lenti-virus of TXNDC15 rescue construct and 8 µg/ml final concentration of polybrene. 72h later, the media was exchanged to remove excess polybrene, and cells were transduced with 150 µl reporter lenti-virus the in presence of 8 µg/ml final concentration of polybrene. 48h after reporter transfection, the cells were harvested, washed, resuspended in 500 µl Dulbecco's Phosphate Buffered Saline (Gibco), and analyzed by flow cytometry or frozen for analysis via Western Blot.

*Flow cytometry*

HEK 293T cells were trypsinized, washed with 1xPBS, and resuspended in 1xPBS for flow cytometry analysis. K562 cells were analyzed directly from 6-well cultures. Cells were analyzed using an Attune NxT Flow Cytometer (Thermo Fisher Scientific, USA). Flow cytometry data was analyzed using FlowJo v10.8 Software (BD Life Sciences, USA) or by Python using the FlowCytometryTools package.

*Conjugation of the ALFA nanobody to HRP for Western blotting*

HRP-maleimide was coupled to the ALFA nanobody through a single engineered C-terminal cysteine residue, as described (Pleiner et al., 2018).

*Preparation of human ER microsomes*

HEK293T cells (WT or endogenous tagged GFP-TXNDC15) were harvested and washed in 1X PBS. Cells were resuspended in 4 times the pellet volume of sucrose buffer (10 mM HEPES, pH 7.5, 250 mM sucrose, 2 mM magnesium acetate, 1X cOmplete EDTA-free protease inhibitor cocktail [Roche]) and lysed by douncing at 4 °C. Cells were diluted to twice the lysis volume in sucrose buffer and pelleted at 3214 xg for 35 min. at 4 °C. Supernatant from the cell lysate was transferred to a new tube and pelleted again at 3214 xg for 35 min. at 4 °C. Samples were pelleted in an ultracentrifuge in an MLA80 rotor (Beckman-Coulter) at 75,000 xg for 1h at 4 °C to isolate the microsomal fraction. The microsomal pellet was resuspended in microsome buffer (10 mM HEPES, pH 7.5, 250 mM sucrose, 1 mM magnesium acetate, 0.5 mM DTT) to an A280 of 75. Microsomes (hRMs) were nucleased to remove contaminating RNAs. The hRMs were mixed with $CaCl_2$ (1 mM) and micrococcal nuclease (0.125 U/µL) on ice before incubating for 6 minutes at 25 °C. Nucleased hRMs were transferred to ice and immediately quenched with EGTA (2 mM). Nucleased hRMs were flash frozen and stored at -80 °C prior to use in *in vitro* translations.

*Mammalian in vitro translation*

Translation reactions were prepared using nucleased rabbit reticulocyte lysate (RRL) supplemented with human derived rough ER microsomes, as previously described (Sharma et al., 2010; Walter & Blobel, 1983). DNA templates for *in vitro* transcription were made by PCR using primers within the SP6 promoter (5' end) and after the stop codon (3' end). Transcription reactions were incubated at 37 °C for 2 hours, and then used directly in a translation reaction, which was incubated for 45 minutes at 32 °C.

*Purification of in vitro translation reactions*

GET1 was purified from *in vitro* translations and analyzed by mass spectrometry. GET1 contained a C-terminal ALFA tag and was purified directly from *in vitro* translations in RRL using an anti-ALFA nanobody strategy, as previously described (Stevens et al., 2023). Similarly, for experiments in which GET1 was assayed for interaction with GFP-TXNDC15, the anti-GFP nanobody strategy was used. For these experiments, GET1 variants also contained a C-terminal ALFA tag. Briefly, *in vitro* translation reactions were pelleted through a sucrose cushion (50 mM HEPES, pH 7.5, 2 mM MgAc, 100 mM KOAc, 20% sucrose [w/v]) for 20 min at 55K RPM at 4 °C in a TLA-55 rotor (Beckman-Coulter). The microsomal fraction was resuspended in half the translation volume of physiological salt buffer (50 mM HEPES, pH 7.5, 2 mM MgAc, 200 mM KOAc) followed by an

equal volume of 2X solubilization buffer (50 mM HEPES, pH 7.5, 2 mM MgAc, 300 mM KOAc, 2% [w/v] GDN). Samples were solubilized for 20 min. at 4 °C and then pelleted at 14000 xg for 10 min. at 4 °C to remove insoluble debris. Solubilized samples were added to magnetic beads containing anti-ALFA nanobody (for mass spectrometry) or anti-GFP nanobody (for interaction analysis), as described below.

Pierce magnetic Streptavidin beads (Thermo Fisher Scientific) were equilibrated in wash buffer (50 mM HEPES, pH 7.5, 2 mM MgAc, 100 mM KOAc, 0.25% [w/v] GDN) followed by incubation with biotinylated anti-ALFA tag nanobody or biotinylated anti-GFP nanobody in wash buffer for 20 min at 4 °C. To block free biotin binding sites, magnetic beads were incubated with wash buffer containing 10 µM $dPEG_{24}$-biotin acid (Quanta Biodesign) for 5-10 min at 4 °C. Beads were washed 2X with wash buffer before mixing with solubilized *in vitro* translation samples. Samples were incubated with head-over-tail mixing at 4˚C for 2 h. Unbound solubilized samples were removed from the beads and then the magnetic beads were washed three times with wash buffer and resuspended in wash buffer containing 250 nM $SENP^{EuB}$ protease in a volume amounting to one tenth or one twentieth of the original *in vitro* reaction volume. Protease elution proceeded for 30 min on ice, and the eluate was analyzed using mass spectrometry or by using SDS-PAGE and autoradiography.

*Whole-cell proteomics*

K562 cells were spinfected with guides targeting TXNDC15, MARCHF6, or a non-targeting control in biological quadruplicate, as described above. After 8 days of knockdown, cells were harvested and washed in 1X PBS followed by trypsin digestion. Samples were then analyzed using mass spectrometry. Briefly, samples were analyzed by Eclipse mass spectrometer coupled to Vanquish Neo. Peptides were separated on an Aurora UHPLC Column (60 cm × 75 µm, 1.7 µm C18, AUR3-60075C18, Ion Opticks) with a flow rate of 0.3 µL/min for a total duration of 135 min. The Raw data files were searched against Uniprot human database UP000005640 using the Proteome Discoverer 3.0 software based on the Chimerys algorithm. Oxidation / +15.995 Da (M) were set as dynamic modifications, and carbamidomethylation / +57.021 Da(C) was fixed modification. The fragment mass tolerance was set to 20 ppm. The maximum false peptide discovery rate was specified as 0.01 using the Percolator Node validated by q-value.

*Chapter 6*

# CONCLUSION

The fundamental question of how a cell makes a membrane protein has been studied for several decades. During this time, our understanding of this process has evolved, especially in the most recent years. Advances in technologies such as functional genomics and single particle cryo-EM have enabled greater understanding of protein biogenesis and allowed us to examine in unprecedented detail and depth how the diversity of the membrane proteome is accommodated by the insertion machinery at the ER. For example, within the last twenty years, we have gained insight into the structure of the Sec61 translocation channel, which helped guide the field's understanding of how membrane proteins can cross and enter the ER. Additionally, in the last three years alone, structures of other critical biogenesis machineries at the ER have emerged, including those of the EMC and the multipass translocon. These structures have not only enabled us to understand in greater detail how membrane proteins are made, but also have opened up more questions that need to be answered.

A main question, which my thesis work addressed, is how these different insertases and protein machineries cooperate during the process of membrane protein biogenesis. Building on previous knowledge that the EMC is required for the biogenesis of several classes of membrane proteins, we showed that for many multipass membrane proteins, the EMC needs to coordinate with components of the multipass translocon. Specifically, we showed that the EMC and the BOS complex physically associate, and we structurally characterized this holocomplex. We mechanistically dissected the requirements for the BOS complex functioning at EMC and showed that charges in the N-terminal soluble domain of $N_{exo}$ membrane proteins dictate this requirement. This means that residues sometimes as far as 50 residues away from the TMD of a multipass membrane protein can influence its path into the ER bilayer. Indeed, the most energetically costly step during membrane protein insertion is the translocation of soluble domains across the lipid bilayer.

We show that not one pathway is sufficient for all membrane proteins, and instead the biophysical properties of the protein dictate the requirements for insertion at the ER. This is a fundamental difference to the prevailing models in the field. However, while my thesis research addressed

several unanswered questions and challenged the prevailing models, it has also highlighted gaps in our understanding of the biogenesis process. For example, it is not clear how ribosome-nascent chain complexes arrive at the EMC. More research needs to be done to elucidate how the EMC gets a first pass at nascent chains targeted via SRP, which will clarify how this localization step occurs. Further, how the EMC hands off the ribosome nascent chain complex to the multipass translocon after the insertion of TMD1 of its clientele is unclear. We show that individual subunits of the EMC directly interact with components of the multipass translocon, but more mechanistic work needs to be done to define how the handoff occurs.

In addition to addressing major questions within the field of membrane protein biogenesis, my thesis work aimed to further our understanding of the quality control of multi-subunit complex assembly. I focused on a model ER-resident complex, GET1/GET2, and sought to understand how GET1 is identified and degraded in the cell. The ultimate goal of this research was to develop an understanding of how the cell regulates complex assembly, particularly when the assembly step fails and subunits are orphaned in the membrane. Through this research, I identified a novel quality control factor, TXNDC15,  and I used genetic modifier screens to place it into the known MARCHF6 ERAD pathway. This expands what we know about MARCHF6, as most work on ERAD has been performed in yeast. Moreover, I showed that MARCHF6-TXNDC15 function in parallel to the HRD1 ERAD pathway to clear orphaned GET1, which suggests some level of redundancy in the E3 ligases in the human ER. Further, I identified endogenous substrates of MARCHF6-TXNDC15, many of which are critical cytoskeletal and ciliary proteins. Interestingly, mutations within TXNDC15 are linked to ciliopathies, pathologies associated with improper development of the primary cilium, and these mutations also impair TXNDC15's ability to engage in quality control. This finding links TXNDC15's role in quality control to known disease phenotypes in humans.

My work in this area suggests that there are likely many additional factors at play in the various branches of ERAD within higher eukaryotes. Like TXNDC15, there are probably adapters to other E3 ligases in the ER that enable their substrate specificity and that allow them to carry out their quality control functions. We know that the quality control systems must be robust in order to maintain cellular proteostasis, particularly in organisms such as humans that have cells that exist throughout the lifespan. How the cellular machinery enables this is still incompletely understood. Genome-wide functional genomic approaches may be one useful tool to address this gap in the

field. Altogether, a greater knowledge of the full repertoire of factors involved in ERAD would allow us to clarify how the ERAD pathways function to maintain the proteostasis of the cell.

# BIBLIOGRAPHY

Adle, D. J., & Lee, J. (2008). Expressional control of a cadmium-transporting P1B-type ATPase by a metal sensing degradation signal. *The Journal of Biological Chemistry*, *283*(46), 31460–31468. https://doi.org/10.1074/jbc.M806054200

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), Article 6. https://doi.org/10.1107/S2059798318006551

Alexiev, B. A., Lin, X., Sun, C.-C., & Brenner, D. S. (2006). Meckel-Gruber syndrome: Pathologic manifestations, minimal diagnostic criteria, and differential diagnosis. *Archives of Pathology & Laboratory Medicine*, *130*(8), 1236–1238. https://doi.org/10.5858/2006-130-1236-MS

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

Arnér, E. S. J., & Holmgren, A. (2000). Physiological functions of thioredoxin and thioredoxin reductase. *European Journal of Biochemistry*, *267*(20), 6102–6109. https://doi.org/10.1046/j.1432-1327.2000.01701.x

Babu, M., Vlasblom, J., Pu, S., Guo, X., Graham, C., Bean, B. D. M., Burston, H. E., Vizeacoumar, F. J., Snider, J., Phanse, S., Fong, V., Tam, Y. Y. C., Davey, M., Hnatshak, O., Bajaj, N., Chandran, S., Punna, T., Christopolous, C., Wong, V., … Greenblatt, J. F. (2012). Interaction landscape of membrane-protein complexes in Saccharomyces cerevisiae. *Nature*, *489*(7417), Article 7417. https://doi.org/10.1038/nature11354

Bai, L., You, Q., Feng, X., Kovach, A., & Li, H. (2020). Structure of the ER membrane complex, a transmembrane-domain insertase. *Nature*, *584*(7821), Article 7821. https://doi.org/10.1038/s41586-020-2389-3

Bañó-polo, M., Martínez-garay, C. A., Grau, B., Martínez-gil, L., & Mingarro, I. (2017). Membrane insertion and topology of the translocon-associated protein (TRAP) gamma subunit. *BBA - Biomembranes*, *1859*(5), 903–909. https://doi.org/10.1016/j.bbamem.2017.01.027

Bar-Nun, S. (2005). The role of p97/Cdc48p in endoplasmic reticulum-associated degradation: From the immune system to yeast. *Current Topics in Microbiology and Immunology*, *300*, 95–125. https://doi.org/10.1007/3-540-28007-3_5

Beguin, P., Hasler, U., Beggah, A., Horisberger, J.-D., & Geering, K. (1998). Membrane Integration of Na,K-ATPase a-Subunits and b-Subunit Assembly. *Journal of Biological Chemistry*, *273*(38), 24921–24931.

Benschop, J. J., Brabers, N., van Leenen, D., Bakker, L. V., van Deutekom, H. W. M., van Berkum, N. L., Apweiler, E., Lijnzaad, P., Holstege, F. C. P., & Kemmeren, P. (2010). A Consensus of Core Protein Complex Compositions for Saccharomyces cerevisiae. *Molecular Cell*, *38*(6), 916–928. https://doi.org/10.1016/j.molcel.2010.06.002

Bergeron, J. J., Brenner, M. B., Thomas, D. Y., & Williams, D. B. (1994). Calnexin: A membrane-bound chaperone of the endoplasmic reticulum. *Trends in Biochemical Sciences*, *19*(3), 124–128. https://doi.org/10.1016/0968-0004(94)90205-4

Bonifacino, J. S., Cosson, P., & Klausner, R. D. (1990). Colocalized Transmembrane Determinants for ER Degradation and Subunit Assembly Explain the Intracellular Fate of TCR Chains. *Cell*, *63*, 503–513.

Bonifacino, J. S., Cosson, P., Shah, N., & Klausner, R. D. (1991). Role of potentially charged transmembrane residues in targeting proteins for retention and degradation within the endoplasmic reticulum. *EMBO*, *10*(10), 2783–2793.

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 (London, England: 1993)*, *23*(9), 1715–1724. https://doi.org/10.1016/j.str.2015.06.025

Breslow, D. K., Hoogendoorn, S., Kopp, A. R., Morgens, D. W., Vu, B. K., Kennedy, M. C., Han, K., Li, A., Hess, G. T., Bassik, M. C., Chen, J. K., & Nachury, M. V. (2018). A CRISPR-based screen for Hedgehog signaling provides insights into ciliary function and ciliopathies. *Nature Genetics*, *50*(3), Article 3. https://doi.org/10.1038/s41588-018-0054-7

Burr, M. L., Cano, F., Svobodova, S., Boyle, L. H., Boname, J. M., & Lehner, P. J. (2011). HRD1 and UBE2J1 target misfolded MHC class I heavy chains for endoplasmic reticulum-associated degradation. *Proceedings of the National Academy of Sciences*, *108*(5), 2034–2039. https://doi.org/10.1073/pnas.1016229108

Cabantous, S., Terwilliger, T. C., & Waldo, G. S. (2004). 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., Bondio, A. Del, 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. *Nature Scientific Reports*, *9*(August), 1–12. 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), 1055–1064. https://doi.org/10.1002/pro.3791

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, B., Retzlaff, M., Roos, T., & Frydman, J. (2011). Cellular Strategies of Protein Quality Control. *Cold Spring Harbor Perspectives in Biology*, *3*(8), a004374. https://doi.org/10.1101/cshperspect.a004374

Chen, Z., Mondal, A., Abderemane-Ali, F., 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*, *619*(7969), Article 7969. https://doi.org/10.1038/s41586-023-06175-5

Cheng, S. H., Gregory, R. J., Marshall, J., Paul, S., Souza, D. W., White, G. A., O'Riordan, C. R., & Smith, A. E. (1990). Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell*, *63*(4), 827–834. https://doi.org/10.1016/0092-8674(90)90148-8

Chih, B., Liu, P., Chinn, Y., Chalouni, C., Komuves, L. G., Hass, P. E., Sandoval, W., & Peterson, A. S. (2012). A ciliopathy complex at the transition zone protects the cilia as a privileged membrane domain. *Nature Cell Biology*, *14*(1), Article 1. https://doi.org/10.1038/ncb2410

Chiti, F., & Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annual Review of Biochemistry*, *75*, 333–366. https://doi.org/10.1146/annurev.biochem.75.101304.123901

Chitwood, P. J., & Hegde, R. S. (2020). An intramembrane chaperone complex facilitates membrane protein biogenesis. *Nature*, *584*(7822), Article 7822. https://doi.org/10.1038/s41586-020-2624-y

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

Christianson, J. C., & Carvalho, P. (2022). Order through destruction: How ER-associated protein degradation contributes to organelle homeostasis. *The EMBO Journal*, *41*(6), e109845. https://doi.org/10.15252/embj.2021109845

Christianson, J. C., Olzmann, J. A., Shaler, T. A., Sowa, M. E., Bennett, E. J., Richter, C. M., Tyler, R. E., Greenblatt, E. J., Harper, J. W., & 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

Christianson, J. C., & Ye, Y. (2014). Cleaning up in the endoplasmic reticulum: Ubiquitin in charge. *Nature Structural & Molecular Biology*, *21*(4), 325–335. https://doi.org/10.1038/nsmb.2793

Chua, N. K., Howe, V., Jatana, N., Thukral, L., & Brown, A. J. (2017). A conserved degron containing an amphipathic helix regulates the cholesterol-mediated turnover of human squalene monooxygenase, a rate-limiting enzyme in cholesterol synthesis. *The Journal of Biological Chemistry*, *292*(49), 19959–19973. https://doi.org/10.1074/jbc.M117.794230

Colombo, S. F., Cardani, S., Maroli, A., Vitiello, A., Soffientini, P., Crespi, A., Bram, R. F., Benfante, R., & Borgese, N. (2016). Tail-anchored protein insertion in mammals function and reciprocal interactions of the two subunits of the TRC40 receptor. *Journal of Biological Chemistry*, *291*(29), 15292–15306. https://doi.org/10.1074/jbc.M115.707752

Cvicek, V., Iii, W. A. G., & Abrol, R. (2016). Structure-Based Sequence Alignment of the Transmembrane Domains of All Human GPCRs: Phylogenetic, Structural and Functional Implications. *PLOS Computational Biology*, *12*(3), e1004805. https://doi.org/10.1371/journal.pcbi.1004805

David, A., Razali, R., Wass, M. N., & Sternberg, M. J. E. (2012). Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. *Human Mutation*, *33*(2), 359–363. https://doi.org/10.1002/humu.21656

Dephoure, N., Hwang, S., Sullivan, C. O., Dodgson, S. E., Gygi, S. P., Amon, A., & Torres, E. M. (2014). Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. *ELIFE*, 1–27. https://doi.org/10.7554/eLife.03023

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 *. *Journal of Biological Chemistry*, *285*(34), 26174–26181. https://doi.org/10.1074/jbc.M110.132548

Doblmann, J., Dusberger, F., Imre, R., Hudecz, O., Stanek, F., Mechtler, K., & Dürnberger, G. (2019). apQuant: Accurate Label-Free Quantification by Quality Filtering. *Journal of Proteome Research*, *18*(1), 535–541. https://doi.org/10.1021/acs.jproteome.8b00113

Dolan, K. A., Dutta, M., Kern, D. M., Kotecha, A., Voth, G. A., & Brohawn, S. G. (2022). Structure of SARS-CoV-2 M protein in lipid nanodiscs. *ELife*, *11*, e81702. https://doi.org/10.7554/eLife.81702

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

Enquist, K., Fransson, M., Boekel, C., Bengtsson, I., Geiger, K., Lang, L., Pettersson, A., Johansson, S., von Heijne, G., & Nilsson, I. (2009). Membrane-integration Characteristics of Two ABC Transporters, CFTR and P-glycoprotein. *Journal of Molecular Biology*, *387*(5), 1153–1164. https://doi.org/10.1016/j.jmb.2009.02.035

Feige, M. J., & Hendershot, L. M. (2013). Quality Control of Integral Membrane Proteins by Assembly-Dependent Membrane Integration. *Molecular Cell*, *51*, 297–309. https://doi.org/10.1016/j.molcel.2013.07.013

Furth, N., Gertman, O., Shiber, A., Alfassy, O. S., Cohen, I., Rosenberg, M. M., Doron, N. K., Friedler, A., & Ravid, T. (2011). Exposure of bipartite hydrophobic signal triggers nuclear quality control of Ndc10 at the endoplasmic reticulum/nuclear envelope. *Molecular Biology of the Cell*, *22*(24), 4726–4739. https://doi.org/10.1091/mbc.E11-05-0463

Gavin, A.-C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dümpelfeld, B., Edelmann, A., Heurtier, M.-A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A.-M., Schelder, M., Schirle, M., … Superti-Furga, G. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature*, *440*(7084), 631–636. https://doi.org/10.1038/nature04532

Gemmer, M., Chaillet, M. L., van Loenhout, J., Cuevas Arenas, R., Vismpas, D., Gröllers-Mulderij, M., Koh, F. A., Albanese, P., Scheltema, R. A., Howes, S. C., Kotecha, A., Fedry, J., & Förster, F. (2023). Visualization of translation and protein biogenesis at the ER membrane. *Nature*, *614*(7946), Article 7946. https://doi.org/10.1038/s41586-022-05638-5

Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., Guimaraes, C., Panning, B., Ploegh, H. L., Bassik, M. C., Qi, L. S., Kampmann, M., & Weissman, J. S. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*, *159*(3), 647–661. https://doi.org/10.1016/j.cell.2014.09.029

Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., & Ferrin, T. E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science*, *27*(1), 14–25. https://doi.org/10.1002/pro.3235

Görlich, D., Hartmann, E., Prehn, S., & Rapoport, T. A. (1992). A protein of the endoplasmic reticulum involved early in polypeptide translocation. *Nature*, *357*(6373), Article 6373. https://doi.org/10.1038/357047a0

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*, *10*(1), 4403. https://doi.org/10.1038/s41467-019-12301-7

Graf, G. A., Yu, L., Li, W.-P., Gerard, R., Tuma, P. L., Cohen, J. C., & Hobbs, H. H. (2003). ABCG5 and ABCG8 are obligate heterodimers for protein trafficking and biliary cholesterol excretion. *The Journal of Biological Chemistry*, *278*(48), 48275–48282. https://doi.org/10.1074/jbc.M310223200

Guna, A., Hazu, M., Pinton Tomaleri, G., & 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: CB*, *28*(8), R498–R511. https://doi.org/10.1016/j.cub.2018.02.004

Guna, A., Page, K. R., Replogle, J. R., Esantsi, T. K., Wang, M. L., Weissman, J. S., & Voorhees, R. M. (2023). *A dual sgRNA library design to probe genetic modifiers using genome-wide CRISPRi screens* (p. 2023.01.22.525086). bioRxiv. https://doi.org/10.1101/2023.01.22.525086

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

Güngör, B., Flohr, T., Garg, S. G., & Herrmann, J. M. (2022). The ER membrane complex (EMC) can functionally replace the Oxa1 insertase in mitochondria. *PLoS Biology*, *20*(3), e3001380. https://doi.org/10.1371/journal.pbio.3001380

Habeck, G., Ebner, F. A., Shimada-Kreft, H., & Kreft, S. G. (2015). The yeast ERAD-C ubiquitin ligase Doa10 recognizes an intramembrane degron. *The Journal of Cell Biology*, *209*(2), 261–273. https://doi.org/10.1083/jcb.201408088

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

Hampton, R. Y., & Sommer, T. (2012). Finding the will and the way of ERAD substrate retrotranslocation. *Current Opinion in Cell Biology*, *24*(4), 460–466. https://doi.org/10.1016/j.ceb.2012.05.010

Harper, J. W., & Bennett, E. J. (2016). Proteome complexity and the forces that drive proteome imbalance. *Nature*, *537*(7620), 328–338. https://doi.org/10.1038/nature19947

Harrison, S. C. (2008). Viral membrane fusion. *Nature Structural & Molecular Biology*, *15*(7), Article 7. https://doi.org/10.1038/nsmb.1456

Hartmann, E., Görlich, D., Kostka, S., Otto, A., Kraft, R., Knespel, S., Bürger, E., Rapoport, T. A., & Prehn, S. (1993). A tetrameric complex of membrane proteins in the endoplasmic

reticulum. *European Journal of Biochemistry*, *214*(2), 375–381. https://doi.org/10.1111/j.1432-1033.1993.tb17933.x

Hassink, G., Kikkert, M., van Voorden, S., Lee, S.-J., Spaapen, R., van Laar, T., Coleman, C. S., Bartee, E., Früh, K., Chau, V., & Wiertz, E. (2005). TEB4 is a C4HC3 RING finger-containing ubiquitin ligase of the endoplasmic reticulum. *The Biochemical Journal*, *388*(Pt 2), 647–655. https://doi.org/10.1042/BJ20041241

Havugimana, P. C., Hart, G. T., Nepusz, T., Yang, H., Turinsky, A. L., Li, Z., Wang, P. I., Boutz, D. R., Fong, V., Phanse, S., Babu, M., Craig, S. A., Hu, P., Wan, C., Vlasblom, J., Dar, V.-N., Bezginov, A., Clark, G. W., Wu, G. C., … Emili, A. (2012). A census of human soluble protein complexes. *Cell*, *150*(5), 1068–1081. https://doi.org/10.1016/j.cell.2012.08.011

Hegde, R. S., & Keenan, R. J. (2022). The mechanisms of integral membrane protein biogenesis. *Nature Reviews Molecular Cell Biology*, *23*(2), Article 2. https://doi.org/10.1038/s41580-021-00413-2

Hegde, R. S., & Lingappa, V. R. (1999). Regulation of protein biogenesis at the endoplasmic reticulum membrane. *Trends in Cell Biology*, *8924*(99), 132–137.

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

Heldin, C.-H., Lu, B., Evans, R., & Gutkind, J. S. (2016). Signals and Receptors. *Cold Spring Harbor Perspectives in Biology*, *8*(4), a005900. https://doi.org/10.1101/cshperspect.a005900

Hennon, S. W., Soman, R., Zhu, L., & Dalbey, R. E. (2015). YidC/Alb3/Oxa1 Family of Insertases. *The 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., & Heijne, G. Von. (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature*, *450*, 1026–1030. https://doi.org/10.1038/nature06387

Hirsch, C., Gauss, R., Horn, S. C., Neuber, O., & Sommer, T. (2009). The ubiquitylation machinery of the endoplasmic reticulum. *Nature*, *458*(7237), 453–460. https://doi.org/10.1038/nature07962

Holmes, K. C., Popp, D., Gebhard, W., & Kabsch, W. (1990). Atomic model of the actin filament. *Nature*, *347*(6288), 44–49. https://doi.org/10.1038/347044a0

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*, e19760. https://doi.org/10.7554/eLife.19760

Huttlin, E. L., Bruckner, R. J., Paulo, J. A., Cannon, J. R., Ting, L., Baltier, K., Colby, G., Gebreab, F., Parzen, H., Szpyt, J., Tam, S., Zarraga, G., Pontano-vaites, L., Swarup, S., White, A. E., Schweppe, D. K., Rad, R., Erickson, B. K., Obar, R. A., … Harper, J. W. (2017). Architecture of the human interactome defines protein communities and disease networks. *Nature*, *545*(7655), 505–509. https://doi.org/10.1038/nature22366

Hyun, S., Maruri-avidal, L., & Moss, B. (2015). Topology of Endoplasmic Reticulum-Associated Cellular and Viral Proteins Determined with Split-GFP. *Traffic*, *13*, 787–795. https://doi.org/10.1111/tra.12281

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

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

Jaskolowski, M., Jomaa, A., Gamerdinger, M., Shrestha, S., Leibundgut, M., Deuerling, E., & Ban, N. (2023). Molecular basis of the TRAP complex function in ER protein biogenesis. *Nature Structural & Molecular Biology*, *30*(6), Article 6. https://doi.org/10.1038/s41594-023-00990-0

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 (New York, N.Y.)*, *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. (2017). Initiation of Quality Control during Poly (A) Translation Requires Site-Specific Ribosome Ubiquitination. *Molecular Cell*, *65*(4), 743-750.e4. https://doi.org/10.1016/j.molcel.2016.11.039

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

Kalies, K.-U., Rapoport, T. A., & Hartmann, E. (1998). The β Subunit of the Sec61 Complex Facilitates Cotranslational Protein Transport and Interacts with the Signal Peptidase

during Translocation. *Journal of Cell Biology*, *141*(4), 887–894. https://doi.org/10.1083/jcb.141.4.887

Kamiyama, D., Sekine, S., Barsi-Rhyne, B., Hu, J., Chen, B., Gilbert, L. A., Ishikawa, H., Leonetti, M. D., Marshall, W. F., Weissman, J. S., & Huang, B. (2016). Versatile protein tagging in cells with split fluorescent protein. *Nature Communications*, *7*(1), Article 1. https://doi.org/10.1038/ncomms11046

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., … MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*(7809), Article 7809. https://doi.org/10.1038/s41586-020-2308-7

Kelleher, D. J., & Gilmore, R. (2006). An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology*, *16*(4), 47R-62R. https://doi.org/10.1093/glycob/cwj066

Kern, D. M., Sorum, B., Mali, S. S., Hoel, C. M., Sridharan, S., Remis, J. P., Toso, D. B., Kotecha, A., Bautista, D. M., & Brohawn, S. G. (2021). Cryo-EM structure of SARS-CoV-2 ORF3a in lipid nanodiscs. *Nature Structural & Molecular Biology*, *28*(7), Article 7. https://doi.org/10.1038/s41594-021-00619-0

Kikkert, M., Doolman, R., Dai, M., Avner, R., Hassink, G., Voorden, S. van, Thanedar, S., Roitelman, J., Chau, V., & Wiertz, E. (2004). Human HRD1 Is an E3 Ubiquitin Ligase Involved in Degradation of Proteins from the Endoplasmic Reticulum *. *Journal of Biological Chemistry*, *279*(5), 3525–3534. https://doi.org/10.1074/jbc.M307453200

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., &

Rothbauer, U. (2009). Modulation of protein properties in living cells using nanobodies. *Nature Structural & Molecular Biology*, *17*(1), 133–138. https://doi.org/10.1038/nsmb.1727

Koch, H. P., Kurokawa, T., Okochi, Y., Sasaki, M., Okamura, Y., & Larsson, H. P. (2008). Multimeric nature of voltage-gated proton channels. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(26), 9111–9116. https://doi.org/10.1073/pnas.0801553105

Komander, D., & Rape, M. (2012). The Ubiquitin Code. *Annual Review of Biochemistry*, *81*(1), 203–229. https://doi.org/10.1146/annurev-biochem-060310-170328

Kreft, S. G., Wang, L., & Hochstrasser, M. (2006). Membrane Topology of the Yeast Endoplasmic Reticulum-localized Ubiquitin Ligase Doa10 and Comparison with Its Human Ortholog TEB4 (MARCH-VI) *. *Journal of Biological Chemistry*, *281*(8), 4646–4653. https://doi.org/10.1074/jbc.M512215200

Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., Punna, T., Peregrín-Alvarez, J. M., Shales, M., Zhang, X., Davey, M., Robinson, M. D., Paccanaro, A., Bray, J. E., Sheung, A., … Greenblatt, J. F. (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature*, *440*(7084), 637–643. https://doi.org/10.1038/nature04670

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes11Edited by F. Cohen. *Journal of Molecular Biology*, *305*(3), 567–580. https://doi.org/10.1006/jmbi.2000.4315

Krshnan, L., Weijer, M. L. van de, & Carvalho, P. (2022). Endoplasmic Reticulum–Associated Protein Degradation. *Cold Spring Harbor Perspectives in Biology*, a041247. https://doi.org/10.1101/cshperspect.a041247

Kumazaki, K., Chiba, S., Takemoto, M., Furukawa, A., Nishiyama, K., 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), Article 7501. https://doi.org/10.1038/nature13167

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

Lambacher, N. J., Bruel, A.-L., van Dam, T. J. P., Szymańska, K., Slaats, G. G., Kuhns, S., McManus, G. J., Kennedy, J. E., Gaff, K., Wu, K. M., van der Lee, R., Burglen, L., Doummar, D., Rivière, J.-B., Faivre, L., Attié-Bitach, T., Saunier, S., Curd, A., Peckham, M., … Blacque, O. E. (2016). TMEM107 recruits ciliopathy proteins to subdomains of the ciliary transition zone and causes Joubert syndrome. *Nature Cell Biology*, *18*(1), 122–131. https://doi.org/10.1038/ncb3273

Lansbury, P. T., & Lashuel, H. A. (2006). A century-old debate on protein aggregation and neurodegeneration enters the clinic. *Nature*, *443*(7113), Article 7113. https://doi.org/10.1038/nature05290

Lenard, J. (2008). Viral Membranes. *Encyclopedia of Virology*, 308–314. https://doi.org/10.1016/B978-012374410-4.00530-6

Li, W., Tu, D., Brunger, A. T., & Ye, Y. (2007). A ubiquitin ligase transfers preformed polyubiquitin chains from a conjugating enzyme to a substrate. *Nature*, *446*(7133), Article 7133. https://doi.org/10.1038/nature05542

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, 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*(10), Article 10. https://doi.org/10.1107/S2059798319011471

Lilley, B. N., & Ploegh, H. L. (2005). Multiprotein complexes that link dislocation, ubiquitination, and extraction of misfolded proteins from the endoplasmic reticulum membrane. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(40), 14296–14301. https://doi.org/10.1073/pnas.0505014102

Lin, H., Li, S., & Shu, H.-B. (2019). The Membrane-Associated MARCH E3 Ligase Family: Emerging Roles in Immune Regulation. *Frontiers in Immunology*, *10*. https://www.frontiersin.org/articles/10.3389/fimmu.2019.01751

Lin, L., DeMartino, G. N., & Greene, W. C. (2000). Cotranslational dimerization of the Rel homology domain of NF-κB1 generates p50–p105 heterodimers and is required for effective p50 production. *The EMBO Journal*, *19*(17), 4712–4722. https://doi.org/10.1093/emboj/19.17.4712

Lippincott-schwartz, J., Bonifacino, J. S., Yuan, L. C., & Klausner, R. D. (1988). Degradation from the Endoplasmic Reticulum: Disposing of Newly Synthesized Proteins. *Cell*, *54*, 209–220.

Lu, Y., Turnbull, I. R., Bragin, A., Carveth, K., & Skach, W. R. (2000). Reorientation of Aquaporin-1 Topology during Maturation in the Endoplasmic Reticulum. *Molecular Biology of the Cell*, *11*(September), 2973–2985.

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), 61–66. https://doi.org/10.1038/nature10362

Marsh, J. A., & Teichmann, S. A. (2015). Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annual Review of Biochemistry*, *84*, 551–575. https://doi.org/10.1146/annurev-biochem-060614-034142

Martin, J. L. (1995). Thioredoxin—A fold for all reasons. *Structure*, *3*(3), 245–250. https://doi.org/10.1016/S0969-2126(01)00154-X

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

Mateja, A., Szlachcic, A., Downing, M. E., Dobosz, M., Mariappan, M., Hegde, R. S., & Keenan, R. J. (2009). The structural basis of tail-anchored membrane protein recognition by Get3. *Nature*, *461*(7262), Article 7262. https://doi.org/10.1038/nature08319

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*, *0*(0). 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*, e56889. 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

McShane, E., Sin, C., Zauber, H., Wells, J. N., Donnelly, N., Wang, X., Hou, J., Chen, W., Storchova, Z., Marsh, J. A., Valleriani, A., & Selbach, M. (2016). Kinetic Analysis of

Protein Stability Reveals Age-Dependent Degradation. *Cell*, *167*(3), 803-815.e21. https://doi.org/10.1016/j.cell.2016.09.015

Meacock, S. L., Lecomte, F. J. L., Crawshaw, S. G., & High, S. (2002). Different Transmembrane Domains Associate with Distinct Endoplasmic Reticulum Components during Membrane Integration of a Polytopic Protein. *Molecular Biology of the Cell*, *13*(12), 4114–4129. https://doi.org/10.1091/mbc.e02-04-0198

Merkwirth, C., & Langer, T. (2009). Prohibitin function within mitochondria: Essential roles for cell proliferation and cristae morphogenesis. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1793*(1), 27–32. https://doi.org/10.1016/j.bbamcr.2008.05.013

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

Meyer, H., Bug, M., & Bremer, S. (2012). Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. *Nature Cell Biology*, *14*(2), Article 2. https://doi.org/10.1038/ncb2407

Miller-Vedam, L. E., Bräuning, B., Popova, K. D., Schirle Oakdale, N. T., 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*, e62611. https://doi.org/10.7554/eLife.62611

Minami, Y., Weissman, A. M., Samelson, L. E., & Klausner, R. D. (1987). Building a multichain receptor: Synthesis, degradation, and assembly of the T-cell antigen receptor. *Proceedings of the National Academy of Sciences USA*, *84*(May), 2688–2692.

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

Mueller, S., Wahlander, A., Selevsek, N., Otto, C., Ngwa, E. M., Poljak, K., Frey, A. D., Aebi, M., & Gauss, R. (2015). Protein degradation corrects for imbalanced subunit stoichiometry in OST complex assembly. *Molecular Biology of the Cell*, *26*(14), 2596–2608. https://doi.org/10.1091/mbc.E15-03-0168

Neal, S., Jaeger, P. A., Duttke, S. H., Benner, C., K.Glass, C., Ideker, T., & Hampton, R. Y. (2018). The Dfm1 Derlin Is Required for ERAD Retrotranslocation of Integral Membrane Proteins. *Molecular Cell*, *69*(2), 306-320.e4. https://doi.org/10.1016/j.molcel.2017.12.012

Nguyen, K. T., Mun, S.-H., Yang, J., Lee, J., Seok, O.-H., Kim, E., Kim, D., An, S. Y., Seo, D.-Y., Suh, J.-Y., Lee, Y., & Hwang, C.-S. (2022). The MARCHF6 E3 ubiquitin ligase acts as an NADPH sensor for the regulation of ferroptosis. *Nature Cell Biology*, *24*(8), Article 8. https://doi.org/10.1038/s41556-022-00973-1

Nicholls, C. D., McLure, K. G., Shields, M. A., & Lee, P. W. K. (2002). Biogenesis of p53 involves cotranslational dimerization of monomers and posttranslational dimerization of dimers. Implications on the dominant negative effect. *The Journal of Biological Chemistry*, *277*(15), 12937–12945. https://doi.org/10.1074/jbc.M108815200

Niesen, M. J. M., Zimmer, M. H., & Miller, T. F. I. (2020). Dynamics of Co-translational Membrane Protein Integration and Translocation via the Sec Translocon. *Journal of the American Chemical Society*, *142*(12), 5449–5460. https://doi.org/10.1021/jacs.9b07820

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*, e57887. https://doi.org/10.7554/eLife.57887

O'Hayre, M., Degese, M. S., & Gutkind, J. S. (2014). Novel insights into G protein and G protein-coupled receptor signaling in cancer. *Current Opinion in Cell Biology*, *27*, 126–135. https://doi.org/10.1016/j.ceb.2014.01.005

Olzmann, J. A., Richter, C. M., & Kopito, R. R. (2013). Spatial regulation of UBXD8 and p97/VCP controls ATGL-mediated lipid droplet turnover. *Proceedings of the National Academy of Sciences*, *110*(4), 1345–1350. https://doi.org/10.1073/pnas.1213738110

Paatero, A. O., Kellosalo, J., Dunyak, B. M., Almaliti, J., Gestwicki, J. E., Gerwick, W. H., Taunton, J., & Paavilainen, V. O. (2016). Apratoxin Kills Cells by Direct Blockade of the Sec61 Protein Translocation Channel. *Cell Chemical Biology*, *23*(5), 561–566. https://doi.org/10.1016/j.chembiol.2016.04.008

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. *The Journal of Cell Biology*, *217*(3), 1143–1154. https://doi.org/10.1083/jcb.201709115

Pleiner, T., Bates, M., Trakhanov, S., Lee, C., Stark, H., Urlaub, H., Schliep, J. E., Chug, H., Bohning, M., Stark, H., Urlaub, H., & Gorlich, D. (2015). Nanobodies: Site-specific labeling for super-resolution imaging, rapid epitope-mapping and native protein complex isolation. *ELIFE*, 1–21. https://doi.org/10.7554/eLife.11349

Pleiner, T., Hazu, M., Pinton Tomaleri, G., Nguyen, V. N., Januszyk, K., & Voorhees, R. M. (2023). A selectivity filter in the ER membrane protein complex limits protein misinsertion at the ER. *The Journal of Cell Biology*, *222*(8), e202212007. https://doi.org/10.1083/jcb.202212007

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), Article 3. https://doi.org/10.1038/nmeth.4169

Rao, M., Okreglak, V., Chio, U. S., Cho, H., Walter, P., & Shan, S. (2016). Multiple selection filters ensure accurate tail-anchored membrane protein targeting. *ELife*, *5*, e21301. https://doi.org/10.7554/eLife.21301

Rapoport, T. A. (2007). Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature*, *450*(November), 663–669. https://doi.org/10.1038/nature06384

Rapoport, T. A., Jungnickel, B., & Kutay, U. (1996). Protein Transport Across the Eukaryotic Endoplasmic Reticulum and Bacterial Inner Membranes. *Annual Review of Biochemistry*, *65*(1), 271–303. https://doi.org/10.1146/annurev.bi.65.070196.001415

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

Replogle, J. M., Norman, T. M., Xu, A., Hussmann, J. A., Chen, J., Cogan, J. Z., Meer, E. J., Terry, J. M., Riordan, D. P., Srinivas, N., Fiddes, I. T., Arthur, J. G., Alvarado, L. J., Pfeiffer, K. A., Mikkelsen, T. S., Weissman, J. S., & Adamson, B. (2020). Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nature Biotechnology*, *38*(8), Article 8. https://doi.org/10.1038/s41587-020-0470-y

Ridnõi, K., Šois, M., Vaidla, E., Pajusalu, S., Kelder, L., Reimand, T., & Õunap, K. (2019). A prenatally diagnosed case of Meckel–Gruber syndrome with novel compound heterozygous pathogenic variants in the TXNDC15 gene. *Molecular Genetics & Genomic Medicine*, *7*(5), e614. https://doi.org/10.1002/mgg3.614

Rivera-Monroy, J., Musiol, L., Unthan-Fechner, K., Farkas, Á., Clancy, A., Coy-Vergara, J., Weill, U., Gockel, S., Lin, S.-Y., Corey, D. P., Kohl, T., Ströbel, P., Schuldiner, M., Schwappach, B., & Vilardi, F. (2016). Mice lacking WRB reveal differential biogenesis requirements of tail-anchored proteins in vivo. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep39464

Roberson, E. C., Dowdle, W. E., Ozanturk, A., Garcia-Gonzalo, F. R., Li, C., Halbritter, J., Elkhartoufi, N., Porath, J. D., Cope, H., Ashley-Koch, A., Gregory, S., Thomas, S., Sayer, J. A., Saunier, S., Otto, E. A., Katsanis, N., Davis, E. E., Attié-Bitach, T., Hildebrandt, F., … Reiter, J. F. (2015). TMEM231, mutated in orofaciodigital and Meckel syndromes, organizes the ciliary transition zone. *The Journal of Cell Biology*, *209*(1), 129–142. https://doi.org/10.1083/jcb.201411087

Sachelaru, I., Petriman, N. A., Kudva, R., Kuhn, P., Welte, T., Knapp, B., Drepper, F., Warscheid, B., & Koch, H.-G. (2013). YidC Occupies the Lateral Gate of the SecYEG Translocon and Is Sequentially Displaced by a Nascent Membrane Protein *. *Journal of Biological Chemistry*, *288*(23), 16295–16307. https://doi.org/10.1074/jbc.M112.446583

Saraiva, J. M., & Baraitser, M. (1992). Joubert syndrome: A review. *American Journal of Medical Genetics*, *43*(4), 726–731. https://doi.org/10.1002/ajmg.1320430415

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*, e06306. https://doi.org/10.7554/eLife.06306

Schorr, S., Nguyen, D., Haßdenteufel, S., Nagaraj, N., Cavalié, A., Greiner, M., Weissgerber, P., Loi, M., Paton, A. W., Paton, J. C., Molinari, M., Förster, F., Dudek, J., Lang, S., Helms, V., & Zimmermann, R. (2020). Identification of signal peptide features for substrate specificity in human Sec62/Sec63-dependent ER protein import. *The FEBS Journal*, *287*(21), 4612–4640. https://doi.org/10.1111/febs.15274

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

Schulman, B. A., & Wade Harper, J. (2009). Ubiquitin-like protein activation by E1 enzymes: The apex for downstream signalling pathways. *Nature Reviews Molecular Cell Biology*, *10*(5), Article 5. https://doi.org/10.1038/nrm2673

Schultz, M. L., Krus, K. L., Kaushik, S., Dang, D., Chopra, R., Qi, L., Shakkottai, V. G., Cuervo, A. M., & Lieberman, A. P. (2018). Coordinate regulation of mutant NPC1 degradation by selective ER autophagy and MARCH6-dependent ERAD. *Nature Communications*, *9*(1), Article 1. https://doi.org/10.1038/s41467-018-06115-2

Schulz, J., Avci, D., Queisser, M. A., Gutschmidt, A., Dreher, L.-S., Fenech, E. J., Volkmar, N., Hayashi, Y., Hoppe, T., & Christianson, J. C. (2017). Conserved cytoplasmic domains promote Hrd1 ubiquitin ligase complex formation for ER-associated degradation (ERAD). *Journal of Cell Science*, *130*(19), 3322–3335. https://doi.org/10.1242/jcs.206847

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., & Selbach, M. (2013). Corrigendum: Global quantification of mammalian gene expression control. *Nature*, *495*(7439), 126–127. https://doi.org/10.1038/nature11848

Scott, N. A., Sharpe, L. J., & Brown, A. J. (2021). The E3 ubiquitin ligase MARCHF6 as a metabolic integrator in cholesterol synthesis and beyond. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, *1866*(1), 158837. https://doi.org/10.1016/j.bbalip.2020.158837

Serdiuk, T., Balasubramaniam, D., Sugihara, J., Mari, S. A., Kaback, H. R., & Müller, D. J. (2016). YidC assists the stepwise and stochastic folding of membrane proteins. *Nature Chemical Biology*, *September*. https://doi.org/10.1038/nchembio.2169

Shaheen, R., Ansari, S., Mardawi, E. A., Alshammari, M. J., & Alkuraya, F. S. (2013). Mutations in TMEM231 cause Meckel-Gruber syndrome. *Journal of Medical Genetics*, *50*(3), 160–162. https://doi.org/10.1136/jmedgenet-2012-101431

Shan, S., & Walter, P. (2005). Co-translational protein targeting by the signal recognition particle. *FEBS Letters*, *579*(4), 921–926. https://doi.org/10.1016/j.febslet.2004.11.049

Shao, S., & Hegde, R. S. (2011). Membrane Protein Insertion at the Endoplasmic Reticulum. *Annual Review of Cell and Developmental Biology*, *27*(1), 25–56. https://doi.org/10.1146/annurev-cellbio-092910-154125

Shao, S., Malsburg, K. Von Der, & Hegde, R. S. (2013). Listerin-Dependent Nascent Protein Ubiquitination Relies on Ribosome Subunit Dissociation. *Molecular Cell*, *50*(5), 637–648. https://doi.org/10.1016/j.molcel.2013.04.015

Shao, S., Rodrigo-Brenni, M. C., Kivlen, M. H., & Hegde, R. S. (2017). Mechanistic basis for a molecular triage reaction. *Science (New York, N.Y.)*, *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

Sharpe, L. J., Howe, V., Scott, N. A., Luu, W., Phan, L., Berk, J. M., Hochstrasser, M., & Brown, A. J. (2019). Cholesterol increases protein levels of the E3 ligase MARCH6 and thereby stimulates protein degradation. *The Journal of Biological Chemistry*, *294*(7), 2436–2448. https://doi.org/10.1074/jbc.RA118.005069

Shemorry, A., Hwang, C., & Varshavsky, A. (2013). Control of Protein Quality and Stoichiometries by N-Terminal Acetylation and the N-End Rule Pathway. *Molecular Cell*, *50*, 540–551. https://doi.org/10.1016/j.molcel.2013.03.018

Shiber, A., Döring, K., Friedrich, U., Klann, K., Merker, D., Zedan, M., Tippmann, F., Kramer, G., & Bukau, B. (2018). Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. *Nature*, *561*, 268–272. https://doi.org/10.1038/s41586-018-0462-y

Shing, J. C., Lindquist, L. D., Borgese, N., & Bram, R. J. (2017). CAML mediates survival of Myc-induced lymphoma cells independent of tail-anchored protein insertion. *Cell Death Discovery*, *3*(November 2016), 16098. https://doi.org/10.1038/cddiscovery.2016.98

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., & Weissman, J. S. (2018). The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. *ELife*, *7*, e37018. https://doi.org/10.7554/eLife.37018

Smalinskaitė, L., & Hegde, R. S. (2023). The Biogenesis of Multipass Membrane Proteins. *Cold Spring Harbor Perspectives in Biology*, *15*(4), a041251. https://doi.org/10.1101/cshperspect.a041251

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), Article 7934. https://doi.org/10.1038/s41586-022-05336-2

Smith, N., Wei, W., Zhao, M., Qin, X., Seravalli, J., Kim, H., & Lee, J. (2016). Cadmium and Secondary Structure-dependent Function of a Degron in the Pca1p Cadmium Exporter. *The Journal of Biological Chemistry*, *291*(23), 12420–12431. https://doi.org/10.1074/jbc.M116.724930

Stefanovic, S., & Hegde, R. S. (2007). 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-Barrett, S., Dickson, A. S., Burr, S. P., Williamson, J. C., Lobb, I. T., van den Boomen, D. J., Lehner, P. J., & Nathan, J. A. (2018). MARCH6 and TRC8 facilitate the quality control of cytosolic and tail-anchored proteins. *EMBO Reports*, *19*(5), e45603. https://doi.org/10.15252/embr.201745603

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* (p. 2023.03.09.531980). bioRxiv. https://doi.org/10.1101/2023.03.09.531980

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), Article 7934. https://doi.org/10.1038/s41586-022-05330-8

Sung, M., Porras-yakushi, T. R., Reitsma, J. M., Hess, S., Huber, F. M., Sweredoski, M. J., Hoelz, A., Hess, S., & Deshaies, R. J. (2016). A conserved quality-control pathway that mediates degradation of unassembled ribosomal proteins. *ELIFE*, *5*, 1–28. https://doi.org/10.7554/eLife.19105

Sung, M.-K., Reitsma, J. M., Sweredoski, M. J., Hess, S., & Deshaies, R. J. (2016). Ribosomal proteins produced in excess are degraded by the ubiquitin–proteasome system. *Molecular Biology of the Cell*, *27*(17), 2642–2652. https://doi.org/10.1091/mbc.e16-05-0290

Swanson, R., Locher, M., & Hochstrasser, M. (2001). A conserved ubiquitin ligase of the nuclear envelope/endoplasmic reticulum that functions in both ER-associated and Matalpha2 repressor degradation. *Genes & Development*, *15*(20), 2660–2674. https://doi.org/10.1101/gad.933301

Tan, J. M. E., van der Stoel, M. M., van den Berg, M., van Loon, N. M., Moeton, M., Scholl, E., van der Wel, N. N., Kovačević, I., Hordijk, P. L., Loregger, A., Huveneers, S., & Zelcer, N. (2020). The MARCH6-SQLE Axis Controls Endothelial Cholesterol Homeostasis and Angiogenic Sprouting. *Cell Reports*, *32*(5), 107944. https://doi.org/10.1016/j.celrep.2020.107944

Thornburg, C. C., Cowley, E. S., Sikorska, J., Shaala, L. A., Ishmael, J. E., Youssef, D. T. A., & McPhail, K. L. (2013). Apratoxin H and apratoxin A sulfoxide from the Red Sea cyanobacterium Moorea producens. *Journal of Natural Products*, *76*(9), 1781–1788. https://doi.org/10.1021/np4004992

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

Tiwari, S., & Weissman, A. M. (2001). Endoplasmic Reticulum (ER)-associated Degradation of T Cell Receptor Subunits: INVOLVEMENT OF ER-ASSOCIATED UBIQUITIN-CONJUGATING ENZYMES (E2s) *. *Journal of Biological Chemistry*, *276*(19), 16193–16200. https://doi.org/10.1074/jbc.M007640200

Tye, B. W., Commins, N., Ryazanova, L. V., Wühr, M., Springer, M., Pincus, D., & Churchman, L. S. (2019). Proteotoxicity from aberrant ribosome biogenesis compromises cell fitness. *ELife*, *8*, e43002. https://doi.org/10.7554/eLife.43002

The UniProt Consortium. (2017). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, *45*(D1), D158–D169. https://doi.org/10.1093/nar/gkw1099

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

Vembar, S. S., & Brodsky, J. L. (2008). One step at a time: Endoplasmic reticulum-associated degradation. *Nature Reviews. Molecular Cell Biology*, *9*(12), 944–957. https://doi.org/10.1038/nrm2546

Vilardi, F., Lorenz, H., & Dobberstein, B. (2011). WRB is the receptor for TRC40/Asna1-mediated insertion of tail-anchored proteins into the ER membrane. *Journal of Cell Science*, *124*(8), 1301–1307. https://doi.org/10.1242/jcs.084277

Vilardi, F., Stephan, M., Clancy, A., Janshoff, A., & Schwappach, B. (2014). WRB and CAML Are Necessary and Sufficient to Mediate Tail-Anchored Protein Targeting to the ER Membrane. *PLOS One*, *9*(1). https://doi.org/10.1371/journal.pone.0085033

Voigt, S., Jungnickel, B., Hartmann, E., & Rapoport, T. A. (1996). Signal sequence-dependent function of the TRAM protein during early phases of protein transport across the endoplasmic reticulum membrane. *Journal of Cell Biology*, *134*(1), 25–35. https://doi.org/10.1083/jcb.134.1.25

Volkmar, N., Thezenas, M., 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*. https://doi.org/10.1242/jcs.223453

von Heijne, G. (2007). The membrane protein universe: What's out there and why bother? *Journal of Internal Medicine*, *261*(6), 543–557. https://doi.org/10.1111/j.1365-2796.2007.01792.x

Voorhees, R. M., & Hegde, R. S. (2016). Structure of the Sec61 channel opened by a signal sequence. *Science*, *351*(6268), 88–91. https://doi.org/10.1126/science.aad4992

Wallin, E., & von Heijne, G. (1995). Properties of N-terminal tails in G-protein coupled receptors: A statistical study. *Protein Engineering*, *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*, 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

Wang, F., Chan, C., Weir, N. R., & Denic, V. (2014). The Get1/2 transmembrane complex is an endoplasmic-reticulum membrane protein insertase. *Nature*, *512*, 441–444. https://doi.org/10.1038/nature13471

Wang, X., Wei, X., Thijssen, B., Das, J., Lipkin, S. M., & Yu, H. (2012). Three-dimensional reconstruction of protein networks provides insight into human genetic disease. *Nature Biotechnology*, *30*(2), 159–164. https://doi.org/10.1038/nbt.2106

Wickles, S., Singharoy, A., Andreani, J., Seemayer, S., Bischoff, L., Berninghausen, O., Soeding, J., Schulten, K., van der Sluis, E. O., & Beckmann, R. (2014). A structural model of the active ribosome-bound membrane protein insertase YidC. *ELife*, *3*, e03035. https://doi.org/10.7554/eLife.03035

Wu, H., & Hegde, R. S. (2023). Mechanism of signal-anchor triage during early steps of membrane protein insertion. *Molecular Cell*, *83*(6), 961-973.e7. https://doi.org/10.1016/j.molcel.2023.01.018

Xu, Y., Anderson, D. E., & Ye, Y. (2016). The HECT domain ubiquitin ligase HUWE1 targets unassembled soluble proteins for degradation. *Cell Discovery*, *2*, 1–16. https://doi.org/10.1038/celldisc.2016.40

Yamamoto, Y., & Sakisaka, T. (2012). Molecular Machinery for Insertion of Tail-Anchored Membrane Proteins into the Endoplasmic Reticulum Membrane in Mammalian Cells. *Molecular Cell*, *48*(3), 387–397. https://doi.org/10.1016/j.molcel.2012.08.028

Yanagitani, K., Juszkiewicz, S., & Hegde, R. S. (2017). UBE2O is a quality control factor for orphans of multiprotein complexes. *Science*, *357*(August), 472–475.

Zattas, D., & Hochstrasser, M. (2015). Ubiquitin-dependent protein degradation at the yeast endoplasmic reticulum and nuclear envelope. *Critical Reviews in Biochemistry and Molecular Biology*, *50*(1), 1–17. https://doi.org/10.3109/10409238.2014.959889

Zavodszky, E., & Hegde, R. S. (2019). Misfolded GPI-anchored proteins are escorted through the secretory pathway by ER-derived factors. *ELife*, *8*, e46740. https://doi.org/10.7554/eLife.46740

Zhang, J.-P., Li, X.-L., Li, G.-H., Chen, W., Arakaki, C., Botimer, G. D., Baylink, D., Zhang, L., Wen, W., Fu, Y.-W., Xu, J., Chun, N., Yuan, W., Cheng, T., & Zhang, X.-B. (2017). Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated

double-stranded DNA cleavage. *Genome Biology*, *18*(1), 35.

https://doi.org/10.1186/s13059-017-1164-8