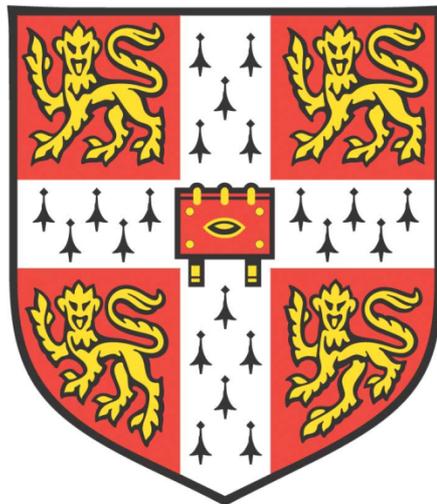


# Membrane protein biosynthesis at the endoplasmic reticulum

**Alina Guna**

*MRC Laboratory of Molecular Biology  
and  
St John's College*



Dissertation submitted for the degree of  
Doctor of Philosophy  
University of Cambridge  
March 2018

*All this happened, more or less*  
- Kurt Vonnegut

## **Declaration**

This dissertation is the result of my own work and includes nothing which is the work done in collaboration except where specifically indicated in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. It does not exceed the prescribed word limit for the Biology Degree Committee.

## Summary

The biosynthesis of integral membrane proteins (IMPs) is an essential cellular process. IMPs comprise roughly 20-30% of the protein coding genes of all organisms, nearly all of which are inserted and assembled at the endoplasmic reticulum (ER). The defining structural feature of IMPs is one or more transmembrane domains (TMDs). TMDs are typically stretches of predominately hydrophobic amino acids that span the lipid bilayer of biological membranes as an alpha helix. TMDs are remarkably diverse in terms of their topological and biophysical properties. In order to accommodate this diversity, the cell has evolved different sets of machinery that cater to particular subsets of proteins. Our knowledge of how the TMDs of IMPs are selectively recognized, chaperoned into the lipid bilayer, and assembled remains incomplete.

This thesis is broadly interested in investigating how TMDs are correctly inserted and assembled at the ER. To address this the biosynthesis of multi-pass IMPs was first considered. Multi-pass IMPs contain two to more than twenty TMDs, with TMDs that vary dramatically in terms of their biophysical properties such as hydrophobicity, length, and helical propensity. The beta-1 adrenergic receptor ( $\beta$ 1-AR), a member of the G-protein-coupled receptor (GPCR) family was established as a model substrate in an *in vitro* system where the insertion and folding of its TMDs could be interrogated. Assembly of  $\beta$ 1-AR is not a straightforward process, and current models of insertion fail to explain how the known translocation machinery correctly identifies, inserts, and assembles  $\beta$ 1-AR TMDs. An *in vivo* screen in mammalian cells was therefore conducted to identify additional factors which may be important for multi-pass IMP assembly. The ER membrane protein complex (EMC), a well conserved ER-resident complex of unknown biochemical function, was identified as a promising hit potentially involved in this assembly process.

The complexity of working with multi-pass IMPs in an *in vitro* system prompted the investigation of a simpler class of proteins. Tail-anchored proteins (TA) are characterized by a single C-terminal hydrophobic domain that anchors them into membranes. Though structurally simpler compared to multi-pass IMPs, the TMDs of TA proteins are similarly diverse. We found that known TA insertion pathways fail to engage low-to-moderately hydrophobic TMDs. Instead, these are chaperoned in the cytosol by calmodulin (CaM). Transient release from CaM allows substrates to sample the ER, where resident machinery mediates the insertion reaction. The EMC was shown to be necessary for the insertion of these substrates both *in vivo* and *in vitro*. Purified EMC in synthetic liposomes catalysed insertion of its TA substrates in a fully

reconstituted system to near-native levels. Therefore, the EMC was rigorously established as a TMD insertase. This key functional insight may explain its critical role in the assembly of multi-pass IMPs – which is now amenable to biochemical dissection.

## Acknowledgements

First and foremost, I have to thank Manu Hegde. He is an exceptional supervisor, scientist, and person. There are few things as exhilarating as watching mastery at work. As such, being Manu's student for the past three years has been an extreme pleasure. I am, and always will be, incredibly grateful for the chance he took taking me into his lab and the subsequent (long) hours he spent patiently teaching me.

Being in the Hegde lab has been the happiest time in my life. This is largely due to past and present lab members. I would specifically like to thank Susan Shao who first took me under her wing and taught me the basics of biochemistry. My lonely year of being the only graduate student in the lab was generously rewarded with the addition of Szymon Juskiewicz and Patrick Chitwood. I cherish working alongside you, and also your participation in out-of-lab shenanigans. I would like to thank Eszter Zavodszky for tolerating me and making the tissue culture room the best chatting nook I'll likely ever find. Finally, a special shout out to Rebecca Voorhees. You made talking about science probably a little too fun. But through our conversations I have learned an indispensable and elegant approach to thinking about problems.

The LMB is truly a spectacular place to work. I am particularly grateful to Sean Munro, my second supervisor, and Liz Miller, our next-door neighbour, for their fruitful insights. The support provided by the members of neighbouring labs and fellow students has also been indispensable. It is important for me to acknowledge the support staff in the workshops and kitchens that have made it a pleasure to do science at the LMB. Most importantly, I am delighted to have made two incredible friends, Suyang Zhang and Ivana Yeow. The late-night dinners, weekend lunches, and Kyoto garden strolls we shared were great.

About halfway through my PhD I went to the Physiology course at the Marine Biological Laboratory in Woods Hole. I think it is safe to say that I came back a very different person. I'm not entirely sure what happened out there, but something clicked. I have a lingering suspicion that Rob Phillips, Jodi Nunnari and Jennifer Lippincott-Schwartz had something to do with this transformation, and for that I am grateful.

Of course, I wouldn't be here at all if it weren't for the financial support I have received from the Gates Cambridge foundation, St John's College and the LMB.

Finally, I have to thank my friends and family for their support, encouragement, and the great fun they have brought into my life. I have been spoiled with having the most wonderful friends. I am particularly lucky to have met Ali Trelle, Cillian O'Fathaigh, Omar Wagih, Chris Rae, Matt Samson, and Margaret Comer who have been with me for this whole fantastic ride. To my oldest friends Jonathan Gray and Emma Cancelliere, I think it's safe to say that we're

never getting rid of each other. Being in England has been bittersweet. On one hand I get to see my grandmother and aunt, both of whom I adore, frequently. On the other hand, I have deeply missed my mom and dad and seeing my siblings Maria, Gabriela, and Stefan grow up. I have incredible gratitude for the freedom and support my parents have always given me despite my seemingly random life choices.

And last but not least – Ryanair. Your deals have been truly irresistible.

## Abbreviations

aa	amino acid
ADP	adenosine diphosphate
ATP	adenosine triphosphate
$\beta$ 1-AR	beta-1 adrenergic receptor
cryo-EM	cryo-electron microscopy
CaM	calmodulin
Cb5	cytochrome b5
cRM	canine pancreas derived rough microsomes
DBC	Deoxy Big CHAP
DSS	disuccinimidyl suberate
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EGTA	ethylene glycol tetraacetic acid
EMC	ER membrane complex
ER	endoplasmic reticulum
ERAD	endoplasmic reticulum associated protein degradation
FACS	fluorescence activated cell sorting
GET	guided entry of tail-anchored proteins
GFP	green fluorescent protein
GPCR	G-protein-coupled receptor
GTP	guanosine 5'-triphosphate
hRM	human cell derived rough microsomes
IMP	integral membrane protein
IP	immunoprecipitation
KD	knockdown
kDa	kilodalton
KO	knockout
min	minutes
mRNA	messenger RNA
PC	phosphatidyl-choline
PE	phosphatidyl-ethanolamine
PK	Proteinase K
PL	proteoliposomes

PMSF	phenylmethylsulfonyl fluoride
Prl	prolactin
RFP	red fluorescent protein
Rh-PE	1,2-dioleoyl-sn-glycero-3-phosphoethanolamine-N-lissamine rhodamine B
RM	rough microsomes
RNA	ribonucleic acid
RNC	ribosome nascent chain
RRL	rabbit reticulocyte lysate
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SND	SRP-independent targeting
SR	SRP receptor
SRP	signal recognition particle
SS	signal sequence
TA	tail-anchored
TMD	transmembrane domain
UV	ultraviolet light

# Table of contents

<b>Summary</b> .....	<b>iv</b>
<b>Acknowledgements</b> .....	<b>vi</b>
<b>Abbreviations</b> .....	<b>vii</b>
<b>Table of contents</b> .....	<b>x</b>
<b>List of figures</b> .....	<b>xiii</b>
<b>List of tables</b> .....	<b>xvi</b>
<b>1 Introduction</b> .....	<b>1</b>
1.1 Introduction.....	1
1.2 Co-translational insertion at the ER.....	2
1.2.1 The signal recognition particle.....	4
1.2.2 The Sec translocation channel.....	7
1.2.3 Accessory Sec translocon factors.....	10
1.3 Post-translational insertion at the ER.....	12
1.3.1 TRC40/Get3 and TA protein targeting.....	13
1.3.2 The Get1/Get2 insertase for TA proteins.....	14
1.3.3 Potential alternative post-translational pathways.....	15
1.4 Other TMD recognition factors.....	17
1.4.1 General cytosolic TMD binding and quality control factors.....	17
1.4.2 Quality control factors at the ER: the Hrd1 and Derlin families...21	
1.5 Overview of thesis. ....	23
<b>2 Mechanisms of multi-pass integral membrane protein insertion</b> .....	<b>24</b>
2.1 Multi-pass IMP TMD diversity and the beta-1 adrenergic receptor ( $\beta$ 1-AR) .....	25
2.2 Insertion and folding of $\beta$ 1-AR <i>in vitro</i> .....	29
2.3 $\beta$ 1-AR TMDs cannot insert in isolation.....	31
2.4 Contextual information is needed for $\beta$ 1-AR TMD insertion.....	32
2.5 Design of a mammalian <i>in vivo</i> screen to identify accessory factors.....	35
2.6 Mammalian <i>in vivo</i> screen results.....	40
2.7 Discussion.....	43
2.8 Materials and methods.....	46

<b>3</b>	<b>TRC-independent post-translational tail-anchored protein insertion</b>	<b>50</b>
3.1	Biodiversity of TA protein TMDs	50
3.2	Experimental assay for post-translational TA protein insertion	53
3.3	Less hydrophobic TA proteins use a TRC-independent insertion pathway	55
3.4	TRC40 cannot mediate the insertion of low hydrophobicity TA proteins before dissociation	59
3.5	Discussion	65
3.6	Materials and methods	68
<b>4</b>	<b>Cytosolic factors maintaining tail-anchored protein insertion competence</b>	<b>71</b>
4.1	SQS and VAMP2 have different cytosolic interactions	72
4.2	The primary cytosolic interaction factor for the SQS TMD is CaM	73
4.3	CaM is sufficient to maintain insertion competence of SQS	77
4.4	CaM dynamically interacts with substrates at physiologic calcium levels	80
4.5	SGTA can mediate SQS insertion in the absence of CaM	82
4.6	Discussion	85
4.7	Materials and methods	86
<b>5</b>	<b>The ER membrane complex (EMC) is essential for tail-anchored protein insertion</b>	<b>89</b>
5.1	SQS insertion is a protein mediated process	90
5.2	Known Sec systems are not involved in SQS insertion	92
5.3	The EMC is essential for SQS insertion <i>in vitro</i>	94
5.4	<i>In vivo</i> consequences of EMC deletion on TA insertion	98
5.5	Less hydrophobic TA proteins need EMC to insert	102
5.6	Discussion	103
5.7	Materials and methods	106
<b>6</b>	<b>Reconstitution of EMC-dependent tail-anchored protein insertion with purified factors</b>	<b>110</b>
6.1	EMC dependency is maintained after reconstitution	110
6.2	Purification and reconstitution of the mammalian EMC	114

6.3	EMC is sufficient for SQS insertion.....	117
6.4	Discussion.....	121
6.5	Materials and methods.....	122
<b>7</b>	<b>Perspectives and future work.....</b>	<b>127</b>
<b>8</b>	<b>References.....</b>	<b>132</b>
<b>9</b>	<b>Publications.....</b>	<b>151</b>
<b>10</b>	<b>Appendix 1: siRNA screen results.....</b>	<b>152</b>

## List of figures

Figure 1.1	Co-translational insertion of IMPs into the ER membrane.....	2
Figure 1.2	Architecture of the co-translational insertion pathway.....	3
Figure 1.3	Structures of the archaeal and mammalian Sec translocon.....	8
Figure 1.4	The TRC post-translational insertion pathway.....	12
Figure 1.5	Post-translational protein recognition in the cytosol.....	18
Figure 2.1	Structural diversity and insertion models of IMPs.....	26
Figure 2.2	Hydrophobic diversity of IMP TMDs.....	27
Figure 2.3	Structural characteristics of $\beta$ 1-AR.....	29
Figure 2.4	<i>In vitro</i> assay of $\beta$ 1-AR insertion and folding.....	30
Figure 2.5	Insertion of isolated $\beta$ 1-AR TMDs.....	31
Figure 2.6	Insertion of truncated $\beta$ 1-AR.....	33
Figure 2.7	$\beta$ 1-AR TMD2 can insert in the context of TMD1.....	34
Figure 2.8	Testing the orientation of inserted truncated $\beta$ 1-AR.....	35
Figure 2.9	Design and logic of <i>in vivo</i> screen for factors affecting $\beta$ 1-AR biosynthesis.....	37
Figure 2.10	Validating <i>in vivo</i> screen design by microscopy.....	38
Figure 2.11	Optimizing conditions for <i>in vivo</i> screen.....	39
Figure 2.12	$\beta$ 1-AR can engage ERAD machinery.....	40
Figure 2.13	Results of <i>in vivo</i> screen.....	41
Figure 3.1	Diversity of TA TMDs.....	52
Figure 3.2	Glycosylation based assay for TA protein insertion.....	54
Figure 3.3	Endogenous TA protein sequences.....	55
Figure 3.4	Insertion and TRC40 interaction of endogenous TA proteins.....	56
Figure 3.5	SQS and VAMP2 insertion in the presence of TRC pathway inhibitors.....	57
Figure 3.6	TA insertion sensitivity to TRC pathway inhibitors.....	58
Figure 3.7	TRC-dependence is a function of TMD hydrophobicity.....	59
Figure 3.8	Schematic of site-specific crosslinking.....	61
Figure 3.9	Preparation of TA-SGTA complexes in the PURE system.....	62
Figure 3.10	Ability of TA substrates to transfer onto TRC40.....	63
Figure 3.11	Preparation of TA-TRC40 complexes in the PURE system.....	64
Figure 3.12	Stability of TA-TRC40 interaction.....	65

Figure 3.13	TRC40 cannot mediate SQS insertion before dissociation.....	66
Figure 4.1	Analysis of interactions made by SQS and VAMP2 in native RRL.....	72
Figure 4.2	CaM is the main interacting partner of SQS.....	74
Figure 4.3	Quantifying all SQS TMD interaction partners in native RRL.....	75
Figure 4.4	Calcium chelation causes SQS to associate with non-CaM chaperones.....	77
Figure 4.5	SQS is not soluble or insertion competent in the PURE system in the absence of chaperones.....	78
Figure 4.6	SQS insertion in membranes is concomitant with its release from CaM.....	80
Figure 4.7	Sensitivity of the SQS-CaM interaction to calcium levels.....	81
Figure 4.8	Analysis of SQS insertion when CaM is inactivated.....	82
Figure 4.9	SQS interactions in nucleated RRL.....	83
Figure 4.10	SQS-SGTA insertion into membranes.....	84
Figure 5.1	SQS insertion depends on intact membranes.....	91
Figure 5.2	Sec61 is not involved in SQS insertion.....	92
Figure 5.3	The Sec61/Sec63 complex is not involved in SQS insertion.....	93
Figure 5.4	SQS insertion is impaired in the absence of EMC5.....	95
Figure 5.5	The EMC subunits.....	96
Figure 5.6	Characterizations of EMC KO and rescue cell lines.....	97
Figure 5.7	SQS cannot insert in the absence of an intact EMC <i>in vitro</i> .....	98
Figure 5.8	Visualizing TA protein fate <i>in vivo</i> .....	99
Figure 5.9	SQS cannot insert in the absence of an intact EMC <i>in vivo</i> .....	100
Figure 5.10	SQS mislocalizes and aggregates in the absence of EMC.....	101
Figure 5.11	Endogenous SQS levels are reduced in the absence of EMC.....	102
Figure 5.12	Analysis of TA proteins for EMC-dependent insertion.....	103
Figure 6.1	Characterization of EMC in reconstituted proteoliposomes with total protein.....	111
Figure 6.2	SQS insertion into $\Delta$ EMC6-PL.....	113
Figure 6.3	Localization of EMC5-GFP.....	114
Figure 6.4	Purification of EMC from mammalian cells.....	115
Figure 6.5	SQS insertion is intact in EMC5-FLAG cells.....	116
Figure 6.6	EMC can be purified in DBC.....	116
Figure 6.7	Characterization of EMC in purified proteoliposomes.....	117

Figure 6.8	EMC is sufficient for SQS, but not VAMP2 insertion into membranes.....	118
Figure 6.9	EMC is sufficient to insert the isolated SQS-CaM complex to near-native levels.....	119
Figure 6.10	SQS insertion depends on an intact EMC.....	119
Figure 6.11	SQS insertion into empty liposomes is susceptible to freeze/thaw cycles.....	120
Figure 7.1	The role of EMC in post- and co-translational insertion.....	128

## List of tables

Table 1	siRNA screen results of highest GFP:RFP ratios.....	41
Table 2	siRNA screen results of lowest GFP:RFP ratios.....	42

# Chapter 1: Introduction

## 1.1 Introduction

Roughly one-fourth of eukaryotic genes code for integral membrane proteins (IMPs) (Krogh et al., 2001). This widely diverse class of proteins are essential for cellular function with critical roles in signal transduction, organelle biogenesis, small molecule transport, intracellular trafficking and cell adhesion. Since IMPs differ in their topologic and biophysical features, the cell has evolved multiple strategies to mediate targeting and insertion into membranes. The majority of IMPs are initially assembled at the endoplasmic reticulum (ER) (Shao and Hegde, 2011a). After reaching maturity in the ER, these proteins remain, are distributed amongst numerous membrane bound organelles in the cell or are trafficked to the plasma membrane (Mellman and Warren, 2000). The defining structural feature of an IMP is the presence of one or more transmembrane domains (TMDs). TMDs are typically stretches of predominately hydrophobic amino acids that span the lipid bilayer as an alpha helix (White and Von Heijne, 2005). They are incredibly diverse, varying significantly in terms of their hydrophobicity, length, helical propensity, and context within a protein. The decisive step in the biogenesis of all IMPs is the integration of the TMDs into the membrane. This key event poses two major biochemical challenges that must be overcome by targeting and insertion machinery.

Foremost, protein synthesis machinery resides in the cytosol, an aqueous environment. This means that IMPs are initially made in a hydrophilic medium where they are intrinsically insoluble. Therefore, before an IMP can be stably inserted into the lipid bilayer, it must first transit through the inhospitable cytosol. Specialized factors are needed to recognize TMDs in the cytosol and shield them until they can engage the appropriate insertion machinery at the destination membrane (Borgese and Fasana, 2011; Keenan et al., 2001). Without such factors, TMDs are susceptible to inappropriate interactions and potentially toxic aggregation.

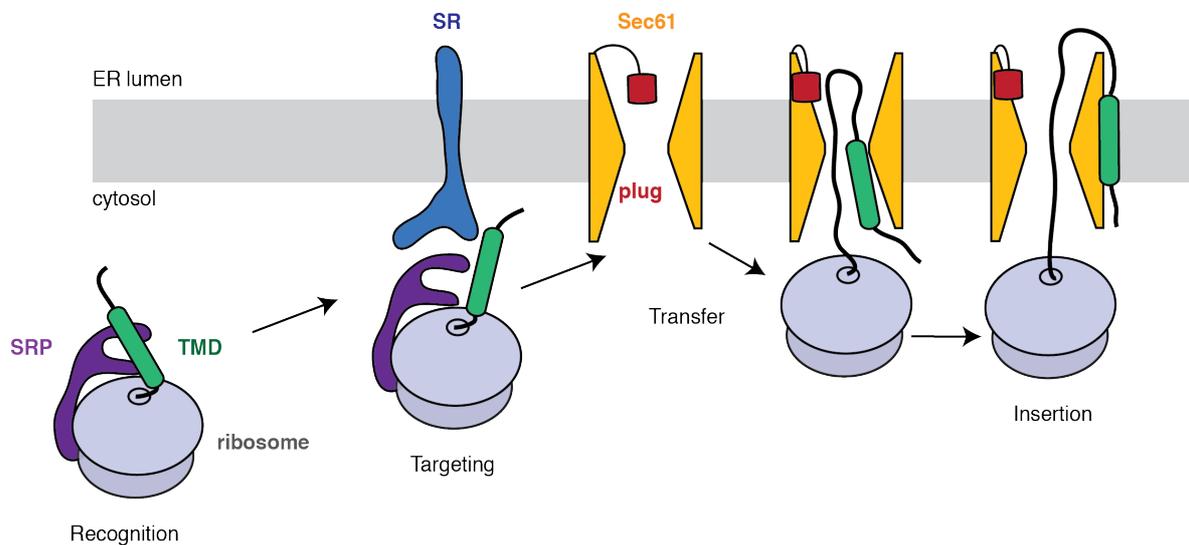
Second, TMDs vary in sequence and biophysical properties, making their recognition in the cytosol and at the membrane a challenge (Papaloukas et al., 2008). The diversity of TMDs makes it difficult for even the best prediction algorithms to accurately discern between bona fide TMDs and hydrophobic segments of soluble proteins (Zhao and London, 2006). However, the machinery in the cytosol and membrane manage to rapidly make this distinction with high fidelity for a large volume of synthesized proteins. Ultimately, TMD recognition is the critical and decisive step made by the membrane embedded insertion machinery, which

partitions TMDs from an aqueous to lipid environment (Cymer et al., 2015; Shao and Hegde, 2011a).

These two central challenges, effective chaperoning and accurate recognition, highlight the essential functions played by TMD recognition factors throughout the biosynthetic process. Two conceptually different strategies have evolved to facilitate the synthesis of IMPs: co-translational and post-translational insertion. Most ER targeted proteins utilize the co-translational pathway, with the known exceptions of tail-anchored and small secretory proteins, which rely on post-translational routes. Though these pathways have been extensively characterized, many questions remain. Namely, it is unclear how the currently known machinery is capable of handling unfavourable TMDs, particularly in the context of assembling multi-pass IMPs.

### 1.2 Co-translation insertion at the ER

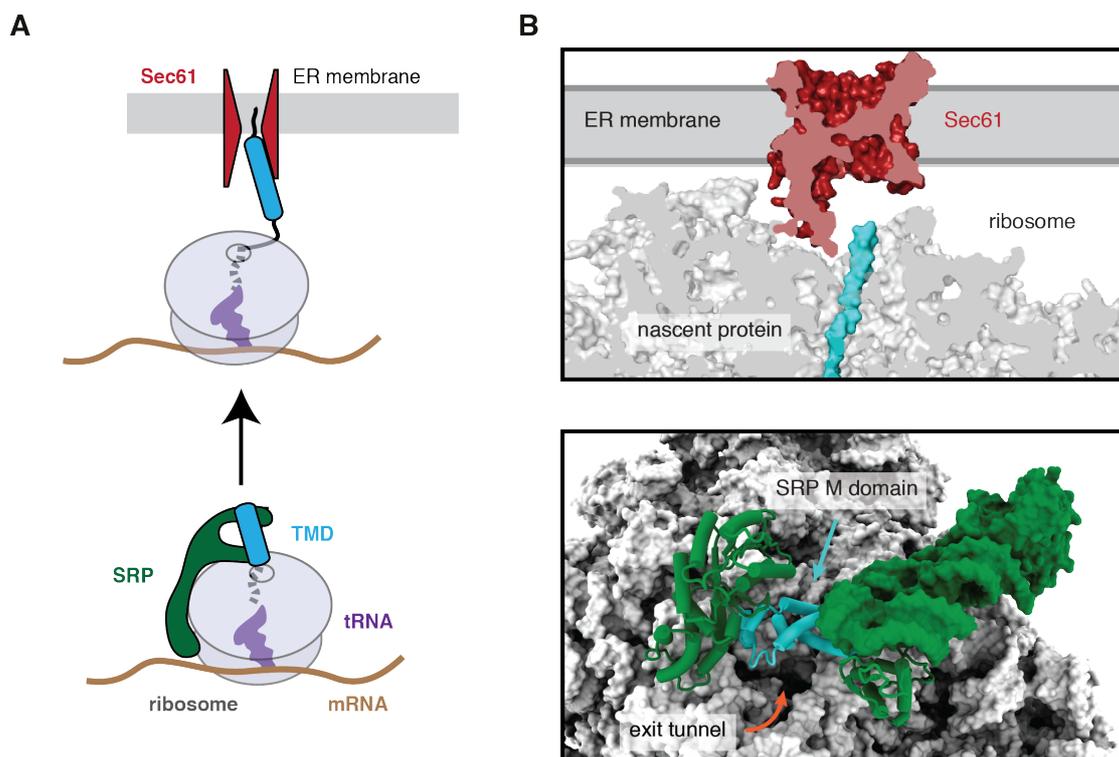
The vast majority of IMPs inserted at the ER utilize the co-translational pathway (Figure 1.1). A set of evolutionarily conserved machinery both in the cytosol and at the ER membrane mediate the recognition, targeting, and insertion of IMPs concurrent with their synthesis



**Figure 1.1 Co-translational insertion of IMPs into the ER membrane.** A hydrophobic element of an IMP (either a signal sequence or TMD, green) is first recognized by the signal recognition particle (SRP, purple) on the ribosome (grey). SRP targets the ribosome-nascent chain complex (RNC) to the ER membrane via the SRP receptor (SR, blue). The GTPase activities of SRP and SR culminate in the transfer of the RNC to the Sec61 translocon (yellow). TMD interaction with the translocon leads to the displacement of the luminal plug (red) and the opening of the lateral gate towards the lipid bilayer. Insertion into the ER membrane depends on accurate recognition of substrate TMDs.

## Chapter 1: Introduction

(Jungnickel and Rapoport, 1995; Rapoport, 2007). The co-translational strategy circumvents the problem of TMD exposure between initial synthesis and membrane insertion by temporal coupling of TMD recognition to protein synthesis and physical coupling to the ribosome (Figure 1.2). This strategy is defined by the signal recognition particle (SRP), which is precisely positioned at the ribosomal exit tunnel, ready to engage emerging hydrophobic elements. Association with SRP immediately after synthesis effectively eliminates exposure of the TMD to bulk cytosol. Following SRP engagement, the ribosome nascent chain (RNC) is delivered to the ER membrane via the SRP receptor (SR). The RNC is then transferred to the Sec translocon. The ribosome engages the Sec translocon such that the ribosome's exit tunnel is aligned with the translocon's central pore (Figure 1.2B) (Beckmann, 1997). This arrangement permits new TMDs emerging from the ribosome to directly interact with the Sec translocon. The molecular details of these step-wise interactions illuminate the limitations and current unknowns of the co-translational pathway.



**Figure 1.2 Architecture of the co-translational insertion pathway.** (A) Diagram of co-translational recognition of a TMD (cyan) by SRP (green) at the ribosome (grey) and subsequent hand-off to the Sec61 translocon. (B) SRP (green and cyan) is intimately associated with the ribosome (grey) and positioned precisely at the ribosomal exit tunnel. The M domain of the SRP54 subunit (cyan) recognizes TMDs (PDB code 3JAN). Following targeting to the ER membrane, the ribosomal exit tunnel aligns with the central pore of the Sec61 translocon (red), permitting minimal exposure of the nascent chain (cyan) to the bulk cytosol (PDB code 3J7Q). Structures rendered with help from Aaron Lewis and Manu Hegde.

### 1.2.1 The signal recognition particle

TMDs emerging from a translating ribosome are first recognized in the cytosol by the signal recognition particle (SRP) (Keenan et al., 2001). In eukaryotes, SRP can also recognize cleavable N-terminal sequences which are comprised of a ~7-15 residue hydrophobic domain. The critical functional features of SRP are its ability to bind the ribosome at the mouth of its exit tunnel, its capacity to recognize hydrophobic sequences, and its GTP-dependent interaction with the ER-localized SRP receptor (SR) (Figure 1.2). These cumulatively allow SRP to mediate targeting and hand-off of the RNC to the Sec61 translocation channel (Akopian et al., 2013; Halic and Beckmann, 2005; Keenan et al., 2001). The core framework for SRP function, developed in the early 1980s (Walter et al., 1984), remains virtually unchanged to the present. However, new technologies have now provided proteome-wide views of SRP's substrate range *in vivo* and structural insights into the molecular mechanisms of hydrophobic element recognition.

Early biochemical analyses of SRP established substrate hydrophobicity as a key parameter mediating recognition and targeting (Walter and Blobel, 1981a; Walter et al., 1981). This was confirmed by *in vivo* experiments showing that proteins whose biogenesis was most affected by SRP depletion were those containing the most hydrophobic signal sequences (Ast et al., 2013; Ng et al., 1996; Ulbrandt et al., 1997). Unaffected proteins were assumed to either not use SRP, or effectively use an alternative mechanism in its absence, a distinction that could not be easily resolved based solely on those studies. Beyond verifying the importance of hydrophobicity and distinguishing membrane proteins as a particularly sensitive class of clients, these studies left ambiguous the full complement of proteins that utilize SRP under unperturbed conditions.

Recent studies have tried to establish the substrate range of SRP in wild type cells by identifying the mRNA positions of all ribosomes affinity purified via SRP (Chartron et al., 2016; Schibich et al., 2016). In *E. coli*, SRP engages essentially all membrane proteins that co-translationally expose a TMD, and a subset of secretory proteins with very hydrophobic signal sequences (Schibich et al., 2016). This matches well with earlier work showing that in *E. coli*, SRP is primarily needed for membrane proteins (Ulbrandt et al., 1997). The substrate range is broader in *S. cerevisiae*, with SRP engaging the majority of ER destined TMD and signal containing proteins (Chartron et al., 2016). The only ER destined proteins not consistently recovered with SRP are those which use post-translational means of insertion.

## Chapter 1: Introduction

Remarkably, a substrate's genetic requirement for SRP (Ast et al., 2013) is not necessarily correlated with whether it engages SRP under normal conditions. Thus, although strong genetic dependence on SRP is a compelling (but not definitive) indicator of its physical engagement, SRP-independence is not equivalent to lack of SRP engagement. Instead, SRP-independence as gauged by genetic studies likely reflects a protein's capacity to avoid aggregation or degradation before using another route to the membrane (Ast et al., 2013; Halic et al., 2004; Panzner et al., 1995). Consistent with these conclusions, acute SRP depletion in yeast showed de-localization from the ER of nearly all SRP-dependent mRNAs and around half of SRP-independent mRNA (Costa et al., 2018). Thus, with the exception of post-translationally delivered proteins, SRP physically engages and is functionally necessary for essentially all ER destined proteins. By contrast, a subset of ER destined soluble proteins can effectively be imported into the yeast ER in the absence of SRP despite using it under normal conditions.

The molecular basis for SRP engagement with hydrophobic elements has come from structural analysis (Figure 1.2B). In eukaryotes, SRP is a ribonucleoprotein complex composed of a ~300 nucleotide RNA scaffold and six protein subunits. The key function component is SRP54 (Ffh in *E. coli*). SRP54 contains a methionine-rich M domain which is responsible for engaging signal sequences and TMDs. When SRP is bound to the ribosome, the M domain is precisely positioned at the ribosomal exit tunnel (Halic et al., 2004; Schaffitzel et al., 2006), ready to sample nascent chains as they emerge. Early crystal structures of the Ffh M domain revealed a hydrophobic groove formed by alpha-helices that were enriched with methionines (Keenan et al., 1998). Subsequent structures showed that hydrophobic peptides bind within the hydrophobic groove of the isolated M domain (Hainzl and Sauer-Eriksson, 2015; Janda et al., 2010; Voorhees and Hegde, 2015). In the current set of available structures, there are slightly different arrangements of the M domain alpha helices, suggesting a scaffold that is sufficiently flexible to conform to different substrates.

More recently, advances in cryo-EM have resulted in the determination of high-resolution structures of native RNC-SRP complexes before and after substrate binding. In addition to allowing earlier X-ray structures of isolated domains to be docked into the native complex, these structures have also revealed the mechanism for how SRP might impose a hydrophobicity threshold for substrate binding, explaining previous biochemical and *in vivo* observations (Voorhees and Hegde, 2015). In an unengaged state, the hydrophobic groove of the M domain was observed to be auto-inhibited by an amphipathic C-terminal 'placeholder' helix. In the engaged structure, this C-terminal domain is displaced by the substrate TMD

## Chapter 1: Introduction

which subsequently resides in the hydrophobic groove of the M domain. The C-terminal domain shifts to a position where it serves as a 'lid' over the substrate bound hydrophobic groove. It is of note that the C-terminal domain was omitted in all earlier X-ray structures, which is why it was not previously observed.

These observations suggest that the hydrophobicity threshold for SRP binding is effectively set by the biophysical properties of the C-terminal helix: its displacement can only occur when its hydrophobicity is surpassed by a putative substrate. The auto-inhibitory helix also prevents the constitutive exposure of SRPs hydrophobic groove and explains why free SRP cannot effectively bind hydrophobic sequences promiscuously. Thus, SRP can only engage substrates when they are presented in a constrained environment near the ribosomal exit tunnel at high local concentrations.

Interestingly, both genome-wide studies and biochemical analysis suggest that SRP can be specifically recruited to the ribosome before the TMD emerges from the exit tunnel (Berndt et al., 2009; Voorhees and Hegde, 2015), or in some cases before the hydrophobic domain is even synthesized (Chartron et al., 2016). This pre-recruitment offers obvious benefits: it allows SRP to capture hydrophobic sequences before they are exposed to the cytosol and without competition from other, more abundant cytosolic TMD binding factors. It remains unclear how SRP can be recruited to ribosomes that are translating membrane bound proteins without relying on direct nascent chain interaction. As ribosome conformation during translation can vary in a codon-specific manner (Lareau et al., 2014), it has been speculated that subtle differences in translation might be exploited to recognize ribosomes decoding stretches of hydrophobic residues (Voorhees and Hegde, 2015). Alternatively, some distinguishing feature of the pioneer round of translation may facilitate SRP pre-recruitment (Chartron et al., 2016). In general, the relationship between SRP and translation elongation remains unclear given that the original finding of SRP halting elongation in heterologous plant-mammal hybrid *in vitro* systems (Walter and Blobel, 1981b) was not recapitulated in subsequent experiments in homologous systems (Chartron et al., 2016; Ingolia et al., 2009, 2011; Wolin and Walter, 1989).

Following SRP engagement of a hydrophobic element in the cytosol, the RNC is directed to the ER membrane via a GTP-dependent interaction between SRP and the SRP receptor (SR) (Gilmore et al., 1982a, 1982b; Meyer et al., 1982; Shan and Walter, 2005; Walter and Johnson, 1994) (Figure 1.1). Both SRP and SR have GTPase activity (Grudnik et al., 2009; Keenan et al., 2001; Shan and Walter, 2005). The eukaryotic SR is an ER-resident heterodimer composed of an  $\alpha$  and  $\beta$  subunit, both of which contain a GTPase domain (Connolly and

## Chapter 1: Introduction

Gilmore, 1989; Miller et al., 1995). It is not totally clear how the GTPase activities of SRP and SR are coordinated, but it is thought that GTP binding occurs during targeting, causing a conformational change that displaces the hydrophobic element from the binding pocket of SRP, while simultaneously increasing the affinity between SRP and SR (Bacher et al., 1996; Rapiejko and Gilmore, 1997). Subsequent GTP hydrolysis results in the dissociation of the SRP-SR complex at the ER membrane permitting SRP recycling into the cytosol (Connolly et al., 1991). The end result of this interaction is the docking of the RNC onto the Sec61 protein-conducting channel.

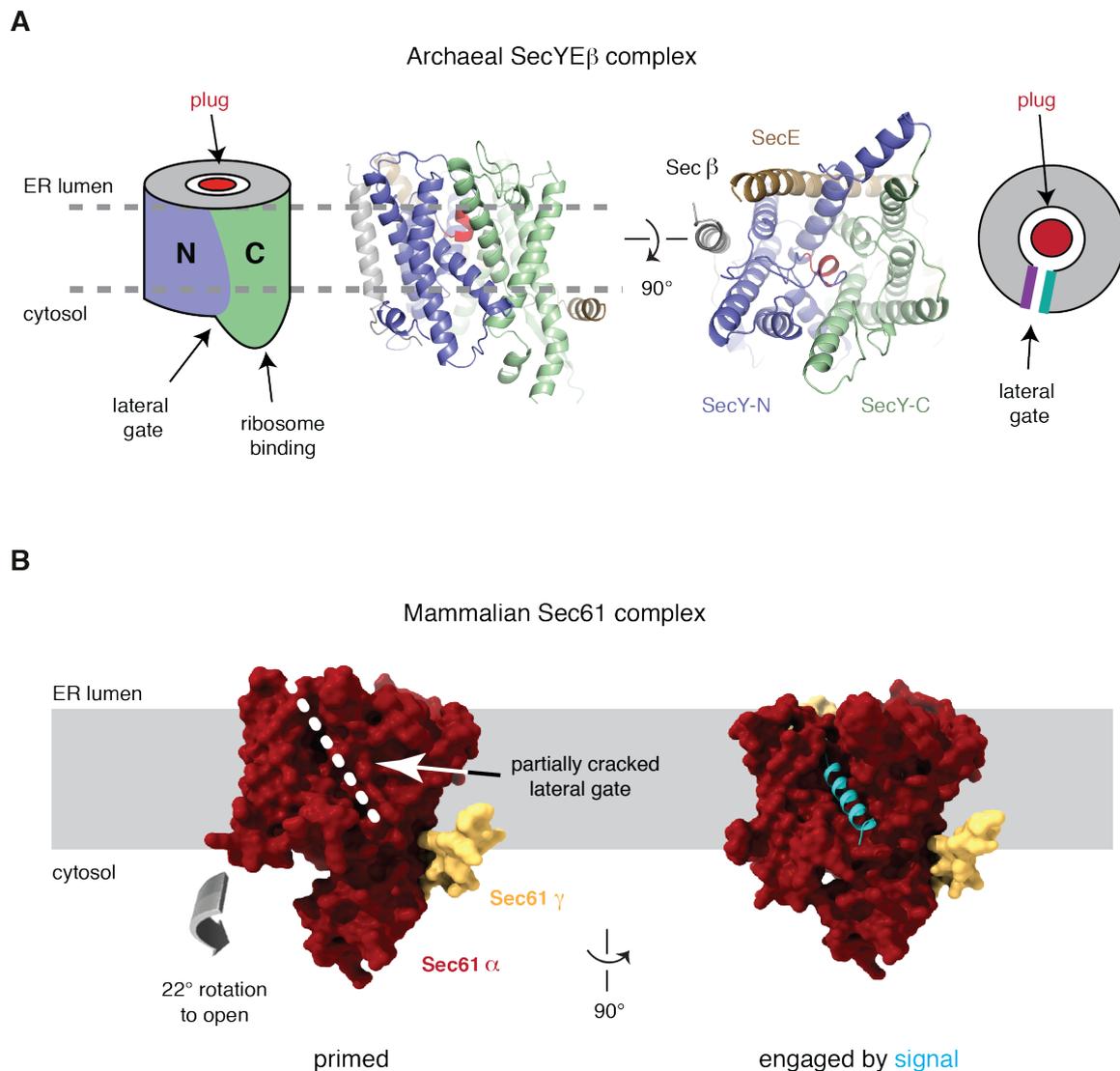
### 1.2.2 The Sec translocation channel

The best understood and longest known route for membrane protein insertion is the universally conserved Sec translocon, which also mediates soluble protein translocation (Rapoport, 2007). While this channel can operate in both post- and co-translational modes, all IMPs delivered via SRP use the latter. The binding of the ribosome to the Sec translocon aligns the ribosomal exit tunnel with the translocon's central pore (Beckmann et al., 1997) (Figure 1.2B). This arrangement allows TMDs emerging from the ribosome to avoid exposure to the cytosol and instead directly engage the translocon. Co-translational recognition and insertion of individual TMDs in succession obviates the challenge of maintaining the solubility and insertion competence of highly hydrophobic, complicated multi-pass IMPs in the cytosol. As such, essentially all multi-pass IMPs are thought to obligately use the Sec translocon.

The central component of the greater Sec translocon is the three-protein Sec61 complex in eukaryotes, and the homologous SecYEG complex in prokaryotes and archaea (Figure 1.3). Extensive biochemical analysis had long established that Sec61 not only provides a pore across the membrane (Mothes et al., 1994), but also recognizes TMDs and grants them lateral access to the lipid bilayer (Martoglio et al., 1995; Mothes et al., 1998; Plath et al., 1998). These insights became indispensable for interpreting ensuing structures of the Sec translocon. The seminal X-ray structure of the isolated archaeal SecY complex showed that SecY (Sec61) forms a pseudo-symmetric clamshell surrounding an hourglass-shaped central pore (Van den Berg et al., 2004) (Figure 1.3A). The slimmest part of the pore is distinguished by a ring of six

## Chapter 1: Introduction

conserved hydrophobic residues. On top of this ring lies a short helix which acts as a plug, occluding the channel when it is in an unengaged state. The back of the clamshell is reinforced by SecE (Sec61 $\gamma$  in eukaryotes). Sec $\beta$  is more peripheral, and while it helps facilitate



**Figure 1.3 Structures of the archaeal and mammalian Sec translocon.** (A) Key structural elements of the Sec translocon based on the crystal structure of the archaeal SecYE $\beta$  complex (PDB code 1RH5). Cartoon diagrams illustrate the important features, with N and C representing the N-terminal and C-terminal halves of the pseudo-symmetric SecY protein. SecE and Sec $\beta$  are omitted from the diagram for simplicity. (B) Structures of the ribosome-bound mammalian Sec61 complex (PDB codes 3J7Q and 3JC2) before (left) and after (right) binding of a hydrophobic signal peptide (cyan). The alpha and gamma subunits of the Sec61 complex are in dark red and yellow respectively. The beta subunit is not seen in the presented view. The position of the lateral gate, which is partially cracked by ribosome binding, is indicated by the dashed white line. Binding of a signal sequence causes a  $\sim 22^\circ$  rotation of the N-terminal half of Sec61 $\alpha$  to open the luminal half of the lateral gate, where the signal peptide subsequently binds. Note that without a bound signal, the channel is closed both towards the lipid bilayer and across the membrane. Structures rendered with help from Manu Hegde.

## Chapter 1: Introduction

translocation, it is not essential (Kalies et al., 1998). Inferring function from the X-ray structure, it was postulated that the plug helix would be displaced when the channel opens, while the seam at the front of the clamshell could act as a lateral gate between the central pore of the channel and the lipid bilayer.

Major advances in single particle cryo-EM have allowed native ribosome-Sec61 complexes to be visualized at close to atomic resolution (Figure 1.3B) (Voorhees and Hegde, 2016a, 2016b; Voorhees et al., 2014). Previously, these were only available at lower resolution (Gogala et al., 2014; Ménétret et al., 2005). In parallel, sub-nanometer views of the translocon in native membrane have been made possible by advances in cryo-tomography (Pfeffer et al., 2015). Structural comparisons of Sec61 $\alpha$  (or SecY) before and after engagement have provided mechanistic insight into substrate recognition. Compared to the quiescent SecY crystal structure (Van den Berg et al., 2004), the ribosome-bound, but substrate-free, Sec61 complex shows several small conformational changes that are thought to ‘prime’ the channel for succeeding substrate recognition (Voorhees et al., 2014). Upon engagement, the ribosome tightly binds two cytosolic loops in the C-terminal half of Sec61 $\alpha$ . This results in subtle shifts in the associated helices. This rigid constraint is propagated to the remaining helices and culminates in a cracked lateral gate, now primed and awaiting the arrival of a substrate.

The structure of the signal sequence-engaged ribosome-Sec61 complex revealed additional conformational changes that accompany substrate binding. As predicted based on earlier photo-crosslinking experiments (Martoglio et al., 1995; Mothes et al., 1998; Plath et al., 1998), the substrate forms an alpha helix that is situated at the Sec61 $\alpha$  lateral gate, exposed to the surrounding lipid. The accommodation of substrates at this site is facilitated by a 22° rigid-body rotation of the N-terminal half of Sec61 $\alpha$ , relative to the hinge at the back of the ‘clamshell’ (Voorhees and Hegde, 2016a). This movement parts the luminal portion of the lateral gate, offering an entry point for substrates into the lipid bilayer. The C-terminal half of Sec61 $\alpha$  remains in place relative to the ribosome, which stays tightly bound. The widening of the lateral gate expands the central pore, simultaneously destabilizing the helical plug. Therefore, substrate recognition is coupled to channel opening, resulting in a conformation where the aqueous pore and lipid bilayer are connected via the lateral gate where the signal peptide is bound. The same overall architecture was seen in a crystal structure of SecY bound to a hydrophobic peptide fused to the SecY partner SecA (Li et al., 2016), and in cryo-tomography structures of the Sec61 complex bound to endogenous substrates in native ER membranes (Pfeffer et al., 2015, 2016). Cumulatively, structural evidence suggests that Sec61

## Chapter 1: Introduction

substrate recognition is a highly conserved process and occurs similarly across species and in the native environment of the ER membrane.

With obvious analogy to the mode of substrate recognition by SRP, it appears that the engaged signal sequence occupies the position at the lateral gate previously occupied by Sec61 $\alpha$ 's helix 2. The interactions between helix 2 and 7 are instead recapitulated by the signal sequence and helix 7. It is attractive to postulate that the hydrophobicity threshold for substrate discrimination is effectively set by Sec61 $\alpha$ 's internal helix 2, with only substrates capable of displacing it gaining access to the lateral gate. Consistent with this view, strengthening or weakening the interactions at the lateral gate via mutations (Trueman et al., 2011) or small molecules (MacKinnon et al., 2014) corresponds to a shift in the hydrophobicity threshold for substrate recognition.

The current model of Sec61 substrate engagement fails to explain how the diverse range of TMDs found in endogenous IMPs are efficiently inserted into the lipid bilayer. Future structural analysis of Sec61 engaged with biophysically diverse signal sequences and TMDs is essential for a complete understanding of this process. This is particularly important for the insertion of multi-pass IMPs. The prevailing model posits that TMDs in the context of multi-pass IMPs are individually and successively recognized by the translocon and subsequently integrated into the lipid bilayer as they emerge from the ribosome (Blobel, 1980; Wessels and Spiess, 1988). However, these TMDs display a broad range of hydrophobicity, with many not recognized by Sec61 for insertion into the lipid bilayer when tested in isolation (Enquist et al., 2009). It is likely that the interactions between such 'weak' TMDs and other TMDs within the same protein (Cymer and von Heijne, 2013; Heinrich and Rapoport, 2003; Skach and Lingappa, 1993), or even in trans (Feige et al., 2015), are important for insertion. How such cooperation might occur, and the role of the Sec61 in integrating this contextual information remains an important direction for future study.

### 1.2.3 Accessory Sec translocon factors

The Sec61 translocon is sufficient to mediate the insertion of most IMPs that take the co-translational route to the ER. However, numerous accessory factors have been postulated to have roles, presumably aiding biosynthesis during or after Sec61 interacts with substrates. Of these, the most well studied is the translocating-chain association membrane (TRAM) protein. TRAM has been implicated in various aspects of translocation, but its biochemical role remains

## Chapter 1: Introduction

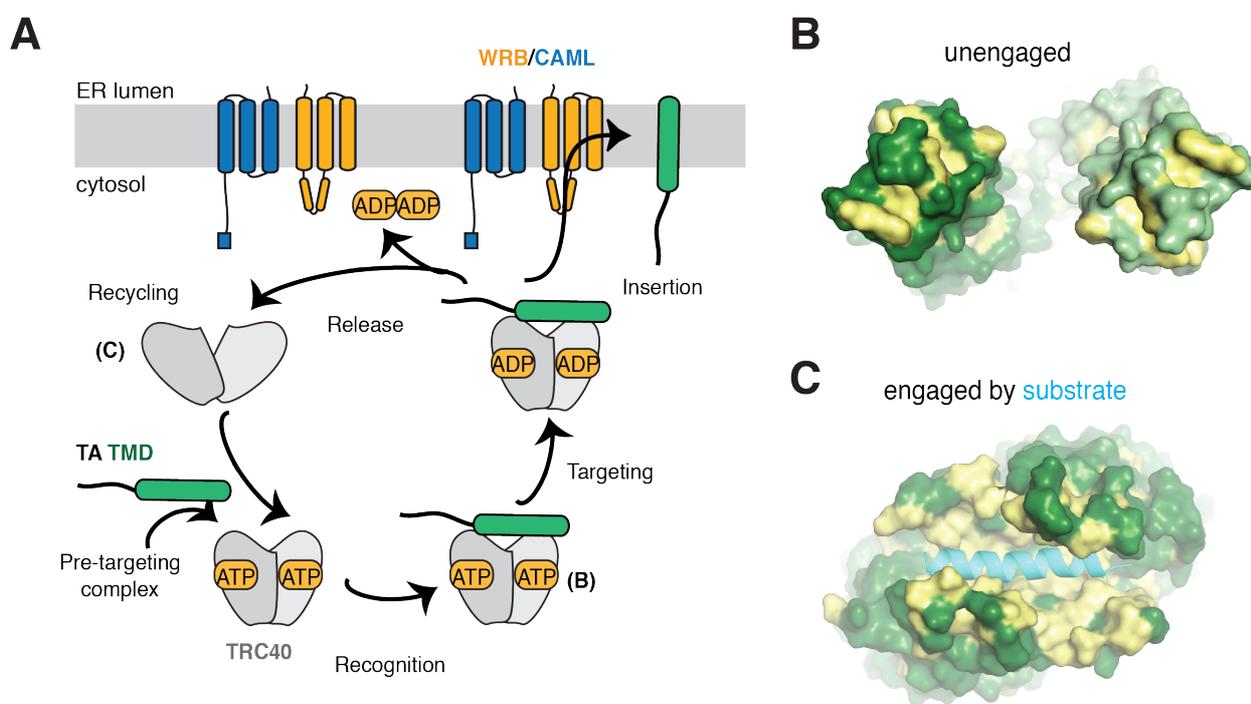
largely obscure. In early experiments with fully purified systems, TRAM was shown to have a stimulatory effect on the translocation of several substrates (Görlich and Rapoport, 1993; Görlich et al., 1992a). However, TRAM dependence is limited as many proteins can be translocated or integrated in its absence (Görlich and Rapoport, 1993; Görlich et al., 1992a; Oliver et al., 1995). Crosslinking analysis indicates that TRAM interacts with both signal sequences and TMDs (Do et al., 1996; Görlich et al., 1992a; Heinrich et al., 2000; High et al., 1993), and may play a specific role in stimulating the insertion of proteins with weakly hydrophobic TMDs (Heinrich et al., 2000; Voigt et al., 1996). In eukaryotic cells, TRAM is not essential.

Another abundant Sec61 associated factor is the translocon-associated protein (TRAP) complex. TRAP is seen prominently in cryo-electron tomography structures of the Sec translocon (Pfeffer et al., 2015, 2016), and weakly in cryo-EM structures, likely due to disorder as a result of partially dissociated TRAP subunits (Ménétrete et al., 2008, 2005; Pfeffer et al., 2014; Voorhees et al., 2014). Biochemical analyses indicate that TRAP is physically associated with Sec61 (Conti et al., 2015; Dejgaard et al., 2010; Shibatani et al., 2005; Snapp et al., 2004), and can chemically crosslink to nascent proteins undergoing transport into the ER lumen (Görlich et al., 1992b; Wiedmann et al., 1987, 1989). TRAP was observed to have a stimulatory effect on the translocation of certain proteins depending on signal sequence efficiency (Fons et al., 2003). However, other studies instead suggest that TRAP may have a role in helping TMDs establish topology within the plane of the membrane (Sommer et al., 2013). Similar to TRAM, TRAP has been associated with various functions, but its exact biochemical role awaits future investigation.

Other putative accessory factors implicated in some aspect of translocation are Sec62, Sec63 and the ribosome-associated membrane protein 4 (RAMP4). The Sec62-Sec63 complex is known to associate with the Sec61 translocon (Meyer et al., 2000) but is thought to be important primarily for the post-translational translocation of small secretory proteins (Lakkaraju et al., 2012; Lang et al., 2012). The Sec62-Sec63 complex has additionally been reported to influence the insertion of C-terminal transmembrane domains of IMPs (Jung et al., 2014). RAMP4 has been shown to be involved in stabilizing IMPs during a concerted ER stress response (Yamaguchi et al., 1999). Multiple reports suggest that perturbing various Sec61 associated factors impacts certain aspects of translocation. However, whether the majority of these observations represent indirect effects, or if these factors have specific roles catering to distinct subsets of substrates is unclear.

### 1.3 Post-translational insertion at the ER

Certain IMPs do not have access to the SRP-dependent pathway (Kutay et al., 1995). This includes small secretory proteins as well as tail-anchored (TA) proteins. TA proteins are defined by a single TMD close to the carboxyl terminus. The proximity of the TMD to the termination codon means translation ends before the TMD exits the ribosomal tunnel (which can accommodate roughly ~40 amino acids of a nascent peptide). Termination prevents TA proteins from being recognized by SRP and instead makes them reliant on a post-translational pathway for delivery to the ER (Figure 1.4A). This pathway was discovered 10 years ago (Favaloro et al., 2008; Schuldiner et al., 2008; Stefanovic and Hegde, 2007) and is now understood in significant detail in both yeast and higher eukaryotes (Chio et al., 2017; Hegde



**Figure 1.4 The TRC post-translational insertion pathway.** (A) After TA substrates are transferred to TRC40 via the pre-targeting complex (see Figure 1.5), the TA-TRC40 complex is targeted to the ER membrane by interactions with the WRB/CAML receptor complex. After ATP hydrolysis, the TA substrate is released for insertion into the lipid bilayer, a process also mediated by the WRB/CAML complex. ATP binding allows for the recycling of TRC40 back to the cytosol. (B, C) Structure of the yeast Get3 homodimer in the nucleotide-free apo state (B, PDB code 2WOO) and the ADP•AlF<sub>4</sub><sup>-</sup> state engaged with the TMD of the yeast protein Pep12 (C, PDB code 4XTR). The helical subdomains of the two subunits are separated in the apo structure. The large hydrophobic surfaces (yellow) come together in the closed conformation, forming a large composite hydrophobic that can accommodate substrate TMDs (cyan). Structures rendered with help from Manu Hegde.

## Chapter 1: Introduction

and Keenan, 2011). The central TMD recognition factor in the cytosol is the widely conserved TRC40 (Get3 in yeast). Apart from preventing aggregation, TRC40 serves the critical function of targeting substrates to the ER membrane via a series of protein-protein interactions with its dedicated receptor, the heterodimer WRB/CAML (Get1/Get2 in yeast). Following release from TRC40, the WRB/CAML heterodimer is sufficient to mediate insertion of TMDs into the lipid bilayer.

### 1.3.1 TRC40/Get3 and TA protein targeting

TRC40/Get3 is the primary factor responsible for delivering TA proteins to the ER membrane. Crystal structures show that Get3 is a symmetric homodimer, with each monomer comprising an ATPase domain and an  $\alpha$ -helical domain (Bozkurt et al., 2009; Hu et al., 2009; Mateja et al., 2009; Suloway et al., 2009; Yamagata et al., 2010). The arrangement of the  $\alpha$ -helical domain is sensitive to the nucleotide state of the ATPase domain as well as association with other factors in the targeting pathway (Figure 1.4A). In the ADP-bound (or apo) state, the two  $\alpha$ -helical domains are separated and folded such that there is no exposure of their hydrophobic surfaces (Figure 1.4B). ATP binding triggers the movement of the  $\alpha$ -helical domains together and they re-configure to form a large, methionine-rich, hydrophobic groove that can accommodate substrate TMDs (Figure 1.4C) (Mateja et al., 2015). It is postulated that in this conformation, each  $\alpha$ -helical domain lends an amphipathic helix that can serve as a flexible lid over the substrate, therefore shielding it from the cytosol.

Despite the high presence of ATP in the cytosol, TRC40 is not independently effective at capturing substrates. It is possible that even in the ATP-bound state the amphipathic lid helices are arranged such that the hydrophobic groove is occluded, preventing promiscuous interactions. Only when TRC40 is in complex with substrate-loading factors can it efficiently bind TMDs that are passed on from the chaperone SGTA (Sgt2 in yeast) (Gristick et al., 2014; Mateja et al., 2015; Mock et al., 2015; Wang et al., 2010). TRC40 and SRP's M domain seem to have analogous modes of operating, with TRC40's  $\alpha$ -helical domains forming a methionine-rich hydrophobic groove that is likely auto-inhibited by an intramolecular amphipathic helix while in a latent state. This strategy may set a hydrophobicity threshold (Rao et al., 2016; Wang et al., 2010) while simultaneously permitting binding only when substrates are at a high local concentration: at the ribosome exit tunnel for SRP and at the substrate-loading complex for

## Chapter 1: Introduction

TRC40. After substrate binding, the same auto-inhibitory helix changes conformation such that it may act as a lid, protecting the TMD in its binding pocket.

### 1.3.2 The Get1/Get2 insertase for TA proteins

The requirement for an alternative insertase to the Sec translocon at the ER membrane was first suggested by the need to insert TA proteins targeted by TRC40 (Stefanovic and Hegde, 2007). Indeed, the receptor for TRC40 also turned out to be the insertase and is composed of the heterodimeric Get1/Get2 complex (WRB/CAML in eukaryotes). Get1/Get2 are both triple-pass multi-spanning IMPs. They were originally identified as a putative receptor for the targeting factor Get3 based on genetic and physical interaction studies in budding yeast (Schuldiner et al., 2008). It was later confirmed in reconstituted systems with recombinant factors that this complex constitutes the minimal insertion machinery (Mariappan et al., 2011). Single-molecule analysis suggests that a single Get1/Get2 heterodimer is sufficient for TA insertion into the lipid bilayer (Zalisko et al., 2017).

Transmembrane domain transfer from Get3 to the Get1/Get2 complex is a highly coordinated event. A molecular model has been deduced from structural analysis (Mariappan et al., 2011; Stefer et al., 2011), and accompanying biochemical studies (Figure 1.4A) (Mariappan et al., 2011; Wang et al., 2014). The Get3-TMD complex is initially recruited to the membrane through an interaction with a short two-helix motif at the end of Get2's flexible N-terminal cytosolic tail. Once tethered at the membrane, the cytosolic coiled-coil of Get1 wedges between the two alpha-helical domains of the Get3 dimer. This splits the two subunits apart, disrupting the hydrophobic groove and thereby inducing substrate release. Get1's coiled-coil protrudes into the ATP binding site of Get3, so ATP hydrolysis or dissociation must occur at some stage between Get2 and Get1 association. This is supported by the observation that a Get3 ATP hydrolysis mutant cannot effectively release substrates (Stefanovic and Hegde, 2007). Subsequent re-binding of ATP to the substrate-free Get3-Get1 complex promotes recycling of Get3 back to the cytosol (Figure 1.4A).

The exact mechanism for how TMDs are inserted into the lipid bilayer by the Get1/Get2 complex is not known. However, this process appears to involve the TMDs of Get1/Get2 (Wang et al., 2014). Mutations at various sites of TMDs of Get1/Get2 disrupts TA insertion while leaving intact the Get3 recruiting function. These transmembrane domains also physically crosslink to TA insertion intermediates, implying that Get1/Get2 recognizes TMDs

## Chapter 1: Introduction

and provides a path from the cytosol to the membrane. Structures of the Get1/Get2 complex, together with structure-guided mutagenesis, will be needed to understand the mechanism of its insertase activity.

### 1.3.3 Potential alternative post-translational pathways

It has been known since the discovery of TRC40 that not every TA protein can effectively engage this factor (Stefanovic and Hegde, 2007). This conclusion was reinforced in subsequent studies showing only partial defects in yeast lacking GET pathway components (Jonikas et al., 2009; Schuldiner et al., 2008), and incomplete defects of TA insertion in TRC40 deficient lysates, cells, and tissues (Casson et al., 2017; Rivera-Monroy et al., 2016; Shao et al., 2017). The capability of certain TA proteins to insert into the membrane both *in vitro* and *in vivo* with some efficiency in the absence of GET factors has led to the proposal of numerous ‘alternative’ pathways.

A protein-free or ‘unassisted’ insertion pathway has been suggested as a viable route for certain ER-targeted TA proteins. Studies demonstrating unassisted insertion have primarily used mammalian cytochrome b5 (Cb5), a TA protein of low hydrophobicity, as a model substrate (Borgese and Fasana, 2011). It was observed that Cb5 can translocate its C-terminus across a pure lipid bilayer with the same rapidity and efficiency as in microsomal membranes, obviating the need for additional membrane insertion machinery (Brambillasca et al., 2005). At least *in vitro*, cytosolic chaperones are not necessary for facilitating the insertion process and are perhaps needed simply to keep substrates in an insertion competent state by preventing aggregation (Colombo et al., 2009). In this model, insertion specificity into the correct membrane is imparted by differences in the composition of the lipid bilayer. Specifically, the ER bilayer has relatively low cholesterol content compared to downstream compartments of the exo- and endocytic system (van Meer, 1989). Indeed, when the cholesterol level in synthetic lipid vesicles is increased to levels higher than those found in the ER membrane, Cb5 fails to insert (Brambillasca et al., 2005). The main limitations of these studies are their complete reliance on *in vitro* analysis. Though potentially illuminating, these purified systems only provide one membrane for insertion, making it difficult to assay targeting specificity.

Previous observations of cytosolic components interacting with certain TA proteins have led to the obvious suggestion that these factors are needed for targeting. Both SRP and the chaperones Hsc70/Hsp40 have been reported to physically interact with TA substrates

## Chapter 1: Introduction

based on crosslinking analysis in crude rabbit reticulocyte lysate followed by antibody immunoprecipitation (Abell et al., 2004, 2007). Despite the proposal that SRP is involved in the delivery of TA proteins to the ER membrane, there is no functional data to support a role for Sec61-mediated translocation (Kutay et al., 1995; Steel et al., 2002; Yabal et al., 2003). In plants, Hsc70 was also reported to interact with the TMD of a TA protein and facilitate its *in vivo* targeting to the nuclear envelope (Brkljacic et al., 2009). In light of recent SRP affinity ribosome profiling studies (Chartron et al., 2016) and the identification of the TRC40 pathway (Favaloro et al., 2008; Schuldiner et al., 2008; Stefanovic and Hegde, 2007) it is likely that neither of these two types of reported TA engagement are physiologically relevant. Physical association based on *in vitro* crosslinking analysis probably reflects numerous minor ‘off-pathway’ interactions given lengthy reaction times and the hydrophobic nature of the TA substrate.

More recently, another putative ER targeting pathway has been described in yeast comprised of the SND (SRP-independent targeting) proteins (SND1, SND2, and SND3). These proteins are thought to be important for the targeting of IMPs that have centrally located TMDs (Aviram et al., 2016). This subset of ER destined proteins is described as unable to be recognized by either SRP (which prefers N-terminal TMDs) or TRC40/Get3 (which recognizes C-terminal TMDs). The SND pathway has also been suggested to compensate in the absence of GET pathway components. The human homolog of SND2, hSND2, has been identified and implicated in the same processes as in yeast (Haßdenteufel et al., 2017). It is speculated that the cytosolic SND1 may somehow mediate substrate recognition, interacting with the ER-localized SND2 and SND3, which may act as a receptor. However, the SND pathway has not been shown to directly mediate targeting or insertion of any TA substrates in yeast or mammals (Aviram et al., 2016; Casson et al., 2017; Haßdenteufel et al., 2017). Indeed, an attempt to provide direct evidence by analyzing targeting of TA proteins by pulse labeling and glycosylation in SND2 depleted strains failed to show an effect on insertion (Aviram et al., 2016). In mammalian cells, deletion of hSND2 had a similarly negligible effect on TA insertion (Haßdenteufel et al., 2017). Additionally, there remains a lack of biochemical evidence that SND proteins can physically interact with proposed substrates. Thus, the exact role of these factors in TMD recognition or post-translational targeting remains to be determined.

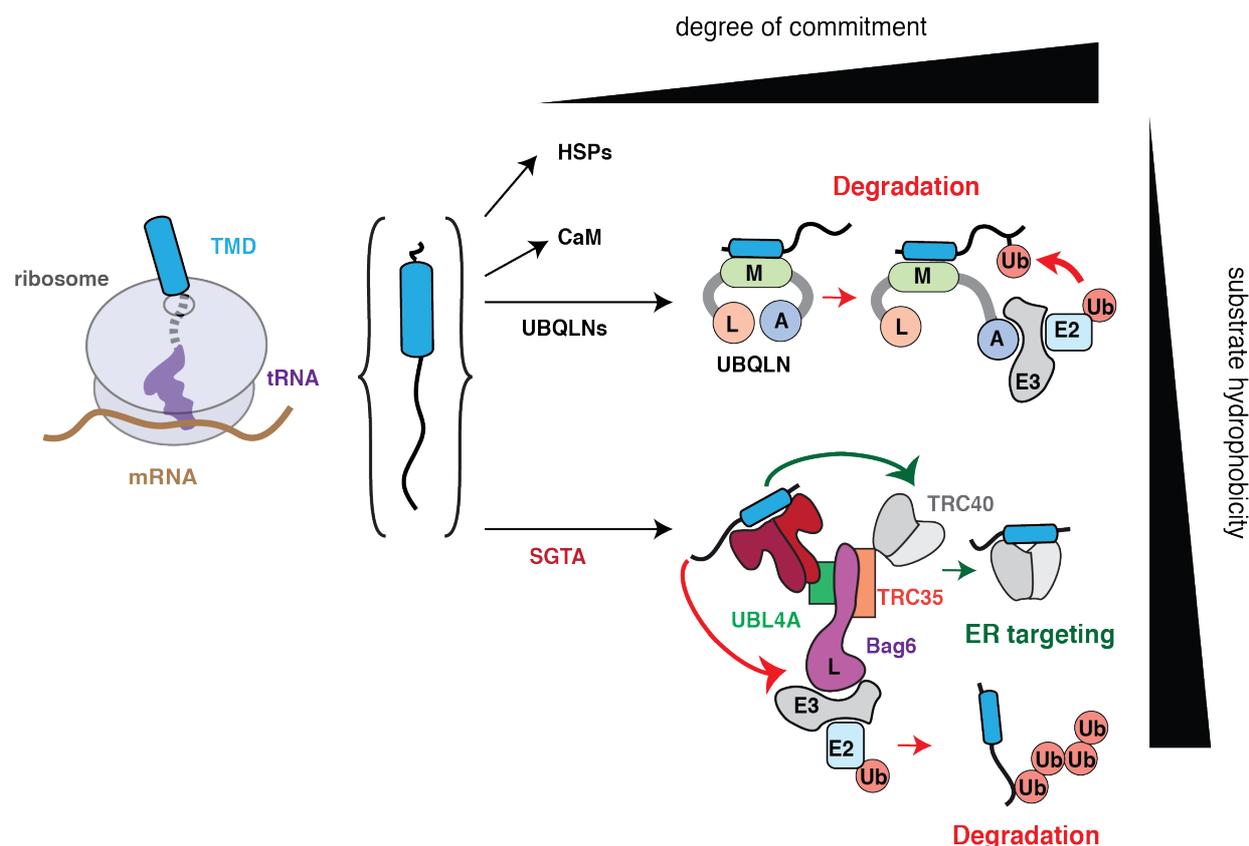
### 1.4 Other TMD recognition factors

In addition to components with central roles in designated targeting pathways, a number of other factors are capable of engaging TMDs (Figure 1.5). In the cytosol, these include general binding factors that generically protect TMDs from inappropriate interactions and those that are linked to protein degradation machinery. Following insertion at the membrane, ER quality control machinery ensures that only correctly folded proteins are trafficked to their final destinations (Ellgaard and Helenius, 2003), with misfolded proteins routed for degradation via ER associated protein degradation pathways (ERAD) (Vembar and Brodsky, 2008). Subsets of ERAD machinery are therefore capable of recognizing and interacting with aberrant TMDs within the plane of the lipid bilayer.

#### 1.4.1 General cytosolic TMD binding and quality control factors

SGTA and Calmodulin (CaM) are general TMD binding factors that do not specifically mediate either membrane targeting or degradation (Figure 1.5). Clients can instead dynamically interact with these factors until they either engage membrane machinery or load onto a dedicated targeting factor.

SGTA has been considered the most upstream factor in the TA targeting pathway to the ER membrane (Shao et al., 2017; Wang et al., 2010). However, it does not have specificity solely for ER destined proteins (Itakura et al., 2016; Wang et al., 2010). Additionally, depletion experiments of SGTA *in vitro* and its deletion in yeast show very mild and selective phenotypes (Jonikas et al., 2009; Shao et al., 2017). Instead, SGTA is best viewed as a broad-specificity TMD recognition factor. The SGTA substrate complex can transiently sample the substrate-loading complex for TRC40 (Mateja et al., 2015; Mock et al., 2015; Rao et al., 2016; Shao et al., 2017; Wang et al., 2010). If the SGTA-bound TMD is sufficiently hydrophobic, it can transfer directly and rapidly to TRC40 for ER targeting. If not, substrates can dissociate from SGTA and attempt to engage machinery for mitochondrial insertion (Itakura et al., 2016), or if this fails, degradation machinery (Shao et al., 2017). Thus, SGTA is a general TMD binding factor that precludes substrate aggregation and queries suitability for TRC40 loading but does not commit substrates to a particular fate. The mechanism of substrate binding by SGTA remains unclear, but it is noteworthy that the binding domain is methionine-rich and predicted to form an alpha-helical bundle, similar to the TMD-binding domains of TRC40 and SRP.



**Figure 1.5 Post-translational protein recognition in the cytosol.** TMD binding proteins arranged approximately based on their relative client hydrophobicity preferences (y-axis) and degree of substrate commitment to a particular fate after binding (x-axis). UBQLNs bind dynamically (via the “M” domain) to allow multiple attempts at membrane insertion, generally for many mitochondrial TMDs. However, the UBA domain (“A” in the figure) of UBQLNs can recruit an E3 ligase that ubiquitinates substrates, thereby committing them for degradation. This results in UBA interaction with substrate ubiquitin, exposing the UBL domain (“L” in the figure), which mediates targeting to the proteasome for degradation. SGTA is a general TMD binding factor that is typically recruited to the Bag6 complex via an SGTA-UBL4A interaction, where a bound substrate has the opportunity to transfer to TRC40. TRC40 is itself recruited to the complex by an interaction with TRC35. If substrate transfer does not occur, substrates dissociate from SGTA and can be captured by Bag6. Substrate binding to TRC40 leads to ER targeting, while binding to Bag6 is commitment to degradation. The UBL domain of Bag6 (“L” in Bag6) recruits the E3 ligase RNF126 for substrate ubiquitination before targeting to the proteasome.

Calmodulin (CaM), which also binds its substrates via a flexible methionine-rich alpha-helical scaffold, has exceptionally broad specificity for a range of peptide sequences (O’Neil and DeGrado, 1990). At physiologic  $\text{Ca}^{2+}$  levels, CaM dynamically engages both signal sequences and TMDs (Martoglio et al., 1997; Shao and Hegde, 2011b). Crystal structures of CaM with a substrate (Finn et al., 1995; Meador et al., 1992; O’Neil and DeGrado, 1990) suggest that the bi-lobed CaM can wrap around a hydrophobic domain to fully protect it from the aqueous environment. CaM can shield hydrophobic segments of up to ~18 residues, which

## Chapter 1: Introduction

is appreciably longer than segments favoured by protein folding factors such as Hsp70 (Zhu et al., 1996). This presumably explains why the latter is used for soluble protein folding, where exposed hydrophobicity patches are typically ~3-5 residues, while the former engages signal sequences and TMDs. Unlike SGTA, CaM does not appear to directly interact with dedicated targeting factors. Therefore, given the high abundance of CaM in the cytosol, it is instead likely a buffer against hydrophobic sequence aggregation. CaM essentially acts as a non-committed factor that is permissive, but not directly involved in, protein targeting.

In addition to these general factors, there are also two known TMD binding factors that can associate with ubiquitination machinery, which subsequently marks substrates for proteasome mediated degradation. The first discovered was Bag6, which contains an N-terminal ubiquitin-like (UBL) domain that recruits the E3 ubiquitin ligase RNF126 (Hessa et al., 2011; Rodrigo-Brenni et al., 2014; Shao et al., 2017). In the context of the complete cytosol, Bag6 appears to have specificity for longer hydrophobic sequences that characterize TMDs. Bag6 interaction with TMDs has a very slow off rate, making it effectively irreversible over physiologic time frames (Shao et al., 2017). The structural basis of Bag6 interaction is not known, but unlike all of the other cytosolic TMD binding factors, Bag6 does not contain a methionine-rich region.

Remarkably, the C-terminal region of Bag6 is an essential structural component of the TRC40 substrate-loading complex (Mock et al., 2015; Shao et al., 2017). This part of Bag6 links the proteins UBL4A, which recruits SGTA to the complex, and TRC35, which recruits TRC40. This ‘mini’ core complex, lacking the TMD binding and UBL domains of Bag6, is fully functional for TA loading onto TRC40. In yeast, the homologs of UBL4A and TRC35 (known as Get5 and Get4, respectively) interact directly in the Get3 loading complex (Gristick et al., 2014; Mateja et al., 2015; Wang et al., 2010). As such, there is no Bag6 homolog in yeast. It appears that Bag6 is therefore an embellishment embedded within the TA pathway in higher eukaryotes to route failed targeting substrates for degradation.

In addition to membrane proteins mislocalized to the cytosol, Bag6 also recognizes membrane proteins dislocated into the cytosol from the ER membrane as a consequence of ERAD (Claessen and Ploegh, 2011; Claessen et al., 2014; Wang et al., 2011). To carry out this function, a subset of the Bag6 pool is recruited to the site of substrate dislocation (Xu et al., 2013), where it presumably captures hydrophobic domains as they emerge into the cytosol. In its absence, dislocated ERAD substrates aggregate in the cytosol. Thus, Bag6 seems to patrol the cytosol for membrane proteins, regardless of their source, and ensures their degradation before aggregation can occur. SGTA may also help prevent aggregation of ERAD substrates

## Chapter 1: Introduction

(Xu et al., 2012), presumably shielding exposed hydrophobic elements alongside Bag6. The broader specificity of Bag6 for non-TA membrane proteins (Hessa et al., 2011; Rodrigo-Brenni et al., 2014) and other particularly hydrophobic segments (Rodrigo-Brenni et al., 2014; Yamamoto et al., 2017) makes it best characterized as a general quality control factor for most mislocalized membrane proteins.

More recently, the Ubiquilins were discovered as a family of TMD binding factors that can also couple with protein ubiquitination machinery (Figure 1.5) (Itakura et al., 2016). Four Ubiquilins (UBQLN1 to UBQLN4) are found in mammals, with the sole yeast homolog known as Dsk2. Unlike Bag6, biochemical analysis indicates that Ubiquilins engage substrates dynamically (Itakura et al., 2016). In this mode of engagement, Ubiquilins allow substrates opportunities to engage insertion machinery at a target membrane (which seems to be mitochondria for many clients). Over time, however, Ubiquilins recruit a yet unidentified E3 ligase to mediate substrate ubiquitination. This is the commitment step for degradation because substrate ubiquitin binds to a ubiquitin associating (UBA) domain in Ubiquilin, preventing substrate dissociation while simultaneously favouring proteasome targeting via Ubiquilin's ubiquitin-like (UBL) domain (Itakura et al., 2016).

The substrate specificity of Ubiquilins is roughly similar to Bag6, but with a preference for TMDs of consistently lower hydrophobicity (Itakura et al., 2016). This may explain why Ubiquilins preferentially bind mitochondrial membrane proteins, whose TMDs are typically less hydrophobic than those destined for the ER. Supporting this notion of specificity, cultured cells lacking Ubiquilins show impaired degradation of mitochondrial membrane precursors that fail import (Itakura et al., 2016). Stimulated B-cells experiencing mitochondrial stress are acutely dependent on UBQLN1, the absence of which leads to mitochondrial precursor accumulation and cell death (Whiteley et al., 2017).

Different Ubiquilin family members might have somewhat different substrate binding properties. For example, in contrast to the other family members, UBLN4 seems to engage ER destined proteins that fail targeting (Suzuki and Kawahara, 2016). The substrate binding region of Ubiquilins is highly methionine-rich and predicted to be mostly alpha-helical, similar to SGTA (Wang et al., 2010), TRC40 (Mateja et al., 2009), CaM (Meador et al., 1992) and SRP (Keenan et al., 1998). *In vitro* and *in vivo*, both the Ubiquilins and Bag6 appear to monitor the cytosol for exposed TMDs, preventing their aggregation and favouring degradation. This function might be particularly important when protein import is impaired during organelle stress (Kang et al., 2006; Whiteley et al., 2017; Wright et al., 2001).

## Chapter 1: Introduction

The relationships between cytosolic TMD binding factors is important to consider in light of the eventual fates of IMPs. The cytosol seems to have several abundant factors that in isolation have remarkably broad substrate ranges for almost any hydrophobic region of a polypeptide. All of them are capable of recognizing, engaging, and shielding TMDs away from solvent. While this overlap provides cells with multiple “back-ups” to minimize un-chaperoned TMDs, it raises the issue of how any degree of specificity is achieved. This is important as the various TMD binding factors couple to different downstream machinery that ultimately dictate substrate fate. Presumably, parameters like relative abundances, localization, and subtle differences of on-rates of substrate binding combine to establish which TMD binding factor is initially favoured. Interactions among the different factors coupled with substrate off-rates would then determine how substrates partition among factors over time. How the combination of these physical parameters determine substrate specificity has been experimentally examined for mammalian TA pathway factors. It seems that a combination of abundance and relative on-rates favour initial SGTA engagement over either Bag6 or TRC40 (Shao et al., 2017). Substrates can then transfer from SGTA to either TRC40 or Bag6, with TRC40 transfer being both faster and of higher priority. Once the substrate is bound to TRC40 or Bag6, the slow rates of dissociation from either commits the substrate to ER targeting or degradation, respectively. Analogous kinetic analysis of substrate flux through the various TMD binding factors at their appropriate relative concentrations will ultimately be needed to explain how they triage nascent membrane proteins between different potential fates.

### **1.4.2 Quality control factors at the ER: the Hrd1 and Derlin families**

While Sec61 and Get1/Get2 recognize TMDs for the purpose of insertion, the ER contains other TMD recognition factors that participate in degradation. These factors are part of the ERAD machinery that is responsible for patrolling the ER membrane and disposing of incorrectly assembled proteins (Vembar and Brodsky, 2008). This machinery must recognize inappropriately ordered TMDs and subsequently extract membrane proteins out of the lipid bilayer for degradation, which occurs in the cytosol and is facilitated by the proteasome. The factors that mediate this recognition are incompletely defined but include the conserved Hrd1 and Derlin families. Other factors are responsible for recognizing misfolded proteins that are either on the luminal or cytosolic face of the ER, but these are not discussed here.

## Chapter 1: Introduction

Hrd1 is a highly conserved ER membrane protein with a cytosolic E3 ligase domain. It was initially discovered as a factor needed to regulate the degradation of HGM-CoA reductase (Bays et al., 2001). In contrast to Sec61's role in forward translocation, Hrd1 is thought to form a channel across the membrane for retro-translocation (Baldrige and Rapoport, 2016; Stein et al., 2014). A recent cryo-EM structure of Hrd1 in complex with its tightly bound partner Hrd3 shows that Hrd1 exists as a homodimer with each subunit containing eight transmembrane segments (Schoebel et al., 2017). Five of these helices form a cytosolically exposed aqueous cavity that spans more than halfway across the membrane. The cavity is sealed lumenally by a two-layer seal of hydrophobic residues and laterally by TMD1 from a neighbouring Hrd1. While the functional state of this structure is uncertain, it is attractive to posit that substrate TMDs can engage Hrd1 at the lateral gate at the position normally occupied by TMD1. The hydrophilic features of the aqueous cavity might favour recognition of partially hydrophilic TMDs, exposure of which could be a cue for inferring misfolding or mis-assembly.

In addition to Hrd1, certain misfolded substrates require a Derlin family member for degradation (Lilley and Ploegh, 2004; Ye et al., 2004). Derlins are six TMD ER proteins that are inactive members of the Rhomboid superfamily of intra-membrane proteases (Greenblatt et al., 2011). Rhomboids must first selectively recognize TMDs before catalysing their cleavage. Inactive Rhomboids (iRhoms) analogously have TMD recognition abilities, which are instead exploited for the purposes of modulating client trafficking and turnover (Freeman, 2014). It is therefore likely that Derlins have adapted similar properties to recognize the TMDs of misfolded proteins and deliver them to Hrd1. This would help rationalize the existence of multiple Derlin family members (three in mammals), as each would have differing specificities allowing for the recognition of diverse degradation substrates. The structural basis for how Derlins can distinguish normal from aberrant TMDs is currently unknown.

### 1.5 Overview of thesis

The goal of this thesis is to investigate how biosynthetic machinery can effectively accommodate the broad biophysical range of IMP TMDs. This is explored both in the context of multi-pass IMPs and TA proteins. A combination of cell-based and *in vitro* approaches is used to address these issues. In the next chapter, a model multi-pass IMP, the beta-1 adrenergic receptor ( $\beta$ 1-AR), is characterized in an *in vitro* biochemical system. Results suggest that the insertion of  $\beta$ 1-AR is a complicated process that cannot be fully explained by current models of insertion, prompting the idea that factors in addition to Sec61 may be important for mediating  $\beta$ 1-AR biosynthesis. This is pursued in the form of a candidate-based screen in mammalian cells using  $\beta$ 1-AR as a reporter. The results are shown after the screen is first presented, characterized, and validated. A notable hit is the ER membrane complex (EMC), whose absence results in aberrant  $\beta$ 1-AR production.

In Chapter 3, the insertion of structurally simpler TA proteins will be considered. Insertion into the ER membrane of a subset of TA proteins of low-to-moderate hydrophobicity is found to be independent of known post-translational insertion routes. How these TA proteins get inserted becomes of immediate interest. The cytosolic machinery responsible for keeping these substrates in an insertion competent state will be identified and mechanistically dissected in an *in vitro* system in Chapter 4.

In Chapter 5, the insertion mechanism of these TA proteins will be explored. The EMC will be shown as essential for TA protein biogenesis in cells, suggesting that it is a critical component of the insertion machinery. Finally, in Chapter 6, the role of the EMC as a TMD insertase will be rigorously established in a reconstituted system with purified components. The EMC is shown to be both necessary and sufficient to mediate the insertion of TA TMDs of low-to-moderate hydrophobicity. The broader implications of this observation and speculation about the subsequent role of the EMC in multi-pass IMP assembly is discussed in the final chapter.

## Chapter 2: Mechanisms of multi-pass integral membrane protein insertion

The majority of integral membrane proteins (IMPs) rely on the co-translational pathway for targeting and insertion at the endoplasmic reticulum (ER) membrane. This means that they are recognized in the cytosol by the signal recognition particle (SRP), and subsequently delivered to the ER membrane where they are transferred to the Sec61 translocon, which mediates the insertion process. Of IMPs, a significant proportion constitute multi-pass IMPs. Structurally, this class of proteins is defined by virtue of having more than two, and up to as many as twenty transmembrane domains (TMDs). Often, multi-pass IMPs must adopt complex folded conformations in the lipid bilayer which are necessary for their functional activity. Final folded structures often depend on specific interactions between TMDs to form tightly packed membrane spanning bundles. However, TMDs of multi-pass IMPs vary significantly in terms of their biophysical features such as hydrophobicity, helical propensity, and the presence of charged residues. How the translocation machinery manages to correctly identify these TMDs, integrate them into the membrane, and facilitate their subsequent folding remains unclear.

To dissect the mechanisms of multi-pass IMP insertion and assembly, we characterized a G-protein-coupled receptor (GPCR), the beta-1 adrenergic receptor ( $\beta$ 1-AR) as a model substrate in an *in vitro* system.  $\beta$ 1-AR TMDs span a considerable hydrophobic range, yet all are ultimately inserted into the ER membrane and assume their correctly folded form. However, *in vitro* analysis showed that individual TMDs do not necessarily insert autonomously, indicating the need for contextual information for their integration into the membrane. Analysis of serial truncations of  $\beta$ 1-AR show that certain TMDs can only successfully insert in the presence of neighbouring TMDs. This implies that insertion is not a simple biophysical process consisting of the serial partitioning of hydrophobic stretches into the lipid bilayer as they reach the Sec61 translocon, and current models fail to explain how this is achieved in mechanistic detail.  $\beta$ 1-AR is thus established as a tractable substrate which can be used to uncover the sequence of events of multi-pass IMP insertion.

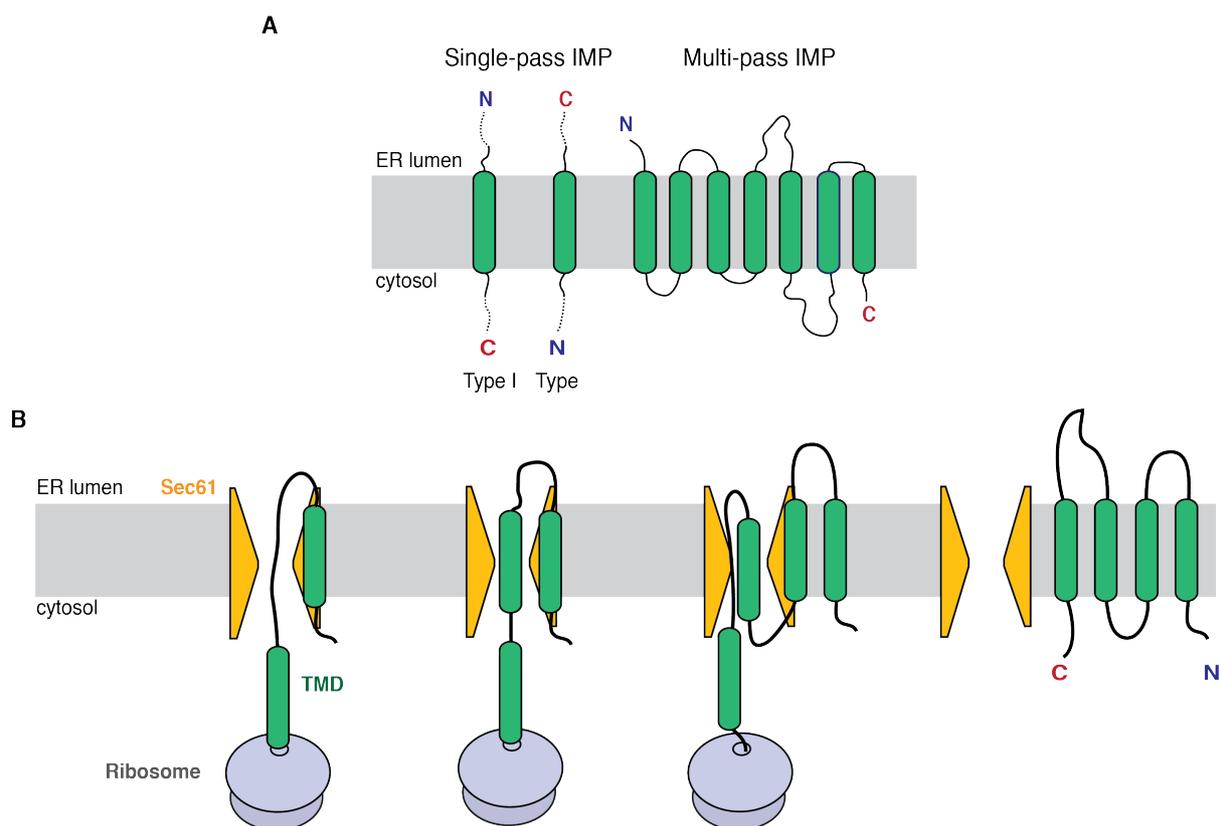
The known physical limitations of the Sec61 translocon, in addition to the behaviour of  $\beta$ 1-AR TMDs suggests that additional factors may be needed for multi-pass IMP biosynthesis. To address this possibility, an *in vivo* siRNA genetic screen was conducted in mammalian cells. This screen was designed to focus specifically on the TMD insertion and assembly process rather than protein targeting. For this, a candidate list of 127 Sec61 complex

## Chapter 2: Mechanisms of multi-pass IMP insertion

associated ER membrane proteins was generated by affinity purification followed by mass spectrometry analysis. A thermostable variant of the  $\beta$ 1-AR was used as a representative multi-pass IMP. Stable cell lines expressing a fluorescent reporter conjugated to  $\beta$ 1-AR were analysed by fluorescence activated cell sorting (FACS) following knockdown of Sec61 associated factors. The screen was validated using microscopy and known chemical manipulations predicted to affect  $\beta$ 1-AR at specific steps of biosynthesis. Of the candidates queried, many did not have an appreciable effect on  $\beta$ 1-AR expression. However, factors previously implicated in multi-pass IMP maturation such as components of the glycosylation machinery, the signal peptidase complex, and the TRAP complex were recovered. Known members of the ER associated protein degradation (ERAD) machinery were also prominent hits, further validating the screening approach. A significant series of hits included multiple subunits of ER membrane protein complex (EMC). The EMC is an abundant ER-resident protein complex genetically implicated in multiple aspect of ER homeostasis. Our results suggest the possibility that the EMC is somehow important for handling the diverse TMDs of multi-pass IMPs, perhaps working in conjunction with the Sec61 complex to promote their insertion.

### **2.1 Multi-pass IMP TMD diversity and the beta-1 adrenergic receptor ( $\beta$ 1-AR)**

IMPs are structurally defined on the basis of number of TMDs and their relative orientation in the ER membrane (Shao and Hegde, 2011a) (Figure 2.1A). Roughly, these can be divided into single-pass and multi-pass IMPs. Single-pass IMPs contain one TMD and can be further subdivided into Type I and Type II to indicate the topology of the N- and C-terminal domains across the membrane. By contrast, multi-pass IMPs are threaded repeatedly across the lipid bilayer. Here, it is thought that the insertion of the first TMD, which can occur either with the N-terminus facing the cytosol or ER lumen, effectively establishes the topology for the subsequent polypeptide chain (Skach, 2009). A significant portion of IMPs contain a signal sequence, which directs their N-terminus into the ER lumen and causes the first TMD to adopt a Type I orientation. Following insertion, TMDs must interact such that the protein adopts a final, functionally active conformation. Integration of both single-pass and multi-pass IMP TMDs into the membrane is thought to occur via the Sec61 translocon. Previous biochemical characterization of the Sec61 translocon in conjunction with recent structural information



**2.1 Structural diversity and insertion model of IMPs.** (A) Integral membrane proteins (IMPs) can be broadly categorized on the basis of the number of TMDs and their relative topology across the lipid bilayer. (B) This schematic cartoon represents the broad thinking about the insertion of multi-pass IMPs into membranes. The sequential insertion model posits that TMDs (green) emerging from the ribosome (grey) are serially passed to the Sec61 translocon (yellow) which partitions them into the membrane

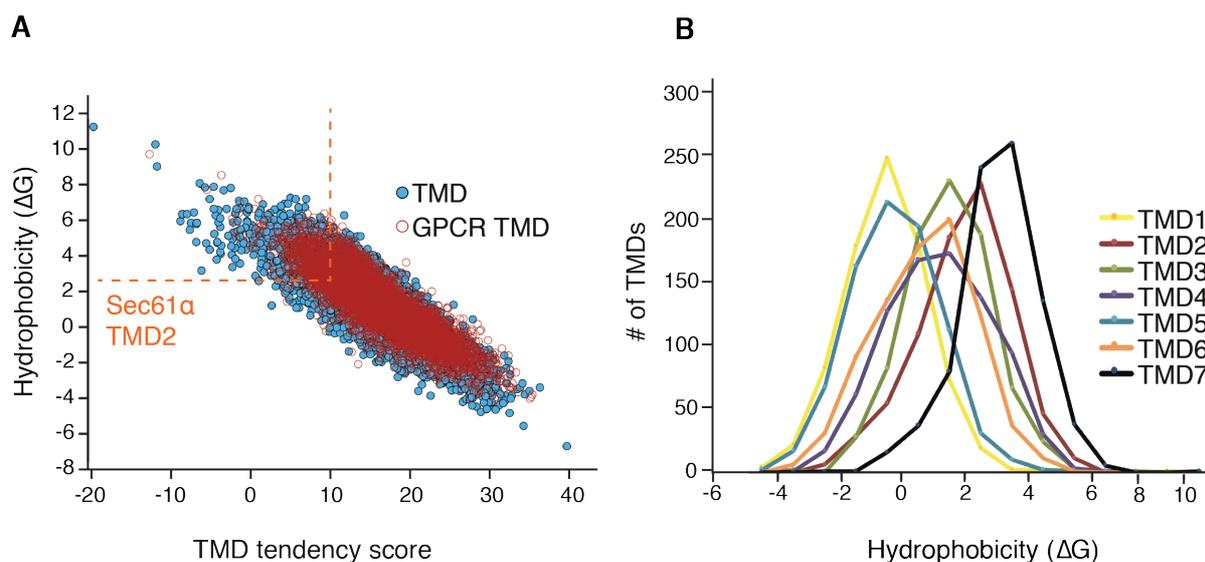
posits that access to the lipid bilayer is contingent on the ability of a TMD to intercalate at the translocon's lateral gate (Voorhees et al., 2014). This model is sufficient to explain the insertion of relatively simple single-pass IMPs. However, numerous characteristics of multi-pass IMP TMDs challenge our current understanding of Sec61-mediated insertion and implicate the involvement of additional factors.

The first models of multi-pass IMP insertion posited that TMDs are successively integrated into the ER membrane by Sec61 as they emerge from the ribosome (Blobel, 1980; Wessels and Spiess, 1988) (Figure 2.1B). This 'sequential insertion' model implies that individual TMDs should be able to autonomously drive their own insertion into the lipid bilayer. For this, TMDs must presumably have a robustly recognizable hydrophobic sequence and local sequence elements that direct their alternating orientation across the membrane. However, eukaryotic multi-pass IMPs are extremely diverse and often contain TMDs that are polar, unusually long or short, or have poor helical propensity (Elofsson and Heijne, 2007). Of these, the most important characteristic is probably TMD hydrophobicity, as emergent TMDs

## Chapter 2: Mechanisms of multi-pass IMP insertion

likely have to displace Sec61 $\alpha$ 's helix 2 to gain access to the lipid bilayer (Voorhees et al., 2014).

To get a more accurate sense of the hydrophobic range spanned by multi-pass IMP TMDs, we interrogated the human proteome using UniProt. A search using the discriminator "Multi-pass membrane protein" (SL-9909) yielded a list of IMPs including the identification and location of any signal sequence, the positions of the predicted TMDs, and the complete protein sequence. Only reviewed entries in the human protein data base were kept. Beta barrel proteins as well as other proteins destined for the mitochondria were removed manually. Members of the GPCR family were annotated as such. TMD hydrophobicity was then calculated using the transmembrane tendency score as defined by Zhao and London (Zhao and London, 2006). The  $\Delta G$  values were also determined (Hessa et al., 2007). These two measures were simultaneously evaluated for Sec61 $\alpha$ 's helix 2. Transmembrane tendency and  $\Delta G$  values are strongly correlated (Figure 2.2A). As expected, TMD hydrophobicity spans a broad range, with the integration of many TMDs into the lipid bilayer predicted as not being energetically



**2.2 Hydrophobic diversity of IMP TMDs.** (A) All TMDs of IMPs inserted into the ER membrane were plotted on the basis of their hydrophobicity according to two scales: TM tendency (x-axis) as defined by Zhao and London (Zhao and London, 2006), and  $\Delta G_{\text{app}}$  (y-axis) as defined by Hessa et al. (Hessa et al., 2005). TMDs which are part of G-protein-coupled receptors (GPCRs) are specifically highlighted (red). TMDs of IMPs span a broad range of hydrophobicity, with many falling short of the hydrophobicity threshold set by Sec61 $\alpha$  translocon's helix 2. (B) The hydrophobicity of GPCR TMDs were analysed based on their relative position in the protein sequence. All GPCRs contain seven TMDs. Certain trends become apparent, such as TMD1 (yellow) being generally more hydrophobic than TMD7 (black).

## Chapter 2: Mechanisms of multi-pass IMP insertion

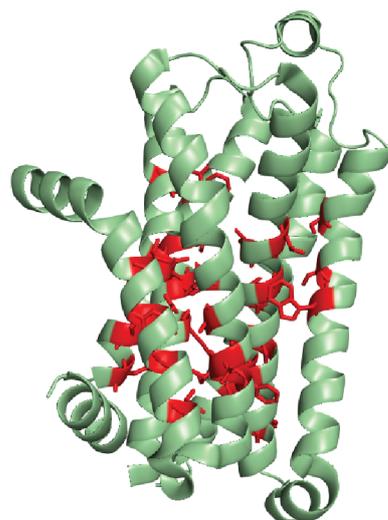
favourable. Notably, a significant number of these TMDs fall below the hydrophobicity threshold of Sec61 $\alpha$ 's helix 2.

To shed light on the sequence of events that leads to the integration and assembly of multi-pass IMP TMDs into the ER membrane, we established an *in vitro* system using a model substrate. We posited that the principles of multi-pass IMP insertion are widely applicable, making the strategy of focusing on a single multi-pass IMP a potentially fruitful one. We chose to specifically focus on the GPCR family. GPCRs are a structurally diverse class of membrane receptors with essential roles in cellular homeostasis and cell signalling. Notably, roughly one third of all marketable drugs target GPCRs (Hauser et al., 2017; Rask-Andersen et al., 2014; Santos et al., 2017). Apart from their biological importance, GPCRs have been extensively studied at the structural-functional level, are highly conserved and widely expressed across eukaryotes, and require accurate folding to carry out the complex conformational changes necessary for signalling. Aberrations in GPCR assembly can have severe consequences for the cell. These characteristics make GPCRs an ideal starting point for studying multi-pass IMP biosynthesis.

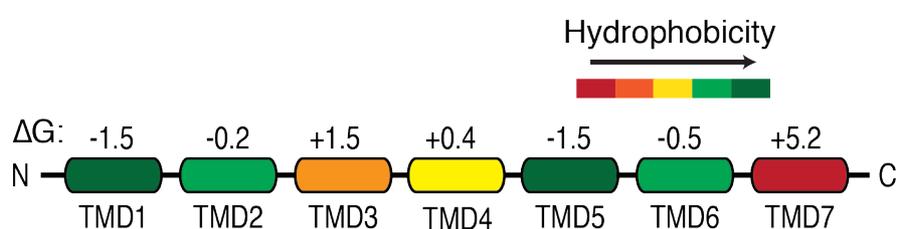
GPCRs are all initially assembled at the ER and trafficked to the plasma membrane where they exert their physiological function. All GPCRs contain seven TMDs and are oriented such that their N-terminus is inserted into the ER lumen, and eventually faces the extracellular space. TMDs of GPCRs are broadly representative of all multi-pass IMPs (Figure 2.2A). Interestingly, the position of a TMD in a GPCR seems to be a predictive factor for TMD hydrophobicity, with TMD1 most likely to be favourably inserted into the lipid bilayer, while TMD7 is consistently poorly hydrophobic (Figure 2.2B).

From the GPCR class of proteins the beta-1 adrenergic receptor ( $\beta$ 1-AR) was selected.  $\beta$ 1-AR is predominately expressed in the heart, where its stimulation results in increased heart rate (Madamanchi, 2007). We chose a modified turkey  $\beta$ 1-AR thermostable variant whose crystal structure is known to 2.7 Å resolution (Figure 2.3A) (Warne et al., 2008). Structural information allows the mapping of conserved intermembrane interaction sites which are essential for the final folded conformation (Venkatakrisnan et al., 2013). As well, the insertion propensity of each TMD can be calculated without relying on TMD prediction algorithms. The range of  $\Delta$ G values of  $\beta$ 1-AR is representative of those found across all multi-pass IMPs (Figure 2.3B).

A



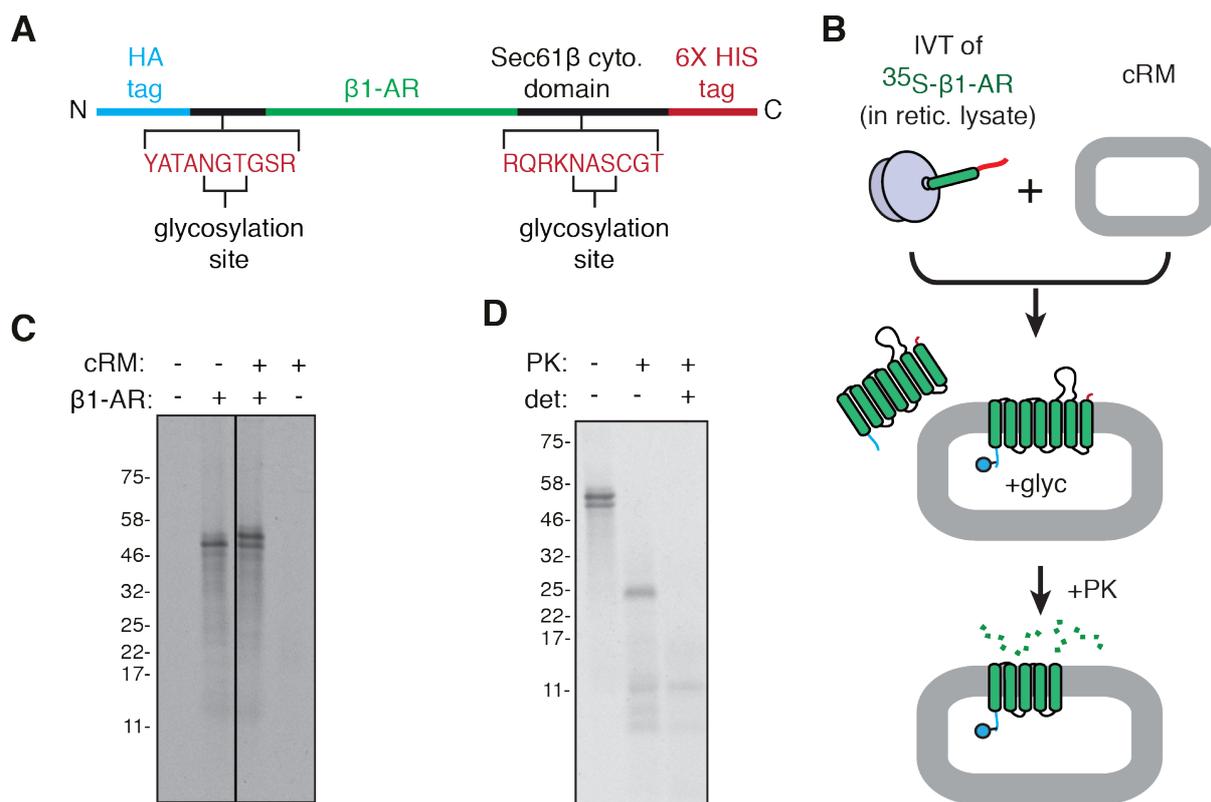
B



**2.3 Structural characteristics of  $\beta 1$ -AR.** (A) Crystal structure of the modified turkey  $\beta 1$ -AR chosen as a model multi-pass IMP to 2.7 Å resolution (Warne et al., 2008), with conserved interacting residues highlighted (Venkatakrisnan et al., 2013) (red). (B)  $\beta 1$ -AR has TMDs of various hydrophobicity which are generally representative of those found in all IMPs (Figure 2.2A).

## 2.2 Insertion and folding of $\beta 1$ -AR *in vitro*

The insertion and folding of  $\beta 1$ -AR was reconstituted *in vitro*. For this, the thermostable variant of the turkey  $\beta 1$ -AR was modified by the addition of an N-terminal HA-tag and a C-terminal His-tag (Figure 2.4A). To monitor insertion into the ER membrane, two glycosylation consensus sequences were inserted preceding and succeeding the  $\beta 1$ -AR sequence. This construct was translated in rabbit reticulocyte lysate (RRL) (Pelham and Jackson, 1976) in the presence of  $^{35}\text{S}$ -methionine, which allows for labelling of newly synthesized proteins and their subsequent detection by autoradiography. Canine pancreas derived ER rough microsomes (cRMs) were used as a source of ER in all insertion reactions (Figure 2.4B). Translation in the presence of cRM results in an upwards shift in the majority of translated  $\beta 1$ -AR due to glycosylation, indicating efficient insertion into the ER membrane (Figure 2.4C).



**2.4 *In vitro* assay of  $\beta$ 1-AR insertion and folding.** (A) Diagram of the  $\beta$ 1-AR reporter cassette used for *in vitro* analysis. (B)  $^{35}$ S-methionine labelled  $\beta$ 1-AR was translated in nucleated reticulocyte lysate and incubated with canine pancreas derived rough microsomes (cRMs). Glycosylation (+ glyc) indicates successful insertion. Folding of  $\beta$ 1-AR is probed by treatment with Proteinase K (PK). The correctly folded  $\beta$ 1-AR should form a tightly folded inaccessible core composed of the first five TMDs. (C, D) Insertion reactions (B) and PK digests with immunoprecipitation against the N-terminal HA tag (C) indicate that the majority of  $\beta$ 1-AR is inserted into cRM in the correctly folded form.

To gauge whether the glycosylated population represents properly folded  $\beta$ 1-AR, insertion reactions were subjected to a protease protection assay (Figure 2.4B). Like all GPCRs,  $\beta$ 1-AR has seven TMDs (TMD1-TMD7) connected by three intracellular (ICL1-ICL3) and three extracellular loops (ECL1-ECL3). For  $\beta$ 1-AR, ECL1 and ECL2 are both very short amino acid sequences, while ECL3 protrudes prominently into the cytosol. Therefore, in the correctly folded form, treatment with Proteinase K (PK) should digest ECL3 and leave a remaining stable core composed of TMDs 1 through 5. To confirm that this is the correct fragment, an additional immunoprecipitation step against the N-terminal HA tag was added. Indeed, following PK digestion, this proteolytic fragment is recovered, confirming its identity and orientation in the membrane (Figure 2.4D). These assays suggest that all the necessary factors for correct  $\beta$ 1-AR synthesis are present and functional in cRM.



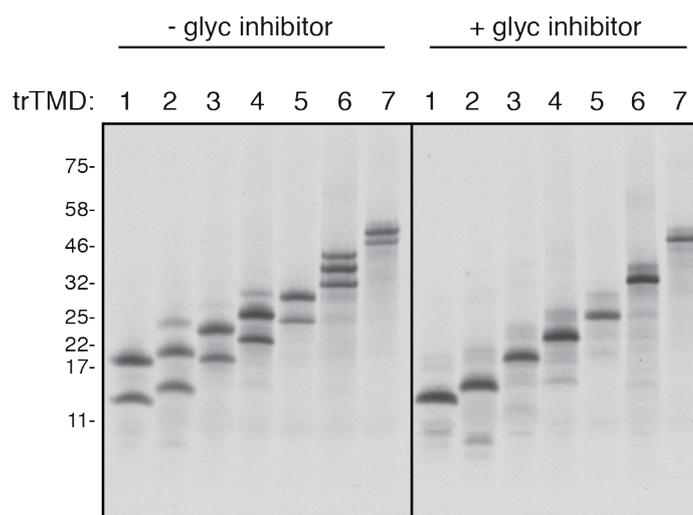
## Chapter 2: Mechanisms of multi-pass IMP insertion

bilayer or fail recognition and be translocated through to the ER lumen. A C-terminal 3F4 was also added to allow for immunoprecipitation of the translated product. Type II cassettes were designed such that the well-established Type I signal anchor protein leader peptide (Lep), which efficiently inserts into cRM (Nilsson and von Heijne, 1993), preceded the test  $\beta$ 1-AR TMD, thereby positioning the test sequence in the Type II orientation upon arrival at the translocon (Figure 2.5B). A 3F4 tag was similarly added on the C-terminus.

To permit co-translational insertion, individual TMDs in their respective cassettes were translated in the presence of cRMs. In order to assess insertion, reactions were subjected to PK digest followed by immunoprecipitation against the N-terminal prolactin epitope for TMDs in the Type I cassette and the C-terminal 3F4 epitope for TMDs in either cassette type.  $\beta$ 1-AR TMDs 1, 4, 5, and 6 all insert successfully into the lipid bilayer (Figure 2.5). However, TMDs 2, 3, and 7 fail to insert when presented in isolation. This failure to recognize TMDs manifests itself as a complete translocation into the ER lumen of TMDs 3 and 7 in the Type I cassette, and residence in the cytosol of TMD2 in the Type II cassette. Notably, behaviour of the isolated TMDs does not correspond with their calculated insertion propensities (ie. TMD4 is shown to insert, while TMD2 does not, despite TMD2 being more hydrophobic). These results suggest that additional contextual information, in addition to that contained within the stretch of amino acids of the TMD itself, is required for the successful insertion of  $\beta$ 1-AR TMDs.

### 2.4 Contextual information is needed for $\beta$ 1-AR TMD insertion

To establish the contextual impact on individual TMD insertion, we analysed serial truncations of  $\beta$ 1-AR (denoted as trTMD1-7) (Figure 2.6). Constructs were subjected to the same co-translational insertion assay as for individual TMDs, followed by a PK digest and immunoprecipitation of the N- and C-terminal fragments. All truncation constructs were in the analogous  $\beta$ 1-AR full length cassette described above (Figure 2.4A), with glycosylation used as the readout for insertion. To aid in the interpretation of the more complex banding patterns generated by PK digestion of truncated  $\beta$ 1-AR constructs, a peptide inhibitor of glycosylation was included in some insertion reactions (Figure 2.6).

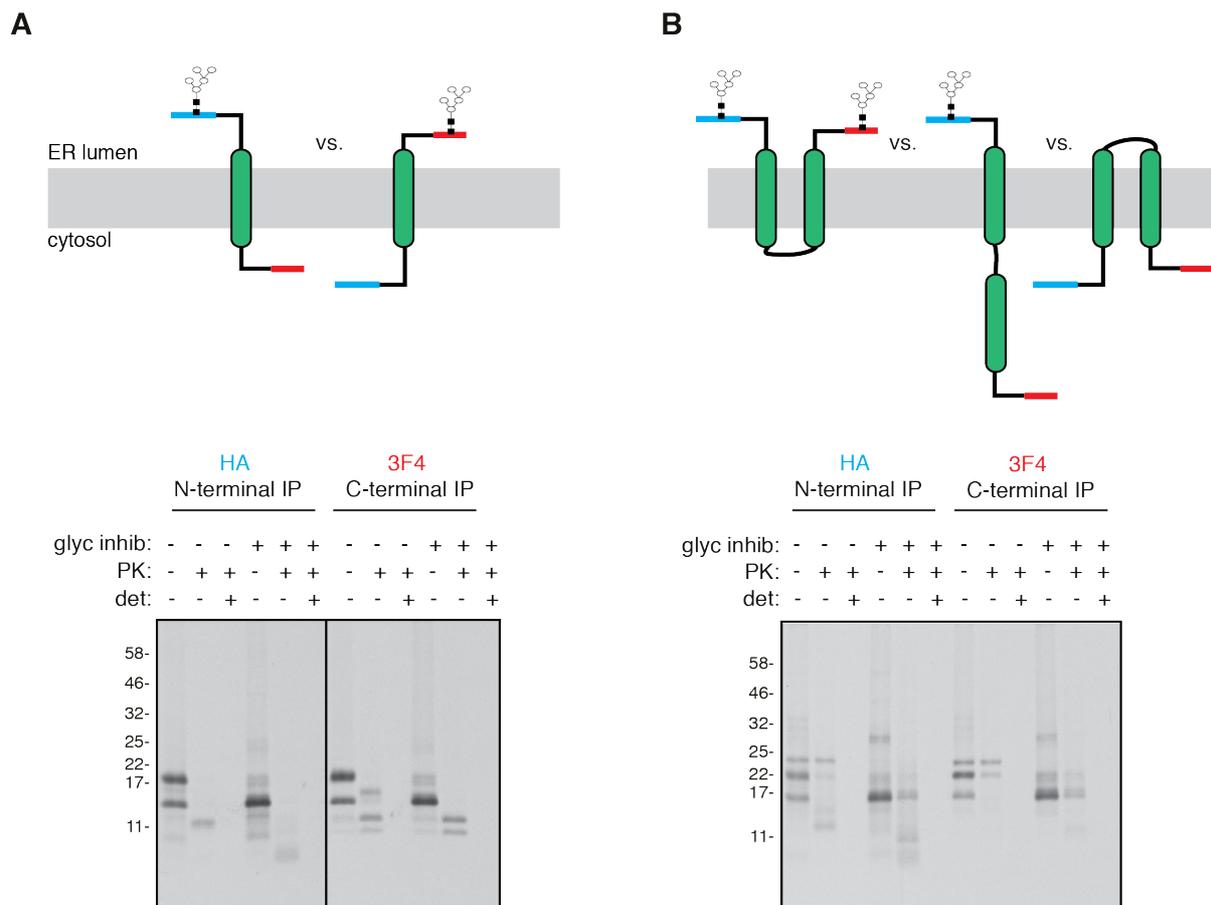


**2.6 Insertion of truncated  $\beta$ 1-AR.**  $\beta$ 1-AR was truncated after each TMD and inserted into the standard cassette (Figure 2.4A), then assayed for insertion (Figure 2.4B) either in the presence or absence of a glycosylation inhibitor.

Despite robust insertion into the membrane, trTMD1 is topologically heterogeneous, with roughly 50% inserted in a ‘backwards’ Type II orientation (Figure 2.7A). Though TMD2 is not capable of inserting in isolation, it can integrate into the membrane in the context of TMD1 (Figure 2.7B). However, trTMD2 still has a heterogeneous topological profile, with a substantial proportion of inserted product oriented incorrectly in the membrane. Indeed, near-homogenous insertion in the correct orientation (~80-90%) is only observed after five or more TMDs (i.e. for trTMD5-7) (Figure 2.8). Even for the full length  $\beta$ 1-AR, there remains a small population that seems to be inserted incorrectly. In all cases, TMDs that cannot insert in isolation (TMDs 2, 3, and 7) seem at least partially capable of doing so when presented in conjunction with their neighbours.

These observations cumulatively raise two key unexplained issues about the insertion of  $\beta$ 1-AR and other multi-pass IMPs. Firstly, since  $\beta$ 1-AR TMDs cannot all insert in isolation, they may briefly reside in either the ER lumen or cytosol until they can interact with either preceding or succeeding TMDs. This has previously been demonstrated for artificially designed multi-pass IMPs with weakly hydrophobic TMDs (Goder et al., 1999; Kauko et al., 2010; Yamagishi et al., 2011). How TMDs may be temporarily stored in these environments, and what dictates their eventual integration into the membrane remains unclear. In addition, it is difficult to discern whether folding of the TMDs begins immediately after the polypeptide integrates into the lipid bilayer or if there is a delay to permit the insertion of the whole protein. Regardless, TMD1 is inserted first, and must wait for TMD7 before it can start making certain

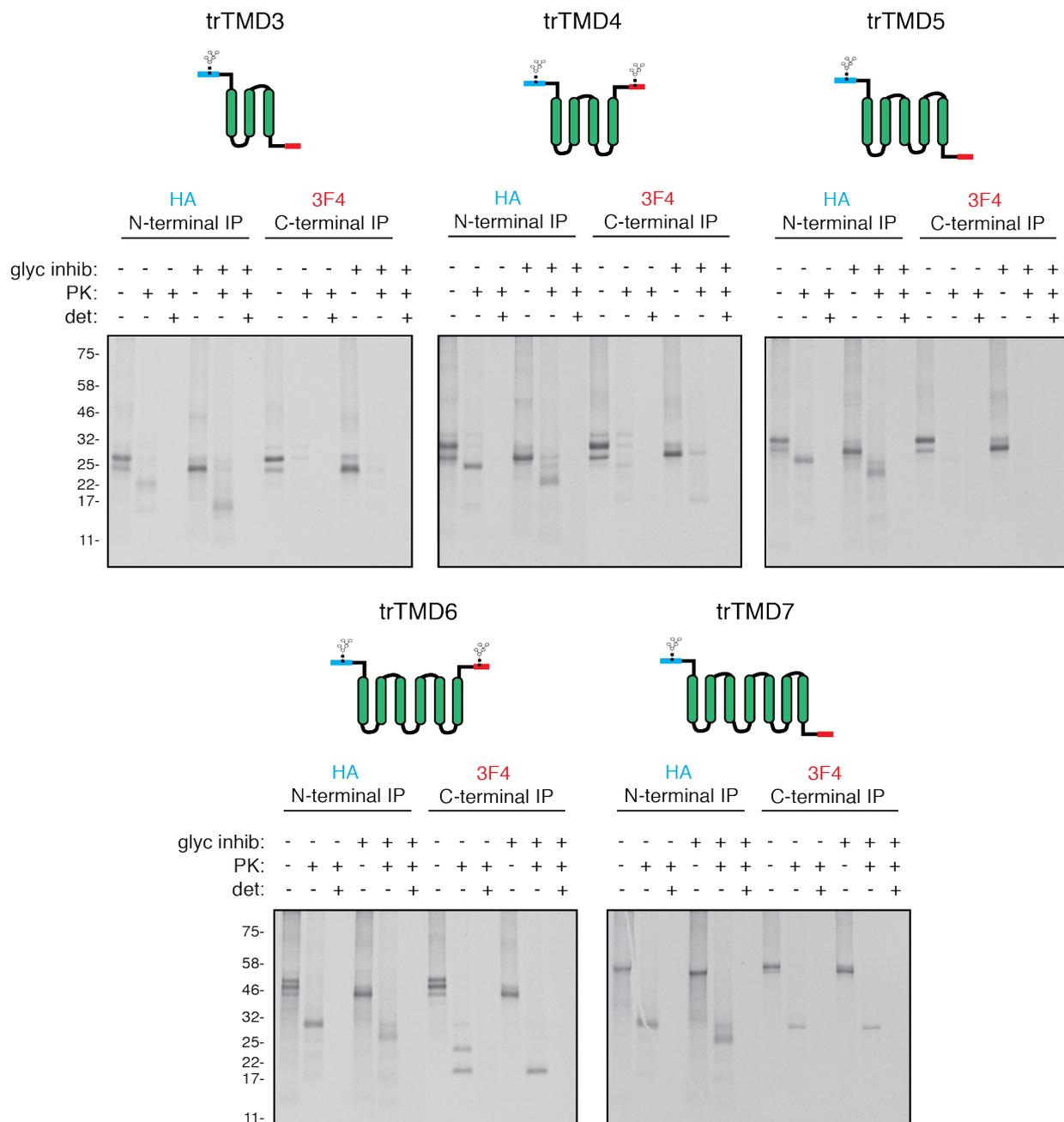
## Chapter 2: Mechanisms of multi-pass IMP insertion



**2.7  $\beta$ 1-AR TMD2 can insert in the context of TMD1.** (A)  $\beta$ 1-AR trTMD1 was assayed for insertion and topology across the membrane (Figure 2.4B). Immunoprecipitation of either the N- or C-terminal tags was used to determine the orientation of the inserted product. (B)  $\beta$ 1-AR trTMD2 insertion was probed the same as in (A). Note that, while TMD2 cannot insert in isolation (Figure 2.5B), it can when preceded by TMD1.

key interactions with it which are necessary for achieving the final folded form (Figure 2.3B) (Venkatakrishnan et al., 2013). How TMD1 and others are ‘held’, whether intermediate folding forms exist, and how these forms are supported in the membrane remain open questions. Secondly, our results suggest that  $\beta$ 1-AR topology may be established relatively late in the biosynthetic process. This would imply that intermediate  $\beta$ 1-AR forms must be able to sample multiple orientations before committing to a particular topology. How topology is established remains unclear. We posit that these additional complications of multi-pass IMP insertion may be resolved by additional factors in the ER that either aid the Sec61 translocon in the insertion process or exist as ‘chaperones’ which can stabilize partially inserted intermediates.

## Chapter 2: Mechanisms of multi-pass IMP insertion



**2.8 Testing orientation of inserted truncated  $\beta$ 1-AR.** The remaining  $\beta$ 1-AR truncations (trTMD3-trTMD7) assayed as in Figure 2.7. Even for the full length  $\beta$ 1-AR (trTMD7) there is a population inserted in the incorrect ‘backwards’ orientation.

## 2.5 Design of a mammalian *in vivo* screen to identify accessory factors

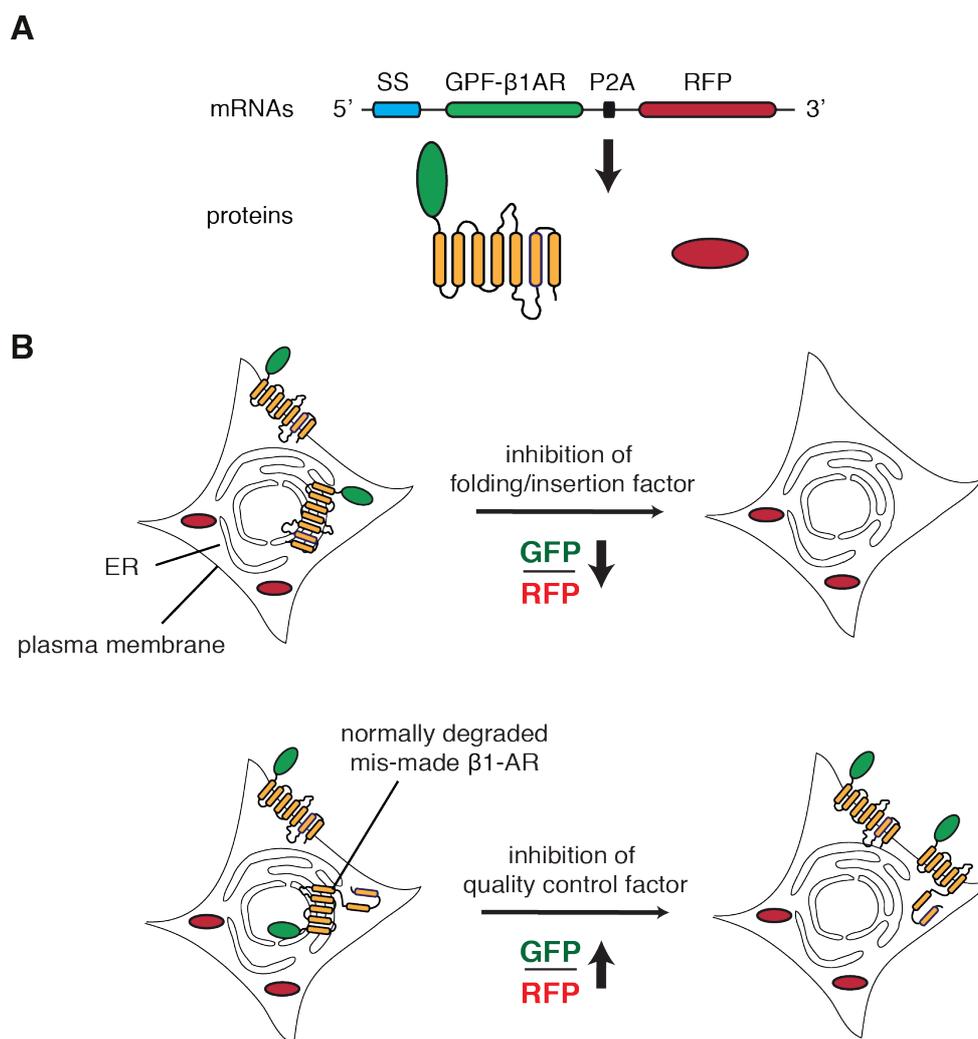
To identify accessory factors that may be important for multi-pass IMP biogenesis, we established an *in vivo* siRNA based screen in mammalian cells. We reasoned that any additional factors needed in multi-pass IMP biosynthesis would be closely associated with the Sec61 translocon. A number of ER-resident proteins have previously been shown to interact with

## Chapter 2: Mechanisms of multi-pass IMP insertion

Sec61 and been implicated in various aspects of translocation. Most notably, these include the translocating-chain association membrane (TRAM) protein, the translocon-associated (TRAP) complex, ribosome-associated membrane protein 4 (RAMP4), Sec62, and Sec63 (Shao and Hegde, 2011a). However, the biochemical roles of all these proteins remains obscure. To this end, a list of 127 candidate Sec61 complex associated ER-resident proteins had previously been generated in the lab by Susan Shao. Briefly, this involved gently solubilizing cRMs to permit native immunoprecipitation of the Sec61 $\beta$  subunit of the translocon, followed by mass spectrometry analysis to identify associated proteins (Appendix 1). It is important to note that this approach biases against cytosolic and ER luminal proteins, which may very well be involved in multi-pass IMP assembly. Presumably, due to the importance of multi-pass IMPs for cellular physiology, completely eliminating a protein necessary for their biosynthesis would be detrimental to cell viability. Therefore, we decided to induce a more acute effect using siRNA knockdown. As such, three different siRNA were acquired for each of the 127 candidate genes (Appendix 1).

To identify potentially subtle effects of accessory proteins on multi-pass IMP biosynthesis, we designed a quantitative screen using FACS to analyse  $\beta$ 1-AR expression at the single-cell level. Reporters were designed to specifically focus on the insertion and assembly process of multi-pass IMPs rather than targeting per se (Figure 2.9A). The thermostable turkey  $\beta$ 1-AR used for *in vitro* analysis was adapted for this purpose. For this, the prolactin signal sequence (SS) was appended to the N-terminal, ensuring efficient targeting to the ER membrane and engagement of the Sec61 translocon, thereby starting the insertion process.  $\beta$ 1-AR conjugated to GFP (GFP- $\beta$ 1AR) was positioned downstream of the SS. This was succeeded by a viral P2A sequence and RFP. The P2A sequence causes the ribosome to skip the formation of a peptide bond, while continuing translation. This results in two separate translation products from one mRNA (Sharma, 2010). The RFP can then be used as an internal control to indicate the amount of originally synthesized  $\beta$ 1-AR.

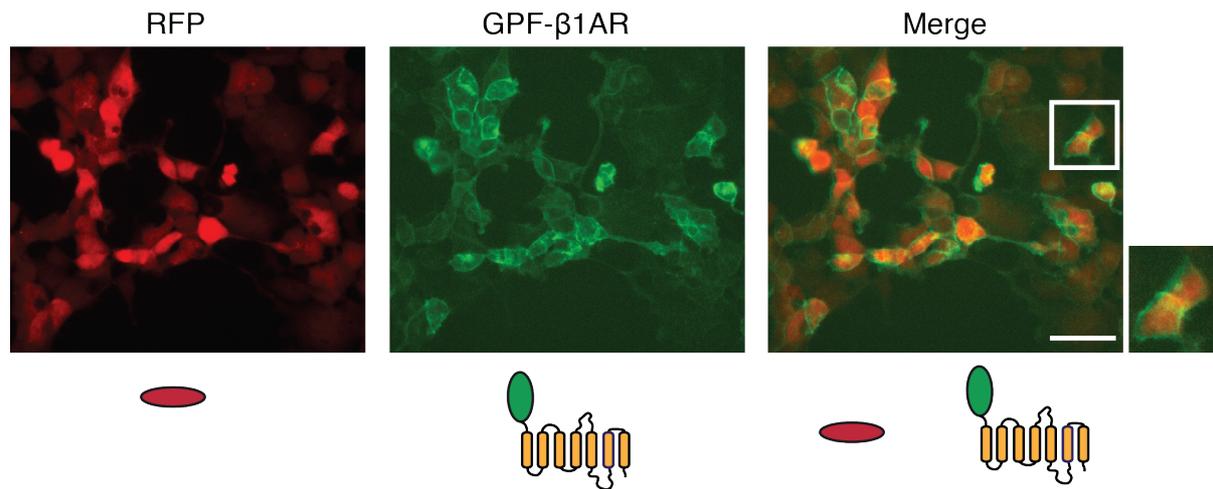
The GFP:RFP ratio therefore represents how much  $\beta$ 1-AR remains in the cell relative to the steady state translation levels (as indicated by RFP) (Figure 2.9B). It can be used as a quick indicator of the potential role of a particular gene on  $\beta$ 1-AR biosynthesis. For example, should a gene that is important for quality control be perturbed (e.g. a component of ERAD), one would expect RFP levels to remain the same, but the amount of GFP- $\beta$ 1AR to increase. This would result in an increase in the GFP:RFP ratio. Conversely, elimination of a protein



**2.9 Design and logic of *in vivo* screen for factors affecting  $\beta$ 1-AR biosynthesis.** (A) Diagram of the construct used for dual-colour FACS analysis of  $\beta$ 1-AR biogenesis in our *in vivo* screen in mammalian cells with knockdowns of Sec61 translocon associated factors. The viral P2A sequence results in skipping of a peptide bond during translation, resulting in two separate translation products (Sharma et al., 2010). In this case, RFP serves as an expression control indicative of the initial amount of translated GFP- $\beta$ 1AR. (B) Expected outcomes of inhibiting factors involved in either  $\beta$ 1-AR biosynthesis (folding/insertion) or quality control. The relative levels of GFP to RFP (GFP:RFP ratio) can be used as a quick indicator of effects on  $\beta$ 1-AR in cells.

necessary for  $\beta$ 1-AR insertion or folding would lead to an increased amount of incorrect  $\beta$ 1-AR, which would presumably get recognized and degraded by ERAD machinery. This would therefore result in a decreased GFP:RFP ratio.

Stable cell lines were generated by the integration of this construct into the Flp-In TRex 293 cell line under an inducible promoter (referred to henceforth as SS-GFP- $\beta$ 1AR -P2A-RFP). To ensure that the construct was behaving appropriately, cells were analysed by microscopy

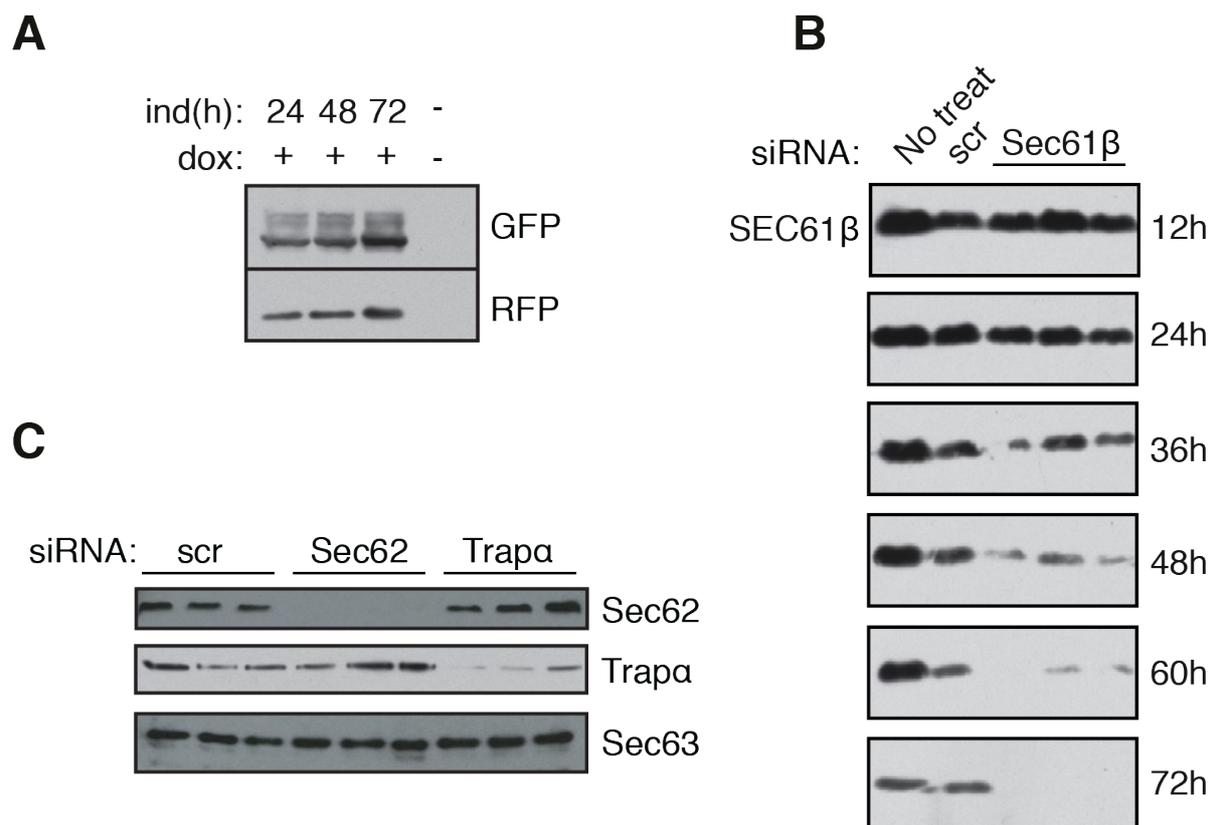


**2.10 Validating *in vivo* screen design by microscopy.** Imaging of induced SS-GFP- $\beta$ 1AR-P2A-RFP stable cell lines (construct described in Figure 2.9A) shows GFP- $\beta$ 1AR localization to the plasma membrane while RFP remains diffusely cytosolic. Scale bar represents 100  $\mu$ m.

following a 24 hour treatment with doxycycline to induce expression (Figure 2.10). As expected, GFP-  $\beta$ 1AR was localized predominately at the plasma membrane, indicating its successful insertion at the ER and subsequent trafficking. RFP was seen only in the cytosol, indicating a functional P2A sequence. All cells had relatively equal expression levels.

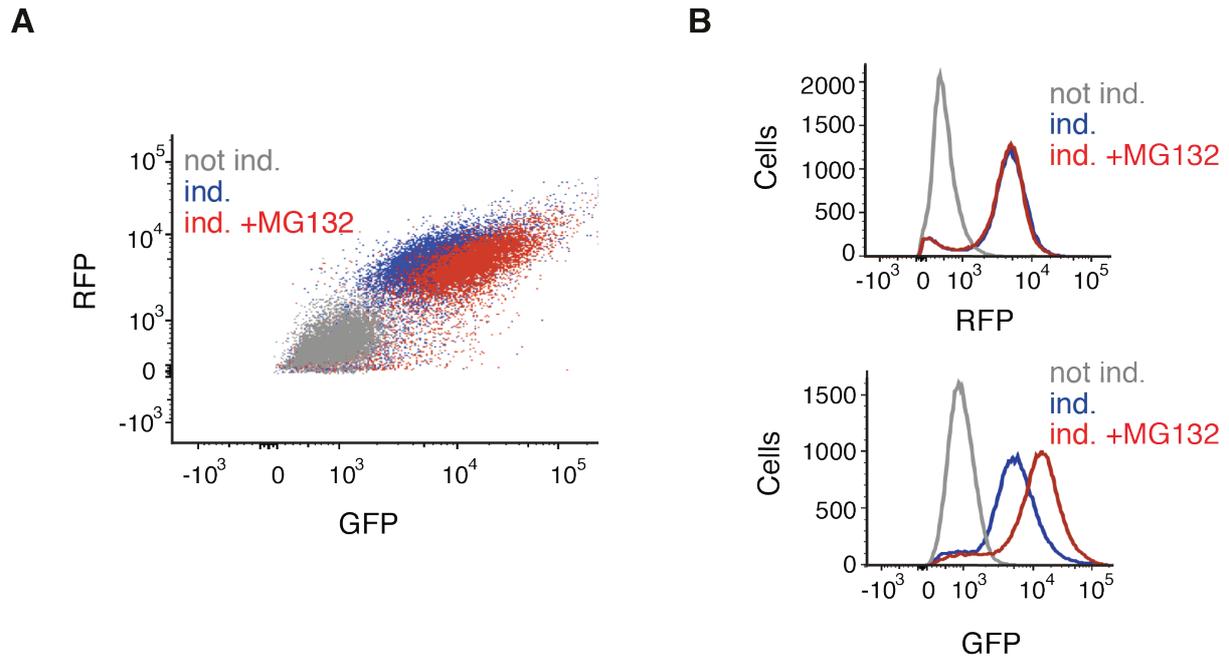
Using the SS-GFP- $\beta$ 1AR-P2A-RFP cell line, optimal conditions for the duration of knockdown in combination with construct induction were determined. We aimed to find conditions that resulted in relatively complete knockdown, while inducing for a sufficiently long time to allow  $\beta$ 1-AR to reach steady state expression levels. For this, cells were induced with doxycycline and harvested after 24, 48, and 72 hours (Figure 2.11A). Samples were then normalized and probed for GFP- $\beta$ 1AR and RFP by blotting. After 24 hours, relatively steady levels of expression are achieved for GFP- $\beta$ 1AR. To determine knockdown time, multiple genes in the list of 127 Sec61 translocon associated factors were knocked down for various time periods followed by immunoblotting (Figure 2.11B). For most, such as Sec61 $\beta$ , Sec62 and Trap $\alpha$ , near complete loss of expression was achieved after 72 hours (Figure 2.11A, B). Obviously, not all proteins queried had available antibodies which could be used to assess knockdown efficiency. As such, for some siRNA tested in the screen, there is the possibility that treatment does not actually reduce gene expression. Regardless, a 72 hour knockdown of a queried gene in combination with a 24 hour induction period of the SS-GFP- $\beta$ 1AR-P2A-RFP were chosen as screen parameters.

An implicit assumption of the screen is that any incorrectly assembled  $\beta$ 1-AR will be degraded through the ERAD pathway. This requires the protein to be extracted from the



**2.11 Optimizing conditions for *in vivo* screen.** (A) SS-GFP- $\beta$ 1AR-P2A-RFP stable cell lines were induced, harvested at various time points, and probed for GFP and RFP levels to establish when steady state expression is achieved. (B) Cells were treated with siRNA against Sec61 $\beta$  and harvested at increments of 12 hours to choose an optimal time for depletion of protein expression, which was determined by blotting. (C) siRNA treatments for 72 hours against Sec62 and Trap $\alpha$ , two ER-resident proteins. Cells were also treated with scrambled (scr) siRNA as a control.

membrane and ultimately degraded in the cytosol by the proteasome (Vembar and Brodsky, 2008). To explicitly test this, cells were acutely treated with the protease inhibitor MG132 following induction and assayed by FACS (Figure 2.12). As expected, this resulted in no change in the level of RFP (Figure 2.12A, B). However, there was an increase in the amount of GFP- $\beta$ 1AR, indicating that even at steady state levels there is a proportion of GFP- $\beta$ 1AR that gets degraded, presumably because it is in an incorrectly made form. The fact that  $\beta$ 1-AR can engage ERAD machinery in this system is promising as it permits the identification of potential accessory factors which affect biosynthesis.



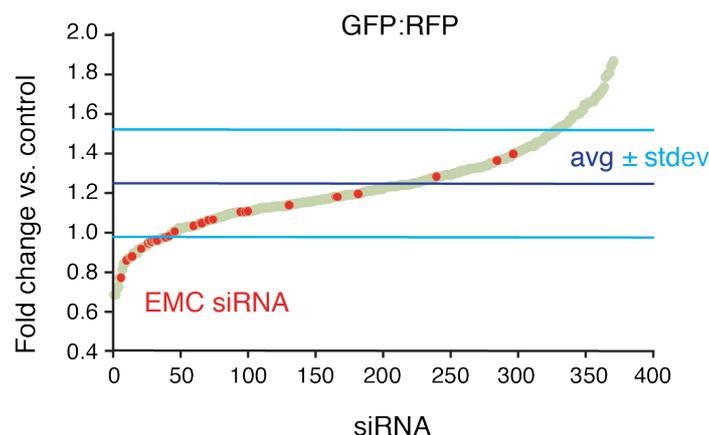
**2.12  $\beta$ 1-AR can engage ERAD machinery** (A) SS-GFP- $\beta$ 1AR-P2A-RFP stable cell lines were induced for 24 hours and treated with 10  $\mu$ M MG132, a general protease inhibitor, for 8 hours prior to analysis by FACS. (B) Changes in the GFP, but not RFP levels, indicate that a sub-population of the GFP conjugated  $\beta$ 1-AR is constitutively degraded when expressed at steady state levels and therefore capable of engaging ERAD machinery.

## 2.6 Mammalian *in vivo* screen results

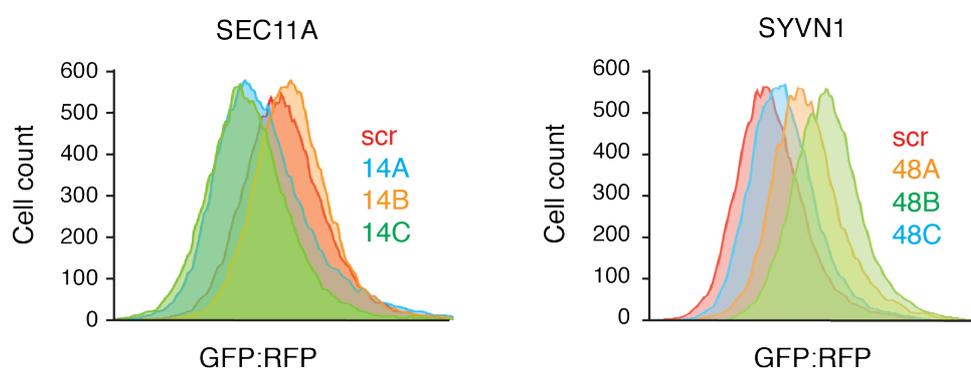
For each siRNA, SS-GFP- $\beta$ 1AR-P2A-RFP cells were knocked down for 72 hours by reverse transfection, and subsequently induced for 24 hours by doxycycline prior to analysis by FACS. For each condition, cells were simultaneously plated and treated with scrambled siRNA as a control. This allowed for normalization to account for slight variations in induction times or plating densities. Immediately prior to analysis by FACS, cells were treated with DAPI stain which permitted the selection of only live cells. Roughly 7,000-10,000 live, expressing cells were analysed for RFP and GFP- $\beta$ 1AR expression for each siRNA queried. For each cell, the RFP and GFP levels were extracted, and a GFP:RFP ratio was calculated.

After all 381 individual siRNA were tested and analysed, the results were plotted as a function of the GFP:RFP ratio relative to the respective scrambled control (Figure 13A, Appendix 1). Representative examples are shown of SEC11A and SYVN1, which resulted in decreased and increased GFP:RFP ratios, respectively (Figure 2.13AB). The mean and standard deviation for all siRNA was calculated. The screen resulted in a number of expected

A



B



**2.13 Results of *in vivo* screen.** (A) Each of the 281 tested siRNA is plotted as a function of the GFP:RFP ratio relative to a same day plated control. Subunits of the ER membrane complex (EMC) are highlighted in red. (B) Examples of siRNA that resulted in either decreases (SEC11A) or increases (SYVN1) in the GFP:RFP ratios, indicative of their role in IMP biosynthesis or degradation respectively (see Figure 2.9B)

hits. For example, perturbations of known ERAD machinery components (SEL1L2, HRD1 and SYVN1) resulted in some of the highest relative GFP:RFP ratios (Ruggiano et al., 2014) (Table 1). However, we were primarily interested in siRNA that caused decreases in the GFP:RFP ratio, as these likely represent factors involved in  $\beta$ 1-AR biosynthesis. We focused on hits that resulted in GFP:RFP values below one standard deviation of the mean

Rank	Gene - siRNA	GFP:RFP
1	SIGMAR1 – 117A	2.69
2	SEL1L2 – 43C	2.59
3	TMED8 – 36C	2.22
4	DERL1 – 39C	2.19
5	SEC22B – 51A	2.12
6	SYVN1 – 48B	2.10
7	SLC39A7 - 104B	1.99
8	TRAM2 – 8C	1.98
9	SEC61A2 – 4A	1.88
10	SEC62 – 5C	1.86

**Table 1 siRNA screen results of highest GFP:RFP ratios.** List of the siRNA which resulted in the highest GFP:RFP ratios. Proteins with roles in degradation (orange) and trafficking (grey) are highlighted.

## Chapter 2: Mechanisms of multi-pass IMP insertion

Rank	Gene - siRNA	GFP:RFP	Rank	Gene - siRNA	GFP:RFP
1	SEC11A – 14C	0.68	21	TMEM85 – 22C	0.91
2	SSR1 – 9B	0.68	22	RRBP1 – 13B	0.92
3	SSR1 – 9A	0.71	23	TMED4 – 32A	0.92
4	SSR1 - 9C	0.72	24	SSR3 – 11C	0.93
5	SEC11A – 14A	0.76	25	LMAN1 – 59A	0.93
6	KIA0090 – 19B	0.77	26	KIA0090 – 19C	0.94
7	RPN2 – 120B	0.81	27	TMED3 – 31B	0.94
8	SPCS2 – 17B	0.84	28	COX4NB – 26B	0.95
9	SPCS2 – 17C	0.84	29	FAM158A – 27B	0.95
10	KIA0090 - 19A	0.85	30	SPCS1 - 16B	0.96
11	OSTC – 123B	0.86	31	TMEM85 – 22B	0.96
12	KRTCAP2 – 124A	0.86	32	MMGT1 – 23A	0.96
13	MMGT1 - 23B	0.87	33	TMEM93 – 24B	0.96
14	C15orf24 - 25A	0.88	34	RRBP1 – 13C	0.96
15	SPCS3 – 18A	0.89	35	TMED6 – 34C	0.96
16	SPCS3 - 12B	0.89	36	TMUB1 – 66B	0.96
17	SPCS3 – 18C	0.89	37	TMED5 – 33C	0.96
18	OSTC – 123C	0.89	38	TMEM85 – 22A	0.97
19	TMED3 – 31A	0.91	39	TMEM93 – 24A	0.97
20	SPCS2 - 17A	0.91	40	OSTC – 123A	0.97

**Table 2 siRNA screen results of lowest GFP:RFP ratios.** List of the siRNA which resulted in the lowest GFP:RFP ratios. Proteins which are part of the TRAP complex (yellow), OST complex (blue), and signal peptidase complex (green) are highlighted. Factors previously implicated in trafficking (grey) are also noted. Hits highlighted in red are components of the EMC complex.

(Table 2). Numerous known factors needed for membrane protein maturation and trafficking were recovered. These included members of the TRAP complex, the signal peptidase complex, and the glycosylation (OST) complex (Shao and Hegde, 2011a). This provided additional confirmation that the screen was resulting in hits indicative of factors necessary for multi-pass IMP biogenesis. A prominent set of hits with low GFP:RFP ratios were members of the ER membrane complex (EMC) (Figure 2.13A, Table 2). The EMC is a poorly characterized complex that has previously been implicated in multi-pass IMP biosynthesis (Bircham et al., 2011; Louie et al., 2012; Richard et al., 2013; Satoh et al., 2015). Our screen indicates that the EMC should be further investigated in this capacity.

### 2.7 Discussion

The biosynthesis of multi-pass IMPs is a complicated process which cannot be fully explained by current insertion models. These conclusions are not novel, as it has long been known that multi-pass IMPs contain TMDs that vary widely in terms of their biophysical properties but must ultimately be accommodated by the Sec61 translocon. To address these issues, we established a tractable *in vitro* system using  $\beta$ 1-AR as a model substrate. This is a first step in permitting the insertion process to be dissected in molecular detail. In addition,  $\beta$ 1-AR is used in an *in vivo* mammalian screen to identify the EMC as a factor which may be crucial for multi-pass IMP assembly.

To aid in the selection of a representative multi-pass IMP, the human proteome was queried to extract information about TMDs contained in this class of proteins. This illustrated the profound differences in TMD hydrophobicity that the translocation machinery must handle.  $\beta$ 1-AR was then chosen as a model substrate and was shown to successfully insert and fold into canine derived rough microsomes. Investigation of the insertion of individual  $\beta$ 1-AR TMDs in isolation showed that three of the seven (TMDs 2, 3, and 7) cannot do so independently. Interestingly, the propensity of insertion of individual TMDs was not necessarily correlated with their hydrophobicity, or whether they adopted a Type I or Type II orientation (Enquist et al., 2009; Hessa et al., 2005, 2007). One early view of TMD insertion into the membrane, which is supported by lipid crosslinking studies, is that TMDs move into the lipid bilayer by passive thermodynamic partitioning through the lateral cleft in the Sec61 $\alpha$  subunit (Heinrich et al., 2000; Martoglio et al., 1995; Rapoport et al., 2004). However, our data suggests that this may not be the case, as clearly not all TMDs are energetically favoured for this process. Newer models of Sec61-mediated insertion, drawing from high resolution structural information, posit that recognition is in some way assessed by a IMP TMD's ability to displace Sec61 $\alpha$ 's helix 2 at the lateral gate for access to the lipid bilayer (Voorhees et al., 2014). However, our bioinformatics analysis shows that a significant proportion of multi-pass IMP TMDs fall well below Sec61 $\alpha$ 's helix 2's hydrophobicity threshold (Figure 2.2A). As such, it remains unclear whether any unifying feature is sufficient to explain if a particular TMD can govern its own insertion into the ER membrane via the Sec61 translocon.

It becomes apparent that additional contextual information may be necessary for inserting TMDs that cannot do so in isolation. Previous studies using chemical inhibition and photo-crosslinking have suggested that TMDs may be able to reside in the translocon for

## Chapter 2: Mechanisms of multi-pass IMP insertion

prolonged time periods and sample different proteinaceous environments prior to their integration into the membrane (Do et al., 1996; McCormick et al., 2003; Meacock et al., 2002; Pitonzo et al., 2009; Sadlish et al., 2005). This leaves open the possibility that TMDs may somehow interact with succeeding or preceding TMD neighbours before insertion. While possible, these interactions must be constrained by the fact that the Sec61 translocon can only physically accommodate one TMD at a time within the central pore (Gogala et al., 2014; Ménétret et al., 2008; Voorhees et al., 2014). It is also possible that TMDs can briefly exist in either the cytosol or ER lumen before being integrated into the lipid bilayer (Goder et al., 1999; Kauko et al., 2010; Yamagishi et al., 2011). Again, this may provide the opportunity to somehow interact with already inserted TMDs, which then drives insertion into the membrane. To test whether insertion of individually incompetent TMDs was stimulated when presented in the context of the proceeding sequence, we analysed truncations of the  $\beta$ 1-AR. Indeed, presenting certain TMDs in the presence of the proceeding polypeptide makes them capable of insertion. This is most readily seen with TMD2, which can only insert when presented after TMD1. It is not clear how TMD1 can affect the insertion of TMD2 and the role of the Sec61 translocon in this process.

Another notable observation from the truncation insertion experiments is that the topology of  $\beta$ 1-AR seems to be established relatively late in the insertion process, implying that the protein can somehow ‘sample’ multiple orientations. How this is achieved within the plane of the lipid bilayer is difficult to visualize. Earlier reports suggest that the first TMD of a multi-pass IMP may be able to ‘flip’ within the Sec61 translocon from a Type I to Type II orientation (Devaraneni et al., 2011) but it is difficult to extrapolate this to longer polypeptide chains. However, this all implies that the sequential insertion model which consists of a succession of translation initiation and termination events cannot explain the insertion of  $\beta$ 1-AR and other multi-pass IMPs (Pitonzo et al., 2009; Sadlish et al., 2005). While this model is sufficient for describing the insertion of substrates such as the Aquaporin 4 channel (AQP4) (Sadlish et al., 2005; Shi et al., 1995), it is becoming clear that this is the exception rather than the rule (Lu et al., 1998; Tu et al., 2000). In fact, the Aquaporin 1 channel (AQP1), another family member of high homology, has a TMD2 that is initially translocated into the ER lumen, resulting in a heterogeneously inserted population that still somehow achieves a correctly folded form (Lu et al., 2000; Pitonzo and Skach, 2006; Skach et al., 1994). Another stipulation of the sequential insertion model is that the first TMD of a multi-pass IMP effectively acts as a signal anchor to definitively determine topology. The fact that only half of rTMD1 is in the correct orientation suggests that this process is not quite so straightforward. These observations

## Chapter 2: Mechanisms of multi-pass IMP insertion

have been made previously when testing the conditions of the sequential insertion model. For example, truncated fusion proteins containing a passive translocation reporter were appended C-terminal to a variety of TMDs as a readout of insertion topology (Skach, 1998). Contrary to model predictions, a large portion of tested TMDs exhibit strange topologic properties that don't mirror their endogenous orientations (Audigier et al., 1987; Moss et al., 1998; Skach and Lingappa, 1993, 1993; Skach et al., 1994; Xie et al., 1996).

Taken together, our *in vitro* experiments prompted us to look for additional factors that may be involved in multi-pass IMP biogenesis. Recent high-resolution structures of the Sec61 translocon make it clear that previous explanations which could reconcile certain observations described above, such as TMDs being inserted 'in bulk', may not be physically feasible. These limitations could be circumvented by the existence of other factors that are capable of interacting with TMDs or insertion intermediates. For this, we conducted a FACS based *in vivo* siRNA screen in mammalian cells using  $\beta$ 1-AR as a reporter. Proteins physically proximal or in association with Sec61 were specifically queried. The reporter construct was designed to contain an N-terminal signal sequence, which would fix the Type I topology of TMD1 and allow us to focus on subsequent TMD insertion and assembly events.

The screen resulted in a number of expected hits, namely members of the ERAD machinery and factors previously implicated in IMP maturation. Of note, the impact of most siRNA on  $\beta$ 1-AR levels was quantitatively relatively mild. This could be due to a number of reasons. Primarily, a thermostable variant of the turkey  $\beta$ 1-AR was used (Warne et al., 2008). While this is convenient for expressing large amounts of protein for structural determination by X-ray crystallography, this variant may be too good at unassisted self-assembly. In which case, removal of accessory factors necessary for multi-pass IMP biosynthesis may have very weak effects. Secondly, there is the possibility of multiple, redundant pathways that support biosynthesis, tempering the effects of loss of any one particular factor.

Despite these relatively weak effects, one prominent and convincing hit was the ER membrane complex (EMC). It is encouraging that multiple subunits of this complex were observed to impact insertion and/or folding of  $\beta$ 1-AR. The EMC has previously been associated with a variety of ER related phenotypes, most notably the biogenesis of multi-pass IMPs. It was first identified in a genetic screen looking for factors that could be involved in protein folding in the ER (Jonikas et al., 2009). Since then, it was proposed to be necessary for the maturation of various multi-pass IMPs in a number of different species. In yeast, EMC2 knockdown leads to decreased expression of the model substrate Mrh1p (Bircham et al., 2011). In more complex organisms such as *C. elegans*, deletion of EMC subunits reduces the cell

## Chapter 2: Mechanisms of multi-pass IMP insertion

surface expression of the nicotinic acetylcholine receptors, five TMD containing proteins that must also oligomerize to be functionally active (Richard et al., 2013). In *D. melanogaster*, loss of EMC subunits is seen to strongly reduce the cell surface expression of the important light receptor rhodopsin-1 (Rh1) (Sato et al., 2015). In HeLa cells, knockdown of EMC2 correlates with decreased the expression of a CFTR cystic fibrosis mutant (Louie et al., 2012). However, all of these studies, as well as our screen, do not specify at which stage of multi-pass IMP assembly the EMC is important, or whether any observed effects are direct.

At this point, we were primed to the notion that the diversity of TMDs of multi-pass IMPs mandate additional factors for successful insertion into the ER membrane. It is possible that the EMC is a factor that may somehow be involved in this process. Either the EMC can directly handle these less favourable TMDs, be involved in directing topology, or serve as a sort of ‘chaperone’ to promote folding. However, a quick glance at the protease digestion patterns of  $\beta$ 1-AR truncations shows that the folding of multi-pass IMPs is a complicated process to study in our *in vitro* biochemical system. Ideally, a simpler system can be used to gain more information about the specifics of how weakly inserting TMDs are handled at the membrane and the possible function of the EMC. For this, we chose to focus on tail-anchored (TA) proteins. This class of proteins is much smaller and structurally simpler, but nevertheless displays a broad diversity of TMDs. TA protein insertion is the subject of the subsequent four chapters.

## 2.8 Materials and methods

### *Plasmids, siRNA and antibodies*

Constructs for expression in rabbit reticulocyte lysate (RRL) were based on the SP64 vector (Promega). The sequence for the thermostable turkey  $\beta$ 1-AR is as previously described (Warne, 2008). Subsequent amino acid positions are relative to this variant. Constructs for expression of the truncated and full length  $\beta$ 1-AR constructs consisted of an N-terminal HA-tag flanked by alanines, followed by a glycosylation consensus sequence (NGT) and then the appropriate  $\beta$ 1-AR truncation. C-terminal to this was appended the cytosolic domain of the Sec61 $\beta$  subunit (aa 1-69) which contains another glycosylation consensus site (NAS) followed by a 6xHis-tag. Note that the glycosylation efficiency of the first site is slightly higher than that for the second, making the interpretation of proportion inserted in a particular orientation slightly more

## Chapter 2: Mechanisms of multi-pass IMP insertion

nuanced (Figure 2.6). Truncations were at the following residues relative to the parent sequence: trTMD1 – 1-46, trTMD2 – 1-81, trTMD3 – 1-118, trTMD4 – 1-159, trTMD5 – 1-213, trTMD6 – 1-248. The Type I cassette that housed TMDs 1, 3, 5 and 7 consisted of the first 150 residues of prolactin (with aa 1-30 containing the signal sequence), followed by an N- and C-terminal flexible linker encompassing the test TMD. Following this was appended the cytosolic domain of the Sec61 $\beta$  subunit (aa 2-59) which contains a glycosylation consensus sequence (NAS) and a 3F4 epitope tag. TMDs constituted the following residues of the parent  $\beta$ 1-AR: TMD1 – 6-40, TMD3 – 83-110, TMD5 – 173-204, TMD7 – 288-316. The Type II cassette containing TMDs 2, 4, and 6 consisted of the Lep protein (aa 1-59 + aa 101-180) followed by a flexible linker containing the test TMD. This was followed by the same C-terminal as for Type I cassettes. TMDs were the following residues of the parent  $\beta$ 1-AR: TMD2 – 43-76, TMD4 – 123-152, TMD6 – 257-287. The *in vivo* reporter was generated by fusing the signal sequence of prolactin (aa 1-30) to GFP followed by  $\beta$ 1-AR. Immediately following was the viral P2A sequence and RFP. Note that the mCherry variant of RFP and the mEGFP variant of GFP were used, but the simpler nomenclature of RFP and GFP are used in the text and figures.

All siRNA including scrambled Silencer Select siRNA was from Thermo Fisher, with the designated catalogue numbers found in Appendix 1. Antibodies against Sec62, Sec63 and Trap $\alpha$  were as previously described (Garrison et al., 2005; Stefanovic and Hegde, 2007).

### ***Mammalian in vitro expression***

Translation reactions in RRL and the preparation of ER derived rough microsomes from canine pancreas (cRM) were minimally modified from standard protocols (Sharma et al., 2010; Walter and Blobel, 1983). Templates for *in vitro* transcription for translation in RRL were generated by PCR using a primer which anneals just upstream of the SP6 promoter and a 3' primer which anneals ~200 downstream of the stop codon (Sharma et al., 2010). Transcription reactions were with SP6 polymerase at 37°C for 1 hour. The reaction was used directly in a translation reaction without further purification as previously described (Sharma et al., 2010). Constructs were translated for 30 min at 32°C. In some reactions, glycosylation was inhibited by the addition of a tripeptide competitive inhibitor at 50  $\mu$ M (Asn-Tyr-Thr). Where indicated in the text, samples were transferred to ice and treated with 1 mg/ml Proteinase K (PK) for 50 min. The reaction was terminated with 5 mM PMSF and left on ice for 5 min before being

## Chapter 2: Mechanisms of multi-pass IMP insertion

transferred to a 5-fold volume of 1% SDS, 0.1 M Tris pH 8 pre-heated to 100 °C. Following certain PK digestion reactions, samples were subjected to an additional immunoprecipitation (IP) step against either the N- or C-terminal tags. For this, samples were diluted 10-fold in IP buffer (100 mM NaCl, 50 mM HEPES, pH 7.6, 1% Triton X-100), and incubated with 2.5 µl 3F4, Prl, or HA antibody and 10 µl Protein A resin for 3 hours at 4 °C. Anti-His tag IPs were done with 10 µl pre-conjugated Ni-NTA beads and diluted 10-fold in His IP buffer (PSB, 250 mM NaCl, 0.5% Triton X-100, 10 mM Imidazole). The resin was washed with the respective buffer three times and eluted with sample buffer before analysis by SDS-PAGE and autoradiography.

### *Cell culture*

Flp-In 293 T-Rex cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) with 10% tetracycline-free foetal bovine serum (FBS) in the presence of 15 µg/ml blasticidin and 100 µg/ml hygromycin. To generate stable cell lines, the SS-GFP-β1AR-P2A-RFP construct was cloned into the pcDNA5/FRT/TO vector and integrated into the FRT locus using the Flp recombinase system as per manufacturer's instructions (Invitrogen). Briefly, cells were grown in a 10 cm tissue culture dish and co-transfected with 6 µg of the construct along with 2 µg of the Flp recombinase expression plasmid (pOG44) using Lipofectamine 2000 (Life Technologies). Cells were selected for 14 days using 100 µg/ml hygromycin. MG132 treatment, where indicated, was for 4 hours at 10 µM.

### *Flow cytometry analysis*

For the screen, cells were reverse transfected in 6 cm tissue culture plates with 10 mM siRNA (see Appendix 1) using the Lipofectamine RNAiMAX reagent as per manufacturer's instructions (Life Technologies). Cells were plated such that they would achieve confluency after approximately 72 hours. Exactly 48 hours after plating, cells were induced with 10 ng/ml doxycycline. Afterwards, cells were washed extensively in ice-cold physiological salt buffer (PSB), detached with 100 mM EDTA and re-suspended in EDTA with 10% FBS. DAPI was added at 1 µg/ml (Sigma Aldrich). Flow cytometry was done on an LSR II (BD Biosciences) and data was analysed using the FlowJo package. Only live, expressing cells were selected for downstream analysis (Itakura et al., 2016).

## Chapter 2: Mechanisms of multi-pass IMP insertion

### *Microscopy*

The localization of GFP-  $\beta$ 1AR and RFP was visualized in fixed cells after induction of the SS-GFP-  $\beta$ 1AR -P2A-RFP Flp-In 293 T-Rex cell line. Briefly, cells were grown on 12 mm (Nunc) polylysine coated cover-slips and induced with 10 ng/ $\mu$ l doxycycline for 24 hours before being fixed in 3.6% formaldehyde. Cells were mounted in SlowFade Gold (Theromo Fisher Scientific) and imaged using an LSM 780 confocal microscope (Zeiss).

## **Chapter 3: TRC-independent post-translational tail-anchored protein insertion**

Tail-anchored (TA) proteins represent an important subset of integral membrane proteins (IMPs). They contain a single C-terminal transmembrane domain (TMD). Due to the proximity of the TMD to the termination codon, TA proteins must be delivered to the endoplasmic reticulum (ER) post-translationally. Similar to the TMDs found in multi-pass IMPs, TA proteins contain TMDs that differ substantially in terms of their biophysical properties (Kalbfleisch et al., 2007). Most TA proteins are thought to use a highly conserved pathway whose key targeting factor is TRC40 (Get3 in yeast), an ATPase which is known to favour strongly hydrophobic TMDs (Favaloro et al., 2008; Mariappan et al., 2010; Schuldiner et al., 2008; Stefanovic and Hegde, 2007). Due to incomplete existing bioinformatics analyses, publicly available databases were first interrogated to curate a list of verified human TA proteins. This permitted the extraction and subsequent analysis of numerous sequence characteristics. Confirming previous results, TA proteins differ significantly from one another in terms of the hydrophobicity of their TMDs (Kalbfleisch et al., 2007).

To determine the behaviour of the TRC pathway in light of this variation, we established an *in vitro* post-translational assay. With this assay, we discover that roughly half of all TA proteins are not effectively inserted via the TRC pathway. The key parameter dictating TRC-dependence is TMD hydrophobicity. Furthermore, in a fully purified system, we show that TA proteins of low hydrophobicity dissociate from TRC40 very rapidly; before they can be effectively targeted to the ER membrane. In this chapter, we establish that alternative means of insertion must exist for a subset of TA proteins of moderate-to-low hydrophobicity which cannot efficiently engage TRC40. This conclusion simultaneously supports the notion that the diverse properties of TMDs necessitates the existence of multiple routes of insertion at the ER membrane.

### **3.1 Biodiversity of TA protein TMDs**

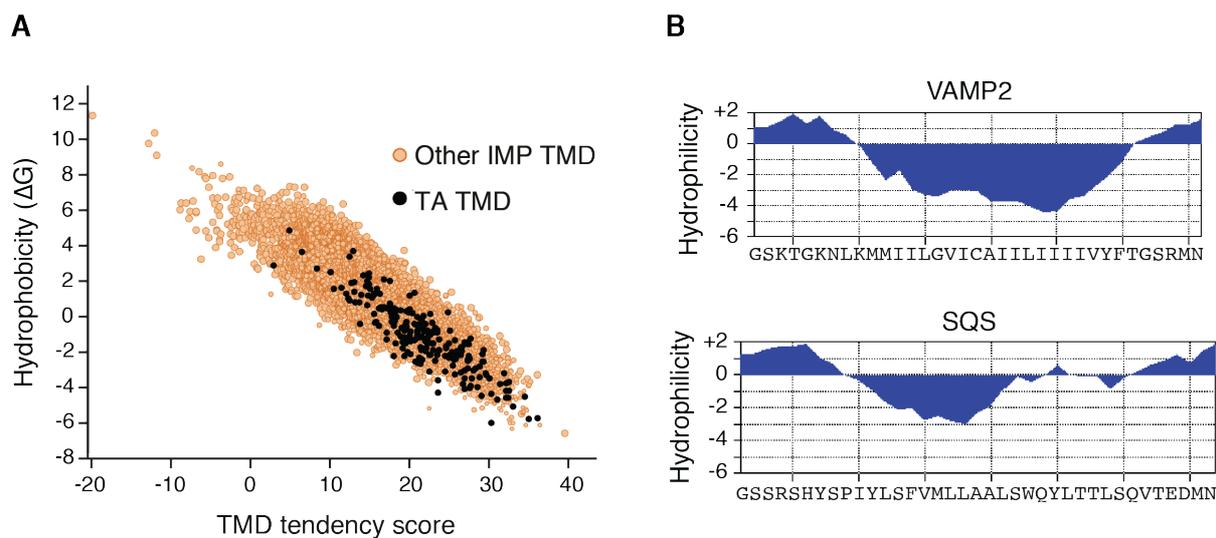
Tail-anchored proteins are a subset of IMPs which are anchored in the membrane by a single C-terminal TMD. Roughly ~3-5% of the eukaryotic proteome encodes TA proteins (Beilharz et al., 2003; Kalbfleisch et al., 2007; Kriechbaumer et al., 2009). They are essential for various

### Chapter 3: TRC-independent post-translational TA protein insertion

cellular processes such as vesicular trafficking, apoptosis, signal transduction, and lipid biosynthesis (Hegde and Keenan, 2011). These functions are generally mediated by the N-terminal cytosolic facing domain which can engage in protein-protein interactions. Two physiologically important classes of proteins with this topology are the Bcl-2 family of apoptosis related proteins and the SNARE proteins of the synaptobrevin/syntaxin families that mediate vesicle targeting and fusion.

By virtue of containing a single TMD, TA proteins are structurally simpler than multi-pass IMPs. However, the proximity of the TMD to the termination codon means that translation ends before the TMD can exit the ribosomal tunnel. This prevents SRP from effectively engaging TA substrates for co-translational delivery to the ER membrane and they instead rely on post-translational means of insertion (Kutay et al., 1993, 1995). How these proteins are captured in the cytosol, chaperoned, and inserted into the membrane has been worked out in some detail. The highly conserved TRC (GET in yeast) pathway mediates the insertion of most TA proteins (Chio et al., 2017; Hegde and Keenan, 2011). The central cytosolic factor is the ATPase TRC40 (Get3 in yeast). Structural studies of Get3 indicate that it exists as a homodimer which contains a deep hydrophobic groove that can bind and shield the TMD of TA proteins (Bozkurt et al., 2009; Hu et al., 2009; Mateja et al., 2009; Suloway et al., 2009; Yamagata et al., 2010). TRC40 then delivers substrates to its dedicated ER-resident receptor (the heterodimer WRB/CAML or Get1/Get2 in yeast) which facilitates substrate insertion into the lipid bilayer (Mariappan et al., 2011; Schuldiner et al., 2008). The surface properties of the substrate-binding groove of Get3 are consistent with biochemical studies showing that there is a strong preference for TMDs of particularly high hydrophobicity (Mariappan et al., 2010; Mateja et al., 2009). Yet, even a cursory glance at the TMD sequences of ER-targeted TA proteins shows that they vary significantly in terms of their hydrophobicity.

Despite the physiological importance of TA proteins, there has not been extensive bioinformatics analysis aimed at identifying the full complement found in eukaryotes. Previous studies have relied on conventional computational methods to predict TA proteins, integrating results from multiple TMD prediction tools such as TMHMM and signal prediction tools such as SignalIP and TargetIP (Emanuelsson et al., 2007; Kalbfleisch et al., 2007; Krogh et al., 2001; Petersen et al., 2011; Sonnhammer et al., 1998). More recently, a machine-learning based technique, TAPPM (Tail-Anchored Protein Prediction Method), has been described which employs Hidden Markov Models (HMMs) to distinguish TA proteins from other membrane



**3.1 Diversity of TA TMDs.** (A) The TMDs of TA proteins (black) were plotted relative to all other TMDs of IMPs (orange) on the basis of their hydrophobicity according to their TM tendency score (x-axis) (Zhao and London, 2006) and  $\Delta G_{app}$  (y-axis) value (Hessa et al., 2005). (B) Hydrophilicity plots (using the Kyte-Doolittle scale) (Kyte and Doolittle, 1982) for selected TMDs which are of low (SQS) and high (VAMP2) TMD hydrophobicity. Note that different hydrophobicity scales are generally well correlated (Zhao and London, 2009).

proteins using only target amino acid sequences (Shigemitsu, 2016). Though this method provides relatively accurate identification of established TA proteins, predicted TA proteins are not validated by cross-referencing with other databases. Notably, these two types of analyses yield different numbers of TA proteins (Kalbfleisch et al., 2007; Shigemitsu et al., 2016).

Due to conflicting information in the literature and incomplete analysis of TMD properties, we first sought to curate all human TA proteins. This was done by querying UniProt using the discriminator “single-pass membrane protein” (SL-9904). Only reviewed entries in the human protein database were retained. Outputs of this search included the complete sequence of the protein, the presence or absence of a signal peptide and transit peptide, and the position of the predicted TMD. Proteins with annotated mitochondrial localization were removed manually, leaving only those with strong evidence of ER/secretory pathway localization. The remaining proteins were subsequently divided into single pass signal anchored proteins or TA proteins. TA proteins were defined as those which had a C-terminal tail of 50 residues or less following the TMD. Defining a TA protein based on the 50 residue C-terminal threshold is rooted in several reasons. First, this is the length needed to engage the Sec61 channel in a looped orientation while still a ribosome nascent chain, as roughly 35 residues are in the ribosomal tunnel while the remaining 15 would span the Sec61 channel (Voorhees et al., 2014). Additionally, ER proximity biotinylation experiments in yeast show that there are essentially

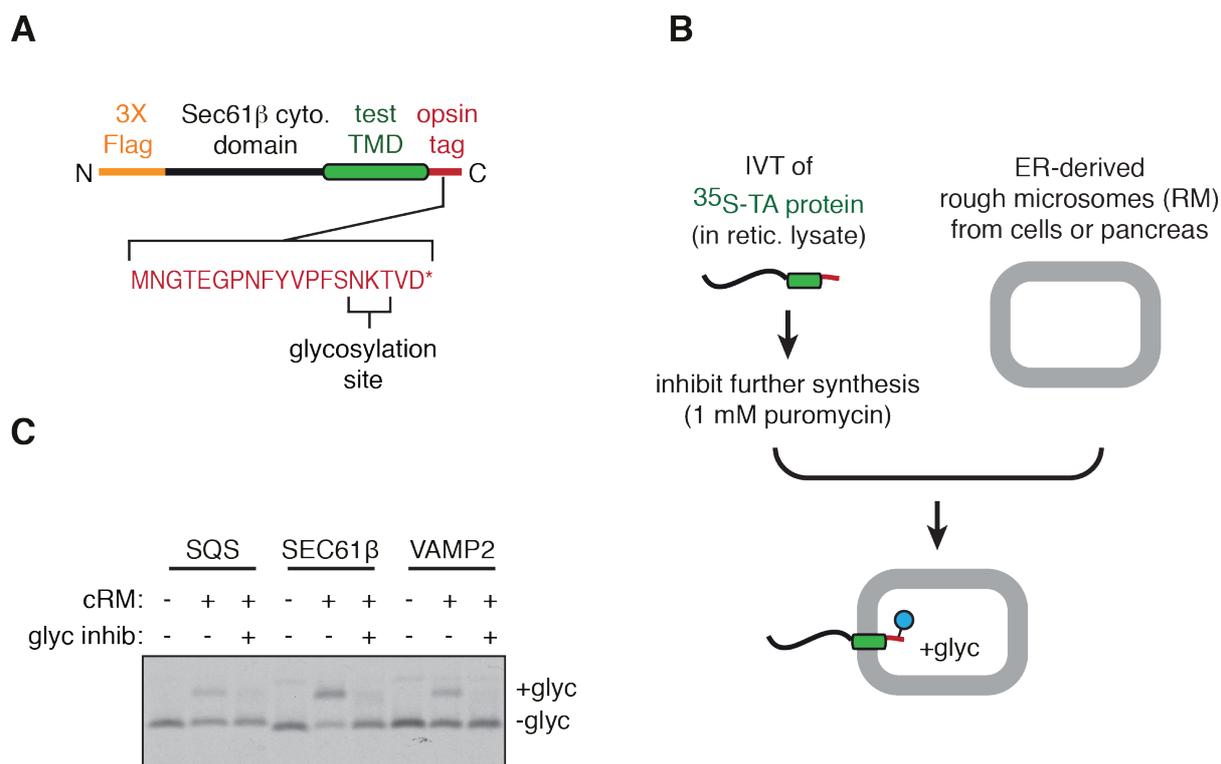
## Chapter 3: TRC-independent post-translational TA protein insertion

no enriched membrane proteins with C-terminal tails of 50 residues or less, indicating that these proteins are not being co-translationally inserted (Jan et al., 2014). The two exceptions are PMP1 and PMP2, however these are likely targeted to the ER via their mRNA (Chartron et al., 2016). Finally, *in vitro* experiments show that 50 residues are clearly tolerated in post-translational translocation experiments (Stefanovic and Hegde, 2007). The TMD for each TA protein was defined as per UniProt annotation. This was adjusted in some instances where the TMD sequence was known to be incorrect based on conflicting existing experimental results. This analysis yielded 235 ER destined and 44 mitochondrial destined TA proteins, which is in the range of previous studies (Kalbfleisch et al., 2007; Shigemitsu et al., 2016).

The TMD hydrophobicity for the full set of ER-destined TA proteins could then be determined (Figure 3.1A). This was accomplished using the transmembrane tendency values for individual amino acids as defined by Zhao and London (Zhao and London, 2006) as well as Kyte-Doolittle (Kyte and Doolittle, 1982), then summing all the values for the residues in the TMD to arrive at a net hydrophobicity score. The delta-G values were also determined (Hessa et al., 2007). It should be noted that although a wide range of scales for determining hydrophobicity exist, there is a strong correlation between all of them (Zhao and London, 2009) and the relative ranking of hydrophobicity between different TA proteins is largely unaffected regardless of the method used. What is clear is that TA TMDs span a wide spectrum of hydrophobicity. However, they tend to be more strongly hydrophobic compared to other TMDs found in multi-pass IMPs (Figure 3.1A). Directly comparing the hydrophilicity plots (computed using the Kyte-Doolittle scale) of a verified TRC pathway substrate such as the Vesicle-associated membrane protein 2 (VAMP2) with the significantly less hydrophobic ER-resident enzyme squalene synthase (SQS) dramatically highlights the range of TMDs the cell must insert post-translationally (Figure 3.1B).

### 3.2 Experimental assay for post-translational TA protein insertion

To directly test whether or how the TRC pathway manages to insert the gamut of endogenous TA proteins, a post-translational *in vitro* glycosylation-based assay was developed (Figure 3.2). The native TA protein Sec61 $\beta$  was modified to generate a generic protein cassette for TA TMDs (Figure 3.2A). For this, a 3X-FLAG tag was appended to the N-terminus while the TMD and C-terminus of Sec61 $\beta$  were removed. This permitted the insertion of any TMD sequence of interest, including 5 flanking residues on each side. A C-terminal opsin tag was also



**3.2 Glycosylation based assay for TA protein insertion.** (A) Diagram of the protein cassette used for most analyses of TA protein insertion. (B) Basic scheme of the *in vitro* post-translational insertion assay, using glycosylation of the C-terminal opsin tag as the readout. In all insertion assays, the TA protein is <sup>35</sup>S-methionine labelled, permitting its detection by autoradiography such that the inserted (+glyc) product can be distinguished from the non-inserted (-glyc) (C) Example of an insertion assay for SQS, Sec61β, and VAMP2 as described in (B). Reactions are performed in the presence of a glycosylation inhibitor where indicated.

appended. The opsin tag has a consensus glycosylation sequence that gets modified in the ER, serving as a simple and highly reliable indicator of insertion (Brambillasca et al., 2005; Buentzel and Thoms, 2017).

Using this cassette, a straightforward scheme was developed to assay insertion of TA TMDs into the ER membrane (Figure 3.2B). First, the TA protein was translated in rabbit reticulocyte lysate (RRL) (Pelham and Jackson, 1976) in the presence of <sup>35</sup>S-methionine, permitting labelling and subsequent detection by autoradiography. To remove the possibility of co-translational delivery, puromycin was added following translation but prior to the addition of membranes. Puromycin causes premature chain termination, ensuring all substrates disengage from the ribosome and thereby inhibiting further translation (Azzam and Algranati, 1973). Canine pancreas derived ER rough microsomes (cRMs) were used as the source of ER. When incubated with cRMs, inserted TA products are indicated by a heavier species due to the addition of glycans to the opsin tag (indicated by '+glyc'). This was verified by the specific

## Chapter 3: TRC-independent post-translational TA protein insertion

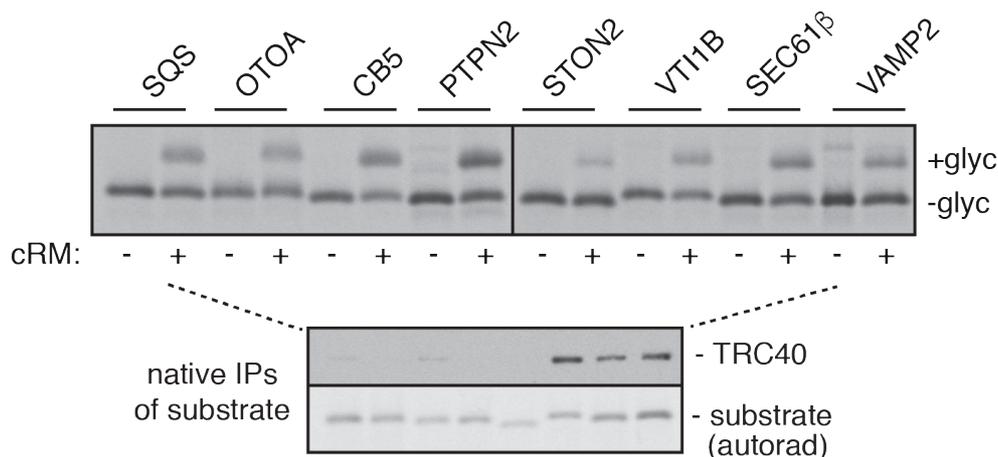
diminishment of this higher molecular weight band if a peptide inhibitor of glycosylation was included in the reaction (Figure 3.2C). The comparison of non-inserted (-glyc) to inserted product could also be compared to calculate how much of the substrate had reached the membrane. In the absence of cRMs, a product slightly larger than the glycosylated product can occasionally be observed. This represents a ubiquitinated product that is generated when insertion does not occur, as verified by de-ubiquitinase treatment and ubiquitin pull-downs (data not shown). TA proteins of low (SQS), moderate (SEC61 $\beta$ ) and high (VAMP2) TMD hydrophobicity were tested to confirm that the assay can be used to examine post-translational insertion (Figure 3.2C).

### 3.3 Less hydrophobic TA proteins use a TRC-independent insertion pathway

Bioinformatics analyses allowed the sampling of endogenous substrates to select eight TA proteins that represent a broad range of TMD hydrophobicity (Figure 3.3). An additional check was done to ensure that each had been experimentally verified, or at least computationally confirmed, as a bona fide TA protein. All eight proteins inserted relatively equally with reasonably high efficiency in the *in vitro* assay (Figure 3.4). This meant that in each case the substrate was being accurately recognized in the cytosol, appropriately delivered to membranes, and successfully inserted into the lipid bilayer by membrane resident machinery.

	TA TMD	relative hydrophobicity
...FYHSTGGS		MNGTEGPN...
SQS	<u>SRSHYSPIYLSFVMLLAALSWOYLTTLSQVTE</u> D	12.5
OTOA	TRTSSSRSPAGALQSWGLWLGCP <u>LLVLM</u> AKLLW	13.5
CB5	DSSSSWWTNWWIPAI <u>SAVAVALMYRL</u> YMAED	15.4
PTPN2	<u>WLYWQPILTKMGFMSVILVGAFV</u> GWRLFFQONAL	16.5
STON2	ALGSIWLMMLPTPFVHPTTL <u>PLLFL</u> LAMLTMFAW	14.2
VT1B	RKVTTNK <u>LLLSII</u> ILLELAILGGLVYYKFFRSH	22.0
SEC61 $\beta$	SPGLKVGVPVVLVMSLLFIASVFMLHIWGKYT	21.6
VAMP2	KTGKNL <u>KMMIIL</u> GVICAILIIIVYFTGSR	27.3

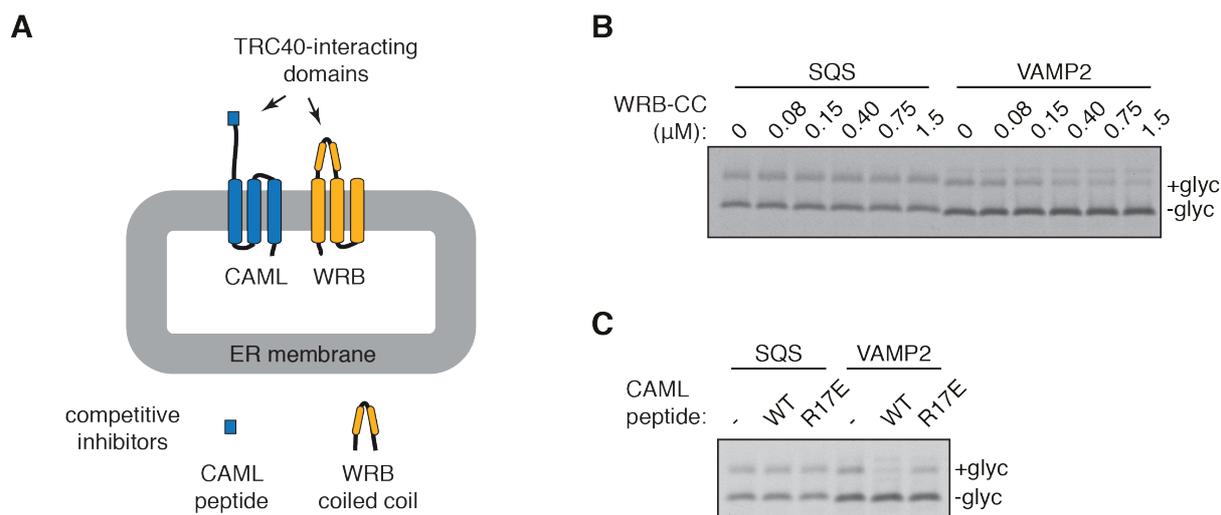
**3.3 Endogenous TA protein sequences.** Sequences of the TMD regions and the immediate flanking residues common to all constructs. These sequences were inserted into the TA protein cassette described in Figure 3.2A. To the right of each sequence is the TMD tendency score of Zhao and London (Zhao and London, 2006). The region defined as the TMD is underlined and was determined using the TMHMM algorithm (Krogh et al., 2001).



**3.4 Insertion and TRC40 interaction of endogenous TA proteins.**  $^{35}\text{S}$ -methionine labelled TA proteins (Figure 3.3) were assayed for post-translational insertion as indicated by glycosylation into cRM (Figure 3.2B). Parallel reactions lacking microsomes for each protein were affinity purified via the substrate's FLAG tag and analysed for TRC40 association (by immunoblot) and substrate (by autoradiography). Identical results were obtained in native RRL and nucleased RRL (shown in figure).

To probe whether TRC40 engages these substrates in the cytosol, parallel reactions lacking microsomes were affinity purified via the FLAG tag. These were analysed for TRC40 association by immunoblotting and verified for equal recovery of substrate by autoradiography. In both native and nucleased RRL (Figure 3.4), only the three most hydrophobic TMDs interacted efficiently with TRC40. The remaining five showed minimal recovery of TRC40. This raised the possibility that substrates with low-to-moderate TMD hydrophobicity don't interact with TRC40 for delivery to the membrane. However, it is also possible that lack of TRC40 recovery could be a result of manipulations during the immunoprecipitation process that might perturb interactions with these less hydrophobic substrates. To more rigorously determine whether a TRC-independent insertion route exists for this select subset of proteins, we sought a means of specifically inhibiting this pathway at the membrane.

It is known that both TRC40 recruitment and substrate release are mediated by the cytosolic fragments of the WRB/CAML complex (Get1/Get2 in yeast) (Mariappan et al., 2011; Wang et al., 2014) (Figure 3.5A). Initial recruitment of TRC40 is via an interaction with a short two-helix motif on the N-terminal tail of CAML/Get2. Substrate release is then triggered by WRB/Get1's cytosolic coiled-coil domain (WRB-CC). The WRB-CC wedges between the two

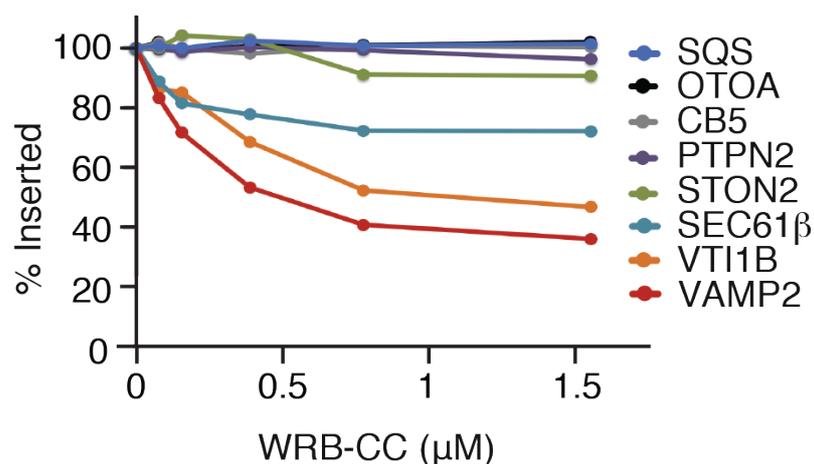


**3.5 SQS and VAMP2 insertion in the presence of TRC pathway inhibitors:** (A) Cartoon diagram of the TRC40 receptor consisting of WRB (yeast Get1 homolog) and CAML (yeast Get2 homolog). The TRC40 interacting domains of each are indicated. These fragments can act to competitively inhibit TRC40 interaction, serving as selective probes of the TRC pathway. (B) Insertion assay for SQS and VAMP2 in the presence of increasing concentrations of the WRB coiled-coil (WRB-CC). (C) Insertion of SQS and VAMP in the absence or presence of the CAML fragment. A point mutation in the CAML peptide (R17E) which prevents interaction is used to demonstrate TRC pathway inhibition specificity.

alpha-helical domains of the Get3 dimer, disrupting the hydrophobic substrate binding pocket. The interaction sites on TRC40 of WRB and CAML partially overlap. Therefore, these fragments can be used to selectively probe the TRC pathway by inhibiting interaction of TRC40 with its receptor.

To assess whether the insertion of certain TA proteins is TRC-independent, post-translational insertion was assayed in the presence of the WRB or CAML cytosolic fragment, which behave like competitive inhibitors. Purified WRB-CC was added in the *in vitro* post-translational insertion assay at increasing concentrations. As expected, a known TRC substrate (VAMP2), showed dose-dependent inhibition (Figure 3.5B). However, the insertion of the low hydrophobicity TA protein SQS was completely unaffected by the presence of WRB-CC even at high concentrations. To confirm that these results reflected selective inhibition of the TRC40 pathway, insertion was also assayed in the presents of the cytosolic CAML peptide (Figure 3.5C). The differential effects on VAMP2 versus SQS insertion remained. A point mutation (R17E) in the CAML peptide that prevents TRC40 interaction is not inhibitory in this assay for either substrate.

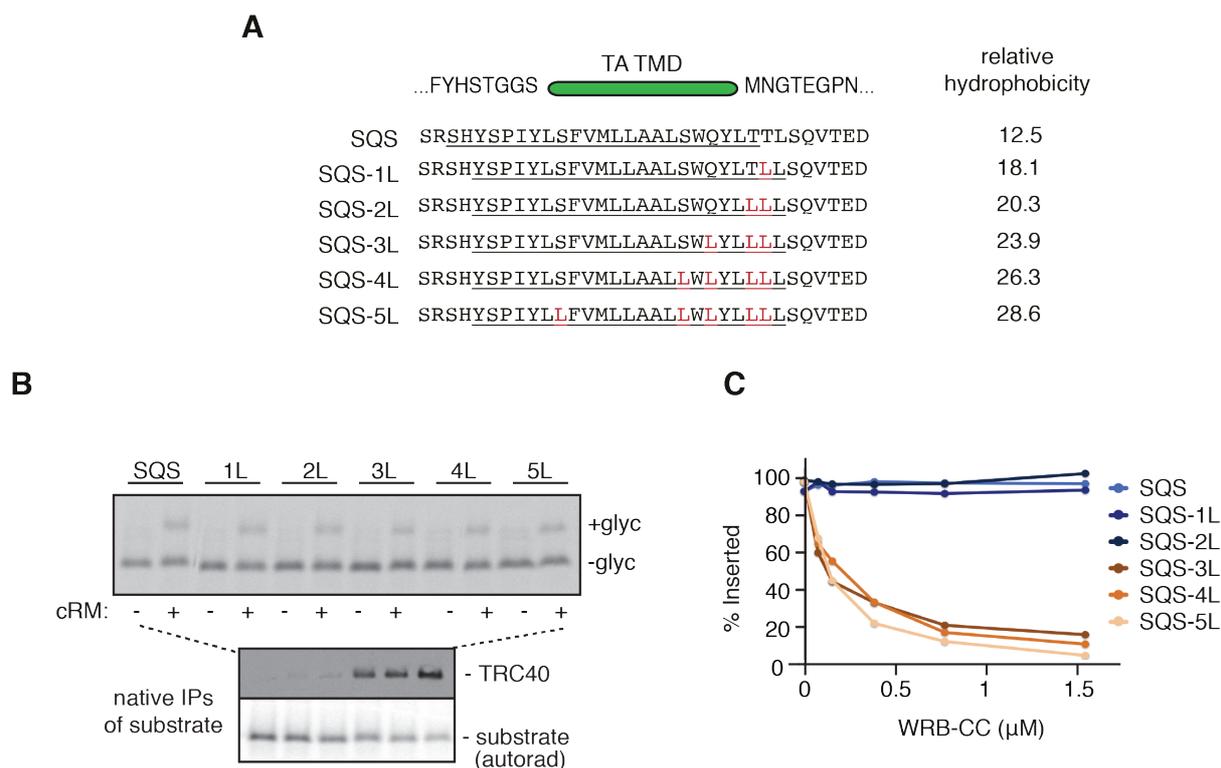
The remaining six selected TA proteins were subsequently assayed for insertion in the presence of WRB-CC. Only the three most hydrophobic substrates which efficiently engaged TRC40 showed a dose-dependent reduction of insertion (Figure 3.6). Those which showed



**3.6 TA insertion sensitivity to TRC pathway inhibitors.** Relative normalized insertion efficiencies for the indicated TA proteins (Figure 3.3) with increasing amounts of the WRB-coiled coil (WRB-CC) (see Figure 3.5).

minimal TRC40 recovery in the immunoprecipitation experiment were completely resistant to inhibition, suggesting that they reach the membrane independently of the TRC pathway. Of the three TA proteins sensitive to WRB-CC inhibition, SEC61β showed a slightly tempered effect. This was also reflected in the fact that SEC61β associates with TRC40 slightly less than VAMP2 and VTI1B based on the immunoprecipitation results (Figure 3.4). Interestingly, the hydrophobicity of the SEC61β TMD lies roughly at the midpoint of all the TA TMDs (Figure 3.1A). Therefore, this likely represents the relative TMD hydrophobicity threshold dictating TRC-dependence versus independence. The partial effect also implies that a small number of proteins of moderate hydrophobicity can use both the TRC pathway and a TRC-independent route to the ER membrane. The interrogation of these eight TA proteins suggests that the TRC pathway only handles relatively hydrophobic proteins. Using the SEC61β TMD as the approximate threshold for TRC-dependence, it is estimated that around half of all TA proteins are inserted via a non TRC pathway.

We wanted to explicitly test that TMD hydrophobicity is the key parameter dictating TRC-dependence versus independence. To do this, the hydrophobicity of a low hydrophobicity TA protein, SQS, was systematically increased through a series of leucine mutations (Figure 3.7A). SQS was shown to be insensitive to inhibition by the WRB-CC/CAML peptide fragment and minimally interacts with TRC40 in the cytosol. These SQS mutants were subjected to the same analysis as endogenous TA proteins: testing ability to engage TRC40 in the cytosol and post-translation insertion in the presence of TRC pathway inhibitors (Figure 3.7B). At a hydrophobicity threshold similar to that of the SEC61β TMD, SQS mutants became sensitive



**3.7 TRC-dependence is a function of TMD hydrophobicity.** (A) The SQS TMD was serially mutated (red residues) to increase its hydrophobicity. These sequences were inserted into the standard TA cassette described in Figure 3.2A. The transmembrane tendency score for each SQS mutant was determined using the method of Zhao and London (Zhao and London, 2006). The TMD is underlined. (B) SQS mutants were tested for insertion and TRC40 association as in Figure 3.4. Identical results were obtained in native RRL and nucleated RRL (shown in figure). (C) SQS mutants subjected to insertion assays in the presence of WRB-CC as in Figure 3.6.

to TRC40 pathway inhibition (Figure 3.7C). This switch from resistance to sensitivity correlated with TRC40 interaction. These observations corroborate those derived from examining endogenous substrates.

### 3.4 TRC40 cannot mediate the insertion of low hydrophobicity TA proteins before dissociation

Analysis in RRL of endogenous TA proteins and SQS leucine mutants indicates that TRC40 minimally associates with low hydrophobicity TMDs, and that these substrates instead rely on alternative means of being targeted and inserted into the ER membrane. Though these results suggest that the TRC pathway does not appear to be used by some TA proteins, they do not confirm that it is strictly incapable of doing so. It is also unclear at which step in the TRC pathway the limitation lies for TA proteins like SQS. We therefore sought to establish where

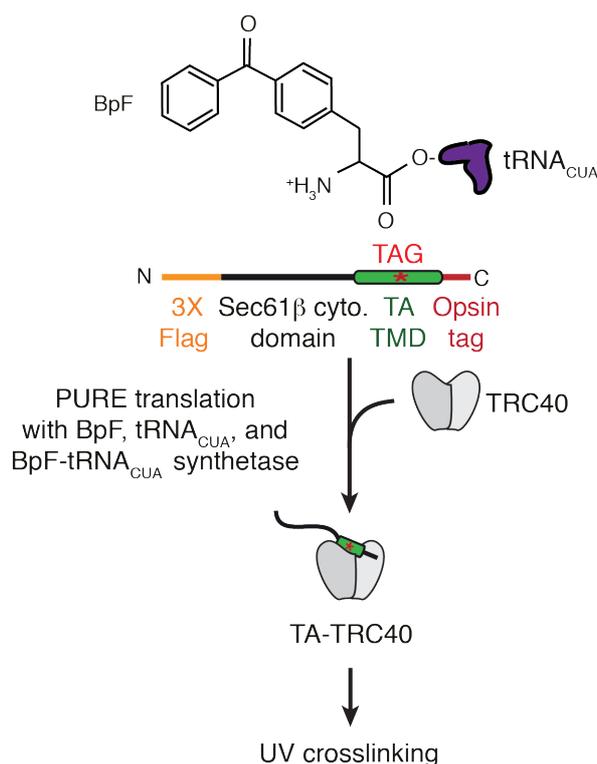
### Chapter 3: TRC-independent post-translational TA protein insertion

the TRC pathway diverges for low compared to high hydrophobicity TMD containing TA proteins.

Effective delivery of TA proteins to the ER membrane by the TRC pathway is a product of a series of successful interactions with pathway components. Primarily, TRC40 can only efficiently capture substrates when it is in complex with other factors. These factors mediate TMD transfer from the chaperone SGTA (Sgt2 in yeast) to TRC40 (Gristick et al., 2014; Mateja et al., 2015; Mock et al., 2015; Wang et al., 2010). One possibility for the differential behaviour of low hydrophobicity substrates is that there is a deficiency in transfer or loading. Once transfer onto TRC40 has occurred, the substrate needs to remain bound until it reaches the WRB/CAML complex. Weakly bound substrates are more likely to fall off as they are being transited through the hydrophilic cytosol. Once at the membrane, substrates have to be released from TRC40 and transfer to the WRB/CAML complex. Therefore, lack of insertion could represent either ineffective release from TRC40 or the inability to interact appropriately with the WRB/CAML complex.

To mechanistically dissect the TRC pathway at higher resolution, we compared TA proteins VAMP2 and SQS in the PURE system (protein expression using recombinant elements), a reconstituted translation system with purified factors (Shimizu and Ueda, 2010). VAMP2 was chosen since it is clearly a TRC-dependent substrate, while SQS seems to not depend at all on this pathway. The PURE system allows proteins to be translated in the presence of specific chaperones, leading to the formation of homogenous TA-chaperone complexes that can be used for downstream functional assays (Mateja et al., 2015). Certain complexes, such as SQS-TRC40 would be extremely difficult to isolate in an RRL based system where the interaction between the two proteins was shown to be extremely feeble. However, generating these complexes in the PURE system is feasible and straightforward (Shao et al., 2017). As in RRL, PURE reactions allow for the radioactive labelling of synthesized proteins by translating in the presence of <sup>35</sup>S-methionine. The protein cassette used in these assays encoded the same protein coding sequence as in the RRL system (Figure 3.8). The only exception was that it was moved to a PURE compatible T7-polymerase based transcription vector. Experiments in the PURE system therefore allowed us to interrogate specific steps of the TRC pathway. Specifically, substrate transfer from SGTA onto TRC40 and the subsequent ability of TRC40 to remain bound to TMDs.

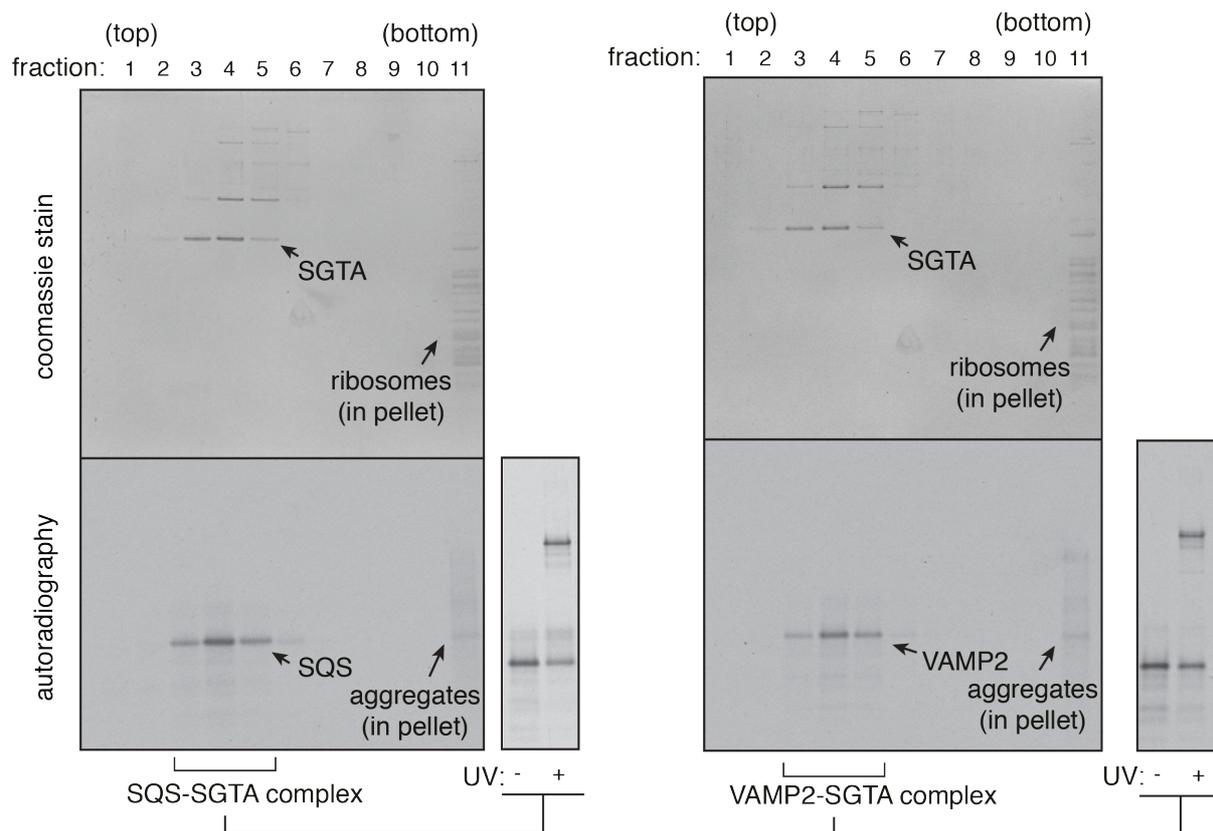
In order to monitor substrate interactions, a site-specific photo-crosslinking strategy was used (Fig 3.8). For this, the non-natural UV-reactive amino acid benzoyl-phenylalanine (BpF) was incorporated into the middle of the TMD in the standard FLAG-TA TMD-opsin



**3.8 Schematic of site-specific crosslinking.** Construct and strategy used to incorporate the non-natural UV-reactive amino acid benzoyl-phenylalanine (BpF) into the middle of a TA TMD. This is achieved by amber codon (TAG) suppression in a homemade PURE system which lacks RF1 (which normally recognizes amber codons) and instead contains the amber-suppressor  $tRNA_{CUA}$ , the BpF- $tRNA_{CUA}$  synthetase, and BpF. Translation in the PURE system in the presence of a TMD binding chaperone, such as TRC40, produces TA-TRC40 complexes. Substrate interactions can be probed via site-specific photo-crosslinking by exposing the BpF containing TMD to UV light.

cassette. This is achieved by amber codon (TAG) suppression in a homemade PURE system. This system lacks RF1, which usually recognized amber codons, and instead contains the amber-suppressor  $tRNA_{CUA}$ , the BpF- $tRNA_{CUA}$  synthetase, and BpF. Interactions can then be examined by exposing the BpF-containing substrate to UV light. This reaction also works with high efficiency on frozen samples, allowing analysis at specific time points (Shao et al., 2017). TA proteins were generally translated in the presence of a specific chaperone, allowing the formation of a defined TA-chaperone complex that can be used for subsequent assays.

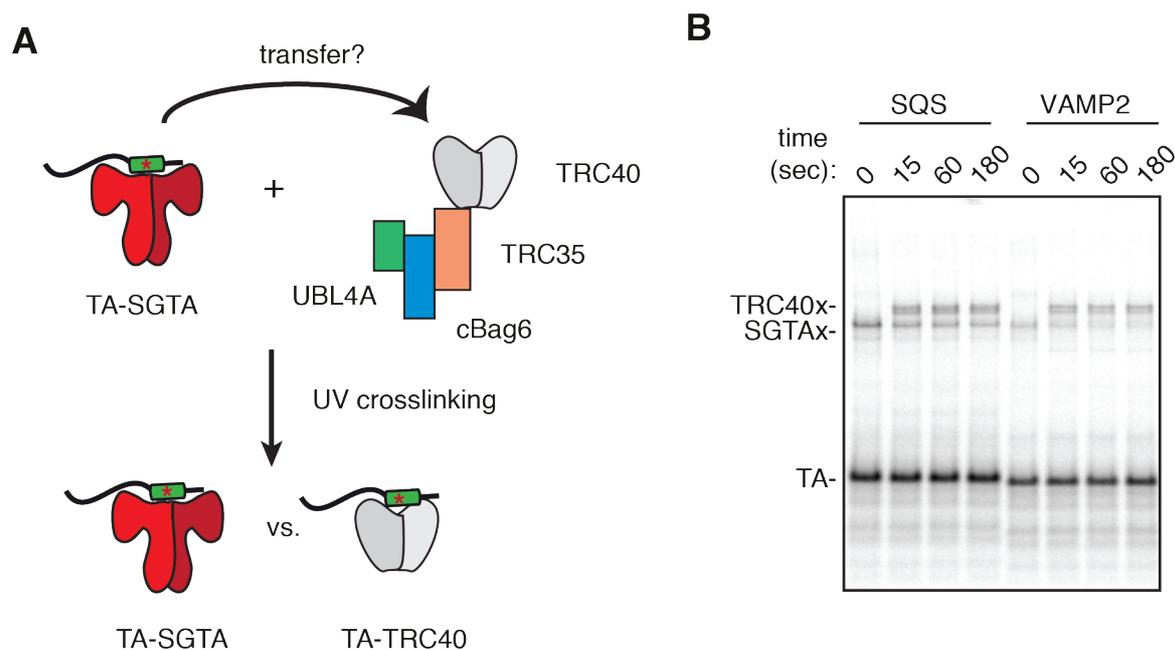
To test whether there are differences in SGTA to TRC40 transfer, TA-SGTA complexes were first assembled in the PURE system (Figure 3.9). Here, either SQS or VAMP were synthesized in the amber-suppressor competent homemade system supplemented with 12  $\mu$ M human recombinant SGTA. The reaction was size separated on a 5-25% sucrose gradient, collected as 11 fractions, and analysed by SDS-PAGE and Coomassie staining or autoradiography. SQS and VAMP both co-fractionate with SGTA in fractions 3-5. Thus, the



**3.9 Preparation of TA-SGTA complexes in the PURE system.** SQS or VAMP2 were translated in an amber-suppressor competent PURE system in the presence of 12  $\mu$ M recombinant human SGTA. Analysis of products on a sucrose gradient shows that SQS co-fractionates with SGTA in fractions 2-4, which are pooled for use in subsequent assays. UV irradiation of the pooled peak fractions verifies that SQS is in complex with SGTA.

soluble fractions represent TA-SGTA complexes, and these were pooled for use in subsequent assays. UV irradiation of the pooled fractions verifies that the substrate contains the photocrosslinking residue and is bound to SGTA.

The isolated TA-SGTA was then incubated with the Bag6 targeting complex in the presence of TRC40 (Figure 3.10A) (Shao et al., 2017). The reaction was supplemented with ATP. The Bag6 targeting complex is essential for TA transfer from SGTA to TRC40 and is composed of Bag6 and the two bridging proteins UBL4A and TRC35 (Mock, 2015; Xu, 2012). Only the 110 C-terminal residues of Bag6 were used (termed cBag6) as they are sufficient to recapitulate the transfer function of the full Bag6 (Shao et al., 2017). The N-terminal of Bag6 is dispensable for TRC40 capture but is required for TA protein ubiquitination, as this domain is responsible for recruiting the E3 ubiquitin ligase RNF126 (Hessa et al., 2011; Rodrigo-Brenni et al., 2014; Shao et al., 2017). Following incubation of the reaction, subsamples were taken at selected time points, flash frozen, and subjected to UV irradiation (Figure 3.10A). After analysis by SDS-PAGE and autoradiography, the amount of TA substrate engaged with



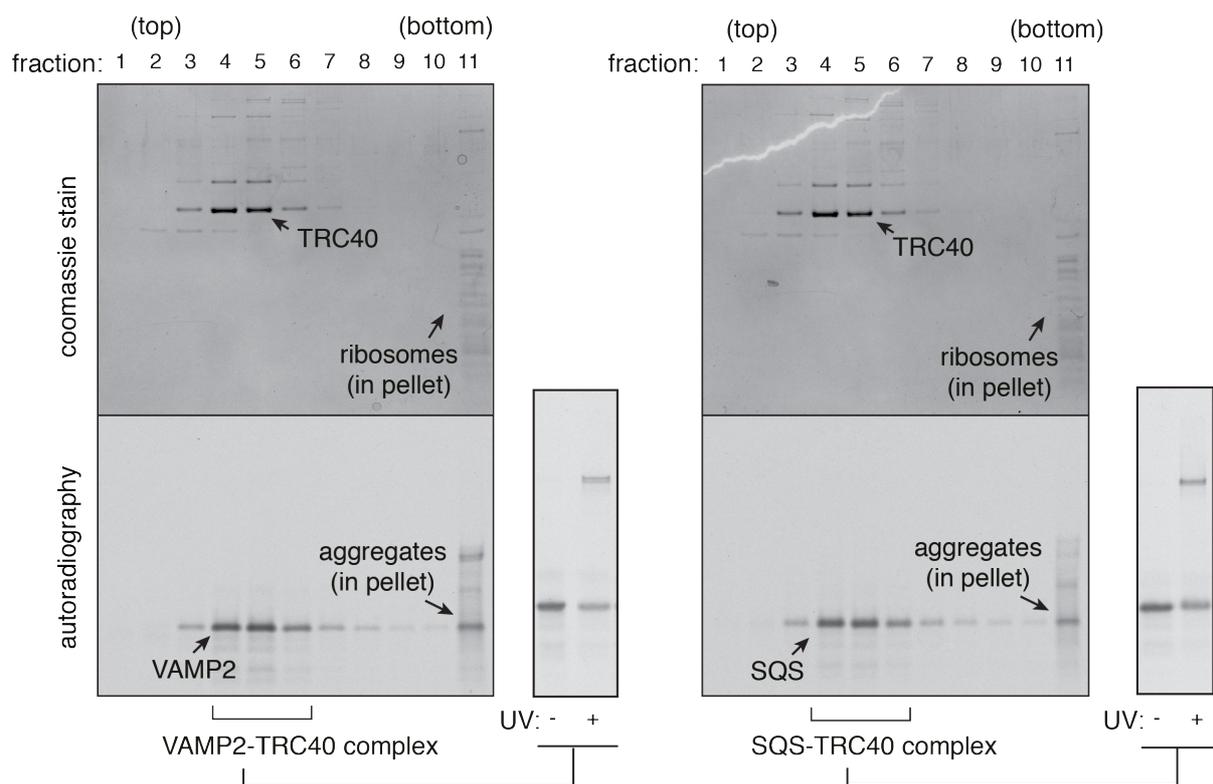
**3.10 Ability of TA substrates to transfer onto TRC40.** (A) Isolated TA-SGTA complexes (Figure 3.9) were incubated with a modified Bag6 targeting complex competent for transfer (composed of TRC35, UBL4A, and cBag6) in the presence of TRC40. Subsamples were removed and subjected to photo-crosslinking by exposure to UV light to determine substrate fate. (B) SQS and VAMP2 transfer onto TRC40 was assayed at the indicated time points. Data generated by Manu Hegde.

either SGTA or TRC40 was quantified. Transfer from SGTA to TRC40 is clearly less effective for SQS than for VAMP (Figure 3.10B). Indeed, SQS seems to transfer incompletely, with a significant portion remaining on SGTA. One interpretation of this result is that the private transfer of SQS is less efficient than that of VAMP2. However, this observation could instead reflect the inability of SQS to remain on TRC40 (i.e. faster off rate), thereby causing it to equilibrate between the two chaperones.

To distinguish between these two possibilities, off-rates of SQS and VAMP2 from TRC40 were compared directly. TA-TRC40 complexes were generated in the PURE system in a similar manner to TA-SGTA complexes (Figure 3.11). Either substrate was translated in the presence of 12  $\mu$ M recombinant zebrafish TRC40 and this reaction was fractionated on a sucrose gradient. Both substrates co-fractionate with TRC40 in fractions 3-6, which were then pooled. A subsample was subjected to UV irradiation to confirm that the isolated complexes represented SQS-TRC40 or VAMP2-TRC40 (Figure 3.11).

To measure the amount of TA substrate that remains on TRC40, the preassembled complex was incubated with 40-fold excess calmodulin (CaM) (Figure 3.12A). CaM is a general TMD binding factor with broad specificity (O'Neil and DeGrado, 1990). In the

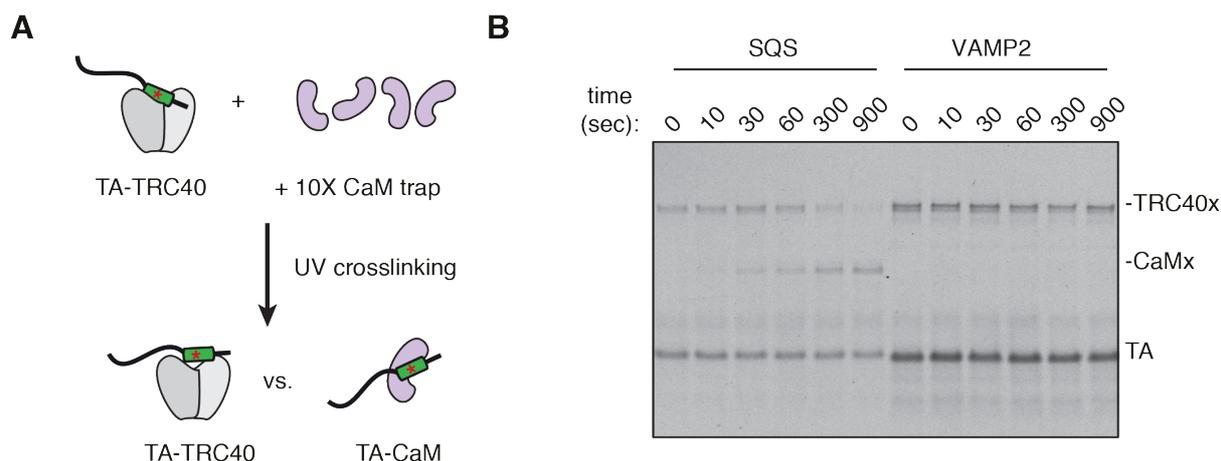
## Chapter 3: TRC-independent post-translational TA protein insertion



**3.11 Preparation of TA-TRC40 complexes in the PURE system.** SQS or VAMP2 were translated as in the presence of 12  $\mu\text{M}$  recombinant zebrafish TRC40 as in Figure 3.9. SQS co-fractionates with TRC40 in fractions 2-4, which are pooled for use in subsequent assays. Association of SQS with TRC40 is verified by UV irradiation.

presence of excess calcium, CaM binds tightly to TMDs (Shao and Hegde, 2011b). Therefore,  $\text{Ca}^{2+}$ -CaM can effectively serve as a trap for exposed TMDs. This property of  $\text{Ca}^{2+}$ -CaM was exploited to trap TMDs were they to dissociate from TRC40. While VAMP2 remained tightly bound to TRC40, SQS rapidly dissociated and was captured by the CaM trap (Figure 3.12B). This is likely due to the inability of TRC40's hydrophobic groove to stably interact with the relatively low hydrophobic TMD of SQS. Based on this observation, one would predict that even if SQS managed to associate with TRC40 in the cytosol, the rapid dissociation of the complex would prevent appreciable insertion via WRB/CAML.

This was explicitly tested in an experiment designed to examine TRC40-mediated targeting and insertion (Figure 3.13A). The isolated TA-TRC40 complexes prepared in the PURE system were presented to cRMs in the presence of a 10-fold excess of the  $\text{Ca}^{2+}$ -CaM trap. In this instance, near irreversible TMD binding with CaM would prevent additional insertion events. Thus, if the TA-TRC40 complex is delivered to its receptor and inserted, the CaM is inert and never sees the substrate. By contrast, dissociation of the TA protein from

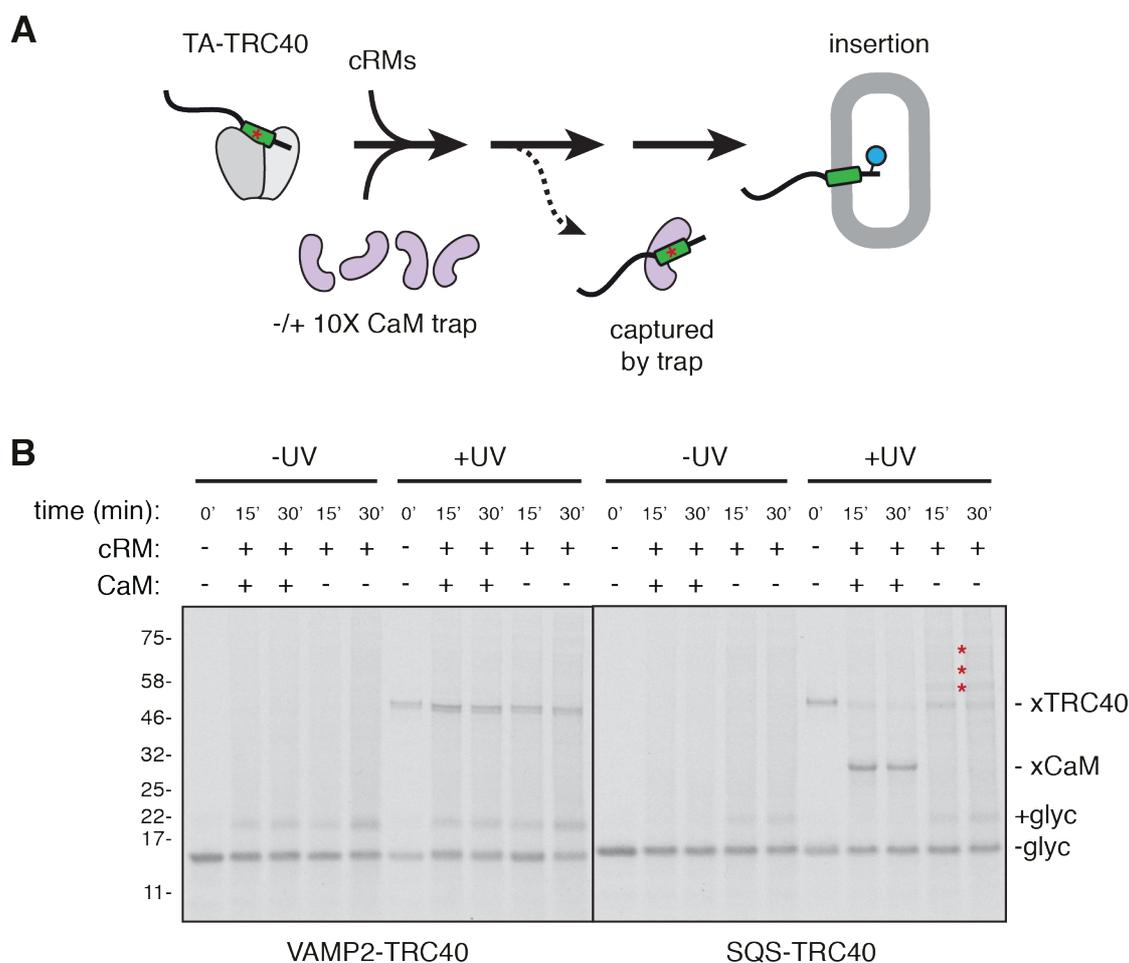


**3.12 Stability of TA-TRC40 interaction.** (A) Scheme used to test the ability of TRC40 to remain associated to TA TMDs. TA-TRC40 complexes (Figure 3.11) were incubated with 10-fold excess purified mammalian calmodulin (CaM). CaM at this concentration is capable of sequestering TMDs should they dissociate from TRC40. Substrate association is determined by exposure to UV light. (B) SQS and VAMP2 association with TRC40 was assayed with subsamples analysed at the indicated time points.

TRC40 before membrane delivery results in capture by the CaM trap. The interactions between the TA protein and TRC40 or CaM were monitored by UV irradiation at specific time points. Successful insertion was indicated by TA glycosylation. VAMP2 inserts comparably well regardless of the CaM trap, and no crosslinking to CaM is observed (Figure 3.13B). Thus, the fate of VAMP2 is either to remain on TRC40, consistent with its very slow off-rate (Shao et al., 2017), or to be inserted into the membrane via the WRB/CAML complex. By contrast, the SQS-TRC40 complex completely fails insertion in the presence of the CaM trap, and SQS is observed to release from TRC40 and bind to CaM. This confirms that SQS dissociates before successful insertion and is consistent with the increased off-rate from TRC40. Of note, in the absence of the CaM trap, SQS does insert into cRMs. This indicates that upon dissociation from TRC40, the released SQS has some alternative route to the membrane. Here, SQS seems to make a heterogeneous set of non-specific interactions, likely due to the absence of any suitable chaperone for this TMD.

### 3.5 Discussion

These data cumulatively illustrate that alternative routes of insertion to the TRC pathway must exist for a significant number of TA proteins which have TMDs of low-to-moderate hydrophobicity. This is not an entirely novel proposal, as it was known very early on that not every TA protein can effectively engage TRC40 (Stefanovic and Hegde, 2007). This, in



**3.13 TRC40 cannot mediate SQS insertion before dissociation.** (A) Schematic of the experiment used to test TRC40-mediated targeting and insertion. (B) SQS and VAMP2 were assayed as in (A). After 15 min or 30 min samples were either directly analysed or first subjected to UV irradiation to track substrate interactions.

addition to the fact that depletion of TRC/GET pathway components have differential effects on TA proteins in both mammalian systems (Casson et al., 2017; Rivera-Monroy et al., 2016) and yeast (Jonikas et al., 2009; Schuldiner et al., 2008) has led to postulation that other ‘alternative’ pathways must exist. However, it remained vague as to which TA proteins remained unaffected by perturbations to the TRC pathway and why. We address both of these issues by combining our bioinformatics analyses with *in vitro* approaches.

An interrogation of the human proteome showed that there are roughly ~235 ER-destined TA proteins using our current bioinformatics approach. This number is reasonable in light of other estimates (Kalbfleisch et al., 2007; Shigemitsu et al., 2016). The main problem with using computational approaches to identify TA proteins is the difficulty in parsing between TA versus signal anchor protein when using a particular C-terminal sequence length as a strict cut-off. It is likely that the cell shows a certain amount of leniency and can accommodate slight deviations in length. Therefore, the best way to definitively establish the

### Chapter 3: TRC-independent post-translational TA protein insertion

full complement of TA proteins would be with experimental validation, which would resolve any ambiguity about particular substrates. After TA proteins are distinguished from signal anchored proteins, there remains the complication of differentiating between those destined for the ER membrane or mitochondria. TA proteins targeted to the outer mitochondrial membrane (OMM) tend to have shorter and less hydrophobic TAs which also exhibit lower helical propensity (Beilharz et al., 2003; Lee et al., 2014; Rao et al., 2016). Some mitochondrial TA proteins are also enriched in basic residues immediately following the TMD, a feature that has been experimentally shown to affect targeting specificity (Horie et al., 2002; Kuroda et al., 1998; Marty et al., 2014; Rao et al., 2016). However, there remains significant overlap between the two groups of TA proteins, with the localization of numerous substrates not yet definitively established. Of note, the mechanism by which TA proteins are inserted into the OMM remains unknown. Both targeting specificity and the mode by which mitochondrial TA proteins reach and are inserted remains an important area of future work.

Despite these shortcomings, our computational method is sufficient to illustrate that ER-destined TA proteins are not uniform in their properties. An essential characteristic that varies widely amongst them is TMD hydrophobicity. TMD hydrophobicity ultimately determines whether a protein gets targeted to the membrane and if the stretch of amino acids can partition from an aqueous to lipid environment (Cymer et al., 2015; Shao and Hegde, 2011a). It remained unclear how TRC pathway can successfully execute its function in light of this diversity.

To experimentally address this, we developed an *in vitro* post-translational insertion assay. This assay was shown to reliably indicate ER insertion with glycosylation as a readout, which also permitted quantification of percent insertion. Though not novel, this assay proved very useful and with slight modifications was used in a variety of other experimental systems. Using this approach, the importance of TMD hydrophobicity was tested directly by examining eight endogenous TA substrates in a standard cassette. Certain TMDs minimally interacted with TRC40 in RRL. This, along with the resistance of their insertion to TRC pathway inhibitors cumulatively showed that they do not depend on this pathway. These TMDs were of particularly low-to-moderate hydrophobicity, hinting that this is the key limitation of the TRC pathway. To test this more rigorously, the hydrophobicity of a low hydrophobicity TA substrate (SQS) was systematically increased until it resembled that of an endogenous substrate with high hydrophobicity (VAMP2). At a similar threshold as that for endogenous substrates, the SQS mutants flipped from being TRC-independent to dependent. Mapping this approximate threshold on our bioinformatics analysis shows that roughly 50% of all TA proteins (~100-

## Chapter 3: TRC-independent post-translational TA protein insertion

110) do not use the well-established TRC pathway for delivery to the membrane. Some TA proteins may be capable of simultaneously using multiple routes of to reach and be inserted into the ER. Analysis in the PURE system showed that observations in RRL were due to TRC40 interaction being very weak with low hydrophobicity TMDs, thereby preventing these substrates from reaching the membrane via this targeting factor. Therefore, what we have showed is that the inherent limitation of TRC40 is its inability to tightly bind TMDs unless they are of high hydrophobicity, rendering it unable to cater to all TA proteins. We can now list the TA proteins which cannot use the TRC pathway and explain which substrate properties determinants of this are. This is a significant contribution which can explain previous *in vitro* and *in vivo* observations.

Naturally, the outstanding question becomes how these TRC-independent substrates get inserted into the ER membrane. Presumably, similar modules as for the TRC pathway must exist in that certain factors need to associate with the TA protein in the cytosol to prevent its aggregation, the substrate must reach the ER membrane, and some ER-resident machinery must mediate the insertion reaction. This other set of factors must be able to correctly recognize and cope with low hydrophobicity TMDs, a function which evidently eludes components of the TRC pathway. This putative ‘alternative’ pathway is experimentally dissected in the following three chapters.

### 3.6 Materials and methods

#### *Plasmids, proteins, antibodies and reagents*

Constructs for expression in rabbit reticulocyte lysate (RRL) were based on the SP64 vector (Promega). Constructs for translation in the PURE system were based on the T7-based PURExpress plasmid provided by New England Biolabs. Transmembrane domains of TA proteins were derived from our described bioinformatics analysis and included the following human proteins with their respective accession numbers. squalene synthase isoform 1 (SQS; NP\_004453.3); otoancorin isoform 2 (OTOA, NP\_733764.1); cytochrome b5 isoform 1 (CB5, NP\_683725.1); tyrosine-protein phosphatase non-receptor type 2 isoform 1 (PTPN2, NP\_002819.1); stonin-2 isoform 1 (STON2, NP\_149095.2); vesicle transport through interaction with t-SNAREs homolog 1B (VTI1B, NP\_006361.1); Sec61 $\beta$  (SEC61B, NP\_006799.1); vesicle associated membrane protein 2 (VAMP2, NP\_055047.1). The TA

## Chapter 3: TRC-independent post-translational TA protein insertion

protein expression construct, described in Figure 3.2A, contained a C-terminal opsin tag with a glycosylation acceptor site to monitor insertion, and an N-terminal 3X FLAG tag for affinity purification.

Constructs used for the purification of recombinant proteins included: His-tagged calmodulin in the pRSETA vector (Shao and Hegde, 2011b), GST-tagged SGTA in the pGEX-6p1 vector (Shao et al., 2017), His-tagged human WRB-coiled coiled (residues 35 to 101) in the pRSETA vector. Rabbit antibodies against TRC40 have been previously described (Stefanovic and Hegde, 2007). His-tagged calmodulin, GST-tagged SGTA, His-tagged WRB fragment, and His-tagged BpA- RS (for PURE system amber-suppression and ‘sense codon suppression’ in the RRL system) were expressed and purified from BL21(DE3) or BL21(DE3) pLysS *E. coli* cells as previously described (Shao et al., 2017). The CAML peptide (ASQRRRAELRRRKLMLNSEQRINRIMGGGWC) and matched R17E mutant were synthesized and purified by Designer BioScience Ltd.

### ***Mammalian in vitro expression***

Translation reactions in RRL and preparation of ER derived rough microsomes from canine pancreas (cRM) were as described in detail previously (Sharma et al., 2010; Walter and Blobel, 1983). Transcription reactions were identical to those described in Chapter 2. TA proteins were translated for 20 min at 32°C. For targeting reactions, 1 mM puromycin was added prior to cRMs and the reactions were incubated at 32°C for an additional 15 min. To assay for WRB-coiled coil or CAML peptide inhibition, purified protein or peptide was added at the time of microsome addition. In some reactions, a tripeptide competitive inhibitor of glycosylation (Asn-Tyr-Thr) was added at 50 µM to verify identity of the glycosylated product.

### ***PURE in vitro translation reactions and photo-crosslinking***

A modified amber-suppression competent homemade PURE translation system based on previously published methods (Shimizu and Ueda, 2010) was prepared as formerly described (Shao et al., 2017). This system replaces RF1 with total *E. coli* tRNA from a strain over-expressing the amber-suppressor tRNA, and contains 50 µg/ml of the purified amber-suppressor tRNA synthetase (BpA-RS) and 0.1 mM of the un-natural photo-crosslinking amino acid benzoyl-phenylalanine (BpA). Note that unlike in RRL, BpA-RS is orthogonal in *E. coli*, so BpA is only incorporated at amber codons. Translation reactions were carried out for 30 min

## Chapter 3: TRC-independent post-translational TA protein insertion

at 37 °C.

Purified SGTA or TRC40 were added to translation reactions at 12  $\mu\text{M}$ . To isolate the TA-TRC40 or TA-SGTA complexes, 5  $\mu\text{l}$  translation reactions were diluted to 20  $\mu\text{l}$  with ice cold physiologic salt buffer (PSB: 50 mM HEPES pH 7.4, 100 mM KOAc, 2 mM  $\text{MgAc}_2$ ) and size fractionated on a 200  $\mu\text{l}$  5-25% sucrose gradient in PSB (supplemented with 100 nM  $\text{CaCl}_2$  in the case of CaM complexes). Centrifugation was for 140 min at 4 °C in a TLS-55 rotor with the slowest acceleration and deceleration settings. Eleven 20  $\mu\text{l}$  fractions were collected from the top, and peak fractions containing TA-chaperone complex were pooled for downstream assays. The concentration of the chaperone in the final insertion assay was between 250 and 500 nM, with the radiolabelled substrate at sub-stoichiometric levels (at least 5-fold lower). Crosslinking analyses in the PURE system were done on ice ~10 cm away from a UVP B-100 series lamp (UVP LLC) for 15 min. After crosslinking, protein sample buffer was added directly for SDS-PAGE analysis.

### *Affinity purification*

TA protein affinity purification was performed with FLAG-TA proteins synthesized in the RRL system (Mariappan et al., 2010). Immediately following the translation reaction, the sample was chilled on ice and incubated with FLAG-M2 agarose beads (Sigma) for 1-2 hours at 4 °C. The resin was washed five times with PSB at 4 °C and the bound proteins were eluted with 0.2 mg/ml 3X FLAG peptide in 150 mM NaCl, 50 mM HEPES pH 7.6 by incubation with mixing for 30 min at 25 °C.

## Chapter 4: Cytosolic factors maintaining tail-anchored protein insertion competence

A significant portion of tail-anchored (TA) proteins are delivered to the endoplasmic reticulum (ER) membrane independently of the TRC pathway. These TA proteins are those whose transmembrane domain (TMD) is not sufficiently hydrophobic to permit effective association with the targeting factor TRC40. The mechanism for non-TRC pathway insertion remains unclear. However, these less hydrophobic substrates are still prone to aggregation in the cytosol after synthesis. Therefore, some type of chaperone must be responsible for keeping them in a soluble, insertion competent state before they reach the ER membrane. Previous *in vitro* studies have shown that certain TA proteins can physically interact with the signal recognition particle (SRP) and chaperones such as Hsc70/Hsp40 (Abell et al., 2004, 2007). However, the biological relevance of these interactions remains ambiguous, and these observations likely represent interactions that are only possible *in vitro* at high concentrations with lengthy reaction times. More recent studies suggesting that the SND proteins may be involved in TA protein targeting similarly lack convincing biochemical evidence that *in vivo* observations represent direct effects (Aviram et al., 2016; Casson et al., 2017; Haßdenteufel et al., 2017). The identification of cytosolic factors needed to cater to non TRC-dependent TA proteins therefore remains an open question.

We sought to identify and characterize the factors involved in this putative TRC-independent pathway. Based on our earlier analysis, squalene synthase (SQS) was used as a model non-TRC substrate. A combination of biochemical techniques revealed that SQS makes a distinct set of interactions in native cytosol compared to the TRC-dependent VAMP2. SQS primarily associates with a ~20 kDa factor that was identified by mass spectrometry as calmodulin (CaM). By contrast, VAMP2 associates with known components of the TRC pathway, namely TRC40, Bag6 and SGTA. The nature of the CaM association with TA substrates was probed in the PURE translation system. CaM was shown to be sufficient for preventing SQS aggregation and maintaining its translocation competence. The CaM interaction appears to be dynamic, allowing for spontaneous binding and release until substrates eventually interact with some translocation machinery at the ER membrane. CaM is capable of this function across the range of physiologic  $\text{Ca}^{2+}$  levels. However, in the absence of CaM, these low hydrophobicity TA substrates still insert with equal efficiency into the ER membrane. This is due to their ability to associate with other general TMD binding factors,

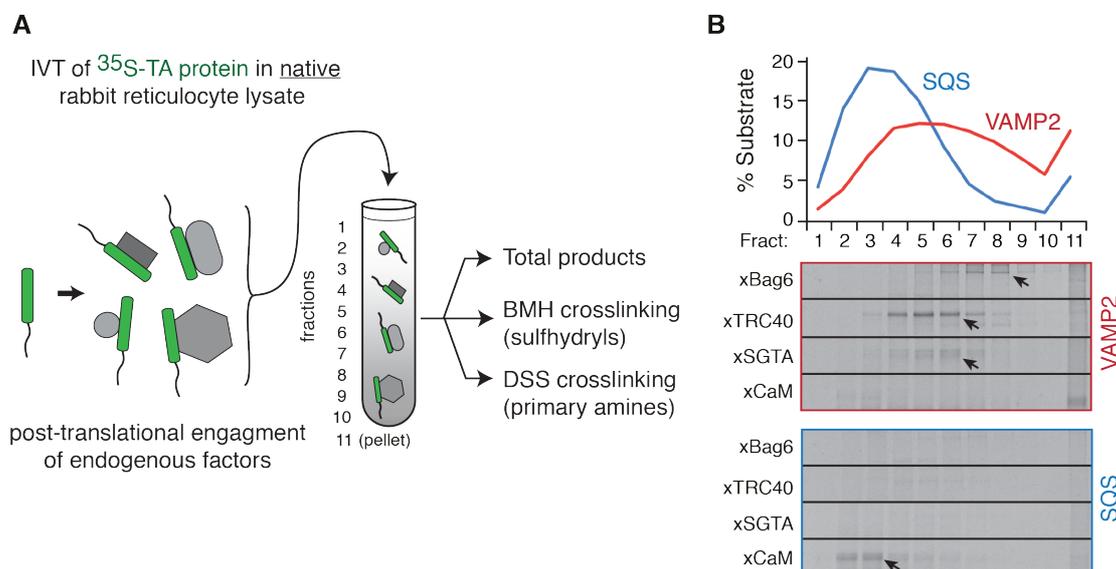
## Chapter 4: Cytosolic factors maintaining TA protein insertion competence

namely SGTA. In a purified system, engagement with SGTA was, like CaM, permissive for TA insertion. Therefore, this non-TRC route does not have a dedicated targeting factor. Rather, substrates seem to be able to utilize any TMD-shielding factor that is capable of dynamic release for attempts at membrane insertion. In native cytosol, the primary factor is the highly abundant CaM.

### 4.1 SQS and VAMP2 have different cytosolic interactions

An unbiased approach was taken to identify cytosolic machinery that post-translationally interacts with low-to-moderately hydrophobic TMD containing TA proteins. SQS was chosen to represent TRC-independent substrates based on earlier evidence that it relies on an alternative route for insertion. Its interacting partners were compared to those of VAMP2, a verified TRC pathway substrate. The *in vitro* post-translation assay developed and described in detail earlier (Chapter 3) was ideal for probing a substrate's cytosolic interaction partners as it allows for translation to be uncoupled from translocation.

A simple scheme was developed to analyze interactions made by SQS and VAMP2 in rabbit reticulocyte lysate (RRL) (Figure 4.1A). A combination of fractionation and chemical



**4.1 Analysis of interactions made by SQS and VAMP2 in native RRL.** (A)  $^{35}\text{S}$ -methionine labelled SQS and VAMP2 were translated in native RRL, separated by size on a sucrose gradient, and each fraction subjected to chemical crosslinking using either amine- or sulfhydryl-reactive cross-linkers. The translation products and their crosslinks were visualized by autoradiography. (B) Graph shows the densitometry profiles of each substrate across the gradient, and the individual panels show regions of the crosslinking gels with relevant interaction partners.

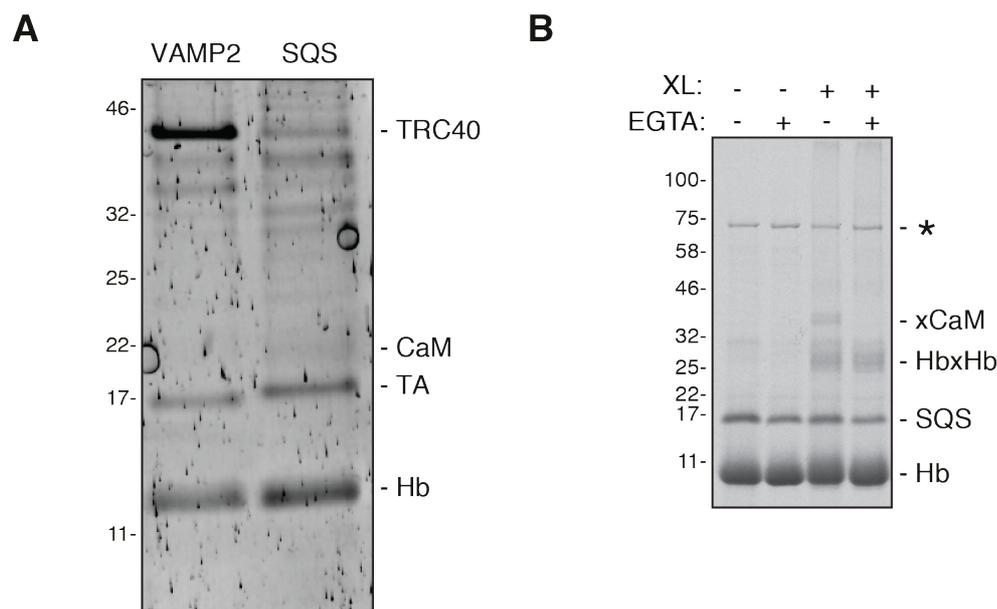
## Chapter 4: Cytosolic factors maintaining TA protein insertion competence

crosslinking strategies were used. The SQS or VAMP2 TMD in the standard TA cassette (containing an N-terminal 3X FLAG tag, and a C-terminal opsin tag; explained in Figure 3.2A) was translated in native RRL. The product of this reaction was size fractionated on a 5-25% sucrose gradient. The individual fractions were then analyzed either directly or after chemical crosslinking with bis-maleimido-hexane (BMH) or di-succinimidyl-suberate (DSS). DSS creates covalent crosslinks between primary amines, such as those at the N-terminal of polypeptides or on lysine groups. Alternatively, BMH reacts to irreversibly conjugate sulfhydryl groups, primarily linking cysteine groups to one another. The translation products and their crosslinks were visualized by autoradiography.

SQS and VAMP2 were both distributed heterogeneously amongst the different fractions (Figure 4.1B). However, these profiles were distinct, indicating that the proteins existed in complex with different sets of interaction partners. Crosslinking analysis confirmed that each substrate associated with proteins of different sizes in the various fractions (Figure 4.1B). As expected, VAMP2 interacted predominately with proteins whose sizes correspond with TRC pathway factors, namely Bag6, TRC40, SGTA, and TRC35 (Mariappan et al., 2010; Shao et al., 2017; Stefanovic and Hegde, 2007). The identities of these were confirmed by immunoprecipitation (data not shown). These crosslinks were extremely weak or non-existent for SQS. Instead, SQS crosslinked primarily to a ~20 kDa product only with DSS. Presumably, the interactions with this ~20 kDa protein is biologically relevant, with this being a factor which is somehow involved in chaperoning SQS through the cytosol. This proposal is addressed experimentally in the following sections of this chapter.

### **4.2 The primary cytosolic interaction factor for the SQS TMD is CaM**

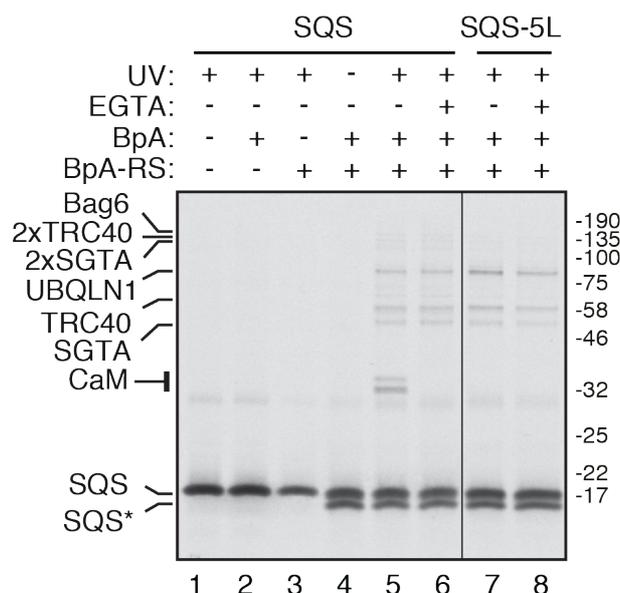
The next goal was to identify the ~20 kDa SQS TMD interacting partner that was uncovered by crosslinking analysis. For this, SQS was translated in native RRL and affinity purified under native conditions (Figure 4.2A). A large-scale translation reaction was cleared of ribosomes and immunoprecipitated against the substrate's N-terminal 3X FLAG tag. After ample washing, the bound species were eluted with the FLAG peptide. VAMP2 was used as a control to confirm that relevant interaction factors were indeed being recovered. A ~20 kD band was



**4.2 CaM is the main interacting partner of SQS.** (A) SQS was translated in native RRL and subjected to a pull down via its FLAG tag under native conditions. Associated proteins were then separated by size and subjected to Sypro staining. Bands at expected sizes based on crosslinking analysis (Figure 4.1B) were excised and identified using mass spectrometry. (B) SQS was translated in native RRL with or without 1 mM EGTA before crosslinking. The major SQS crosslinking partner (CaM) is not seen with EGTA. Haemoglobin (Hb), its inter-subunit crosslink (Hb-Hb), and an unspecified translation product (\*) are indicated.

seen specifically in the SQS sample, albeit faintly. This band was excised and identified by mass spectrometry analysis as calmodulin (CaM). By the same method, the main interacting partner for VAMP2 was identified as TRC40, as expected. Notably, CaM does not contain any cysteine residues. This explains why this crosslink was only revealed when the SQS translation sucrose gradient fractions were treated with DSS, but not BMH.

CaM is a highly conserved, ubiquitous and abundant cytosolic protein that has previously been shown to interact with both signal sequences and TMDs (Martoglio et al., 1997; Shao and Hegde, 2011b). It also acts as an intracellular calcium sensor, a property that is exploited to serve numerous different cellular processes (Means and Dedman, 1980; Means et al., 1991). It contains two EF-hand domains that can each bind two calcium ions. Calcium binding induces a dramatic conformational shift that allows CaM to completely wrap around and bind a target sequence (Finn et al., 1995; Meador et al., 1992; O'Neil and DeGrado, 1990). A flexible, methionine-rich binding groove allows it to accommodate a range of hydrophobic peptide segments. Therefore, the identification of CaM as the primary interaction factor for the



**4.3 Quantifying all SQS TMD interaction partners in native RRL.** SQS and SQS-5L were translated in native RRL containing a photo-reactive cross-linker (BpA) and BpA tRNA synthetase (BpA-RS), which permits incorporation of BpA at tyrosine codons in the TA TMD. Incorporation of BpA results in slightly faster migration of the substrate (SQS\*). Where indicated, 1 mM EGTA was included to chelate endogenous  $\text{Ca}^{2+}$ . Reactions were incubated for 15 min and immediately subjected to crosslinking by exposure to UV light. Following removal of aggregates by centrifugation, SQS or SQS-5L were recovered by immunoprecipitation against its FLAG tag. Visualization by autoradiography allowed for the quantification of all interacting partners. Data generated by Manu Hegde.

low-to-moderately hydrophobic TMDs of TA proteins is not completely surprising, and in line with its reported functions.

Many of CaM's activities, including its ability to bind substrates, is regulated by calcium (Shao and Hegde, 2011b). To test whether this may also be the case for its capacity to bind the TMD of SQS, crosslinking analysis was repeated under conditions of high EGTA (Figure 4.2B). EGTA is a commonly used chelating agent that can bind divalent ions but is more selective for  $\text{Ca}^{2+}$ . Extreme calcium depletion abolishes CaM's ability to interact with the TMD. This behaviour is identical to how CaM interacts with small secretory proteins (Shao and Hegde, 2011b).

Despite showing that CaM is the main interactor for the SQS TMD in the cytosol using multiple traditional biochemical methods, it is possible that a dedicated factor may have eluded detection. This is possible if this 'other' factor engages in low affinity interactions with the SQS TMD. The long and relatively mechanically stressful protocols needed for size fractionation and immunoprecipitation may be unfavourable for maintaining such associations. To detect interactions with the fewest assumptions and in the most native environment possible, we devised a new photo-crosslinking approach (Figure 4.3A). The promiscuous and highly

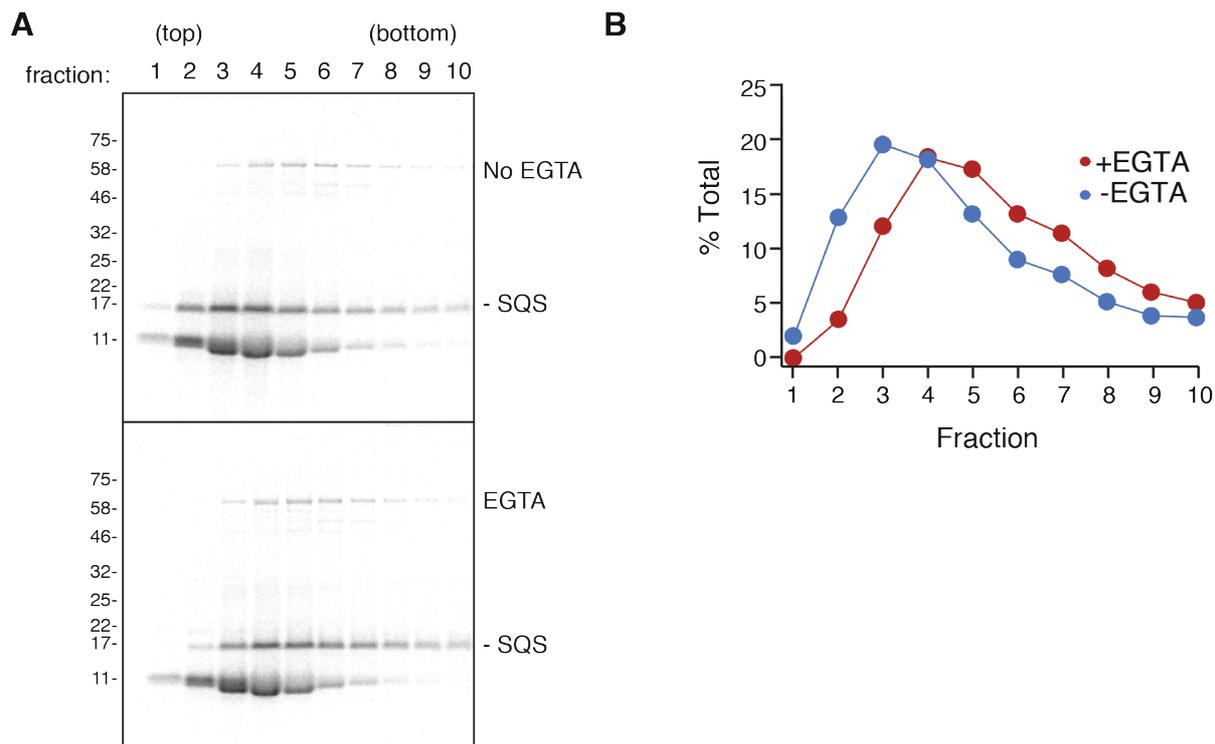
## Chapter 4: Cytosolic factors maintaining TA protein insertion competence

reactive photo-activatable crosslinker benzoyl-phenylalanine was co-translationally incorporated at multiple locations within and surrounding the TMD of SQS using sense codon suppression. Immediately after translation, the complete undiluted sample was irradiated with UV light to induce crosslinking to any factors that should contact the TMD.

For this SQS and SQS-5L were translated in native RRL supplemented with photo-reactive crosslinker benzoyl-phenylalanine (BpA) and BpA tRNA synthetase (BpA-RS) similar to the PURE system reaction (Figure 3.8). SQS-5L is a SQS mutant containing a TMD with high hydrophobicity and is therefore a TRC-dependent substrate that should interact with the same set of factors as VAMP2 (Figure 3.7). In this system, BpA-RS will charge BpA onto tyrosine-tRNA (Tyr-tRNA), thereby competing with endogenous Tyr tRNA synthetase. This will result in BpA incorporation at Tyr codons, three of which are in the TMD of SQS. Control reactions omitted BpA or BpA-RS or included 1 mM EGTA in the translation reaction to chelate endogenous  $\text{Ca}^{2+}$ . These reactions were immediately subjected to UV crosslinking. Following this, aggregates were removed by centrifugation and SQS was recovered by anti-FLAG immunoprecipitation. The major crosslinking partners for both substrates were verified by immunoprecipitation (data not shown), or sensitivity to  $\text{Ca}^{2+}$  in the case of CaM. Crosslinking to two molecules of SGTA or TRC40 occurs because these factors bind substrate as dimers.

This method allowed for the quantification of interacting partners following autoradiography (Figure 4.3B). SQS crosslinking products comprise 45% CaM, 9.7% SGTA, 12.2% TRC40, 12.2% UBQLN1, and 1.1% Bag6 of total crosslinked products in the entire lane (Figure 4.4B). The only unaccounted crosslinking bands that are visible (between the TRC40 and Ubiquilin1 bands) represent 2.9% and 3% of all crosslinking products. The remaining 14.2% of signal in the lane could not be attributed to any discernible band, and presumably represents non-specific heterogeneous crosslinking products. Therefore, CaM is the major interaction partner in complete lysate, and is specific to SQS relative to SQS-5L.

As another independent means of establishing that the majority of SQS associates with CaM was to translate it in native RRL in the absence or presence of 1 mM EGTA and analyse the reactions by size fractionation (Figure 4.4A). The SQS-CaM complex is expected to be roughly ~40-50 kDa and should therefore migrate in lighter fractions within the gradient. However, inactivation of CaM should force SQS to either associate with other factors or start to aggregate. This will manifest itself as more of it migrating deeper into heavier fractions of the gradient. Based on crosslinking results in native cytosol roughly 45% of SQS associates with CaM, meaning that there should be a corresponding portion of SQS that shifts upon



**4.4 Calcium chelation causes SQS to associate with non-CaM chaperones. (A)** SQS was translated in RRL in the absence or presence of 1 mM EGTA to chelate endogenous  $\text{Ca}^{2+}$ . Following size separation on a 5-25% sucrose gradient, SQS levels in each fraction were quantified by phosphorimaging. For reference, the peak of native haemoglobin tetramer (60 kDa) is in fraction 4. **(B)** The percent of total SQS in each fraction.

CaM inactivation. (Figure 4.3) As such, SQS was translated under both  $\text{Ca}^{2+}$  conditions and the resulting reaction separated by size on a 5-25% sucrose gradient. The percent of SQS in each fraction was calculated after exposure to a phosphorimager (Figure 4.4B). Under physiologic  $\text{Ca}^{2+}$  levels, nearly half of SQS migrates in fractions 1-3, indicating a native size of less than 60 kDa. This likely represent SQS in complex with CaM and is in line with earlier results (Figure 4.3A). Critically, this population shifts in the presence of EGTA, indicating that it can no longer associate with CaM. This provides additional evidence that CaM is the primary interaction partners for the SQS TMD.

### 4.3 CaM is sufficient to maintain insertion competence of SQS

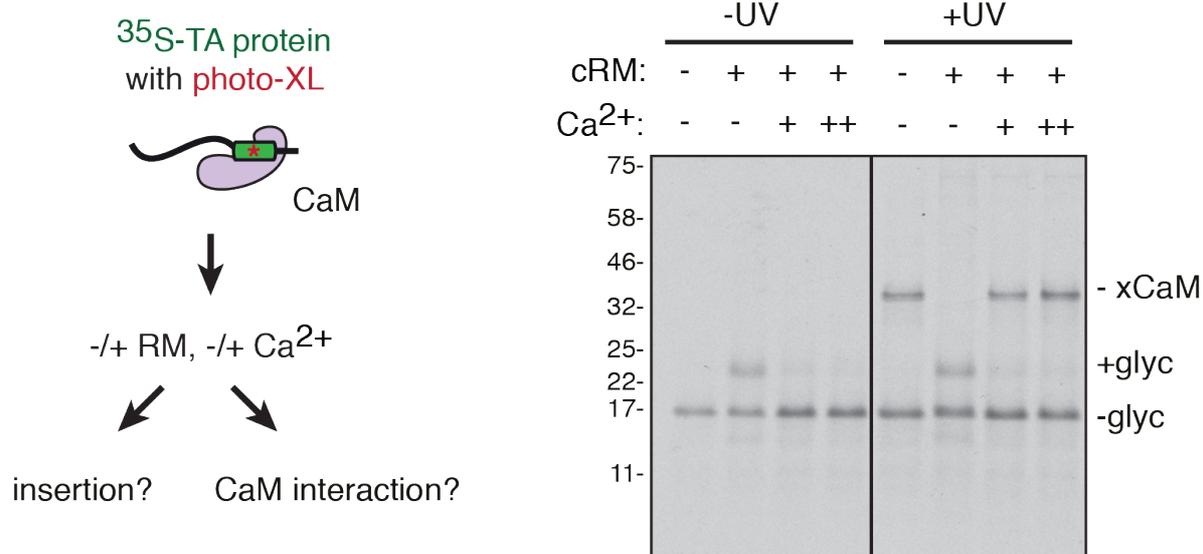
CaM interacts with a significant portion of SQS, making it reasonable to assume that this factor is the one responsible for TMD shielding in the cytosol and targeting to the membrane. However, the nature of CaM's interaction with TA proteins requires further investigation before drawing any conclusions about its possible role. Thus, the relevance of the SQS-CaM association was studied in the PURE system (described extensively in Chapter 3). First, we



## Chapter 4: Cytosolic factors maintaining TA protein insertion competence

was translated in the presence of 12  $\mu\text{M}$  recombinant CaM in the homemade amber-suppression PURE system (Figure 4.5B). SQS contained the non-natural UV-reactive amino acid benzoyl-phenylalanine (BpF) in the middle of the TMD, as previously described (Figure 3.8). Analysis of the reaction on a sucrose gradient shows that, instead of aggregating, SQS is kept soluble and co-fractionates with CaM in fractions 2-4 (Figure 4.5B). The SQS-CaM complex was isolated by pooling these fractions. UV irradiation verified that SQS contains the photo-crosslinking residue and is associated with CaM. Of note, these are the same fractions that represent the SQS-CaM complex in native RRL (Figure 4.4). Therefore, both in native lysate and in a fully purified system, CaM is capable of maintaining SQS solubility by effectively shielding its TMD to prevent aggregation.

While CaM prevents SQS aggregation, this does not necessarily mean that the interaction is meaningful for the insertion reaction. For example, TRC40 can also prevent SQS aggregation in the PURE system, but this is irrelevant for mediating SQS's insertion into the membrane (Chapter 3). It was therefore important to establish whether this was also the case for CaM. For this, SQS was first translated in the PURE system with CaM (where it should be kept soluble as a "SQS-CaM complex") or without (where it should exist as aggregates and is referred to as a the "SQS only" sample). The isolated "SQS-CaM complex" or "SQS only" translations were incubated with or without canine pancreas rough microsomes (cRM) (Figure 4.5C). In one reaction, the "SQS only" sample was supplemented with CaM post-translationally in conjunction with the addition of cRM (deemed "SQS only with CaM". After this incubation period, samples were either analyzed directly or treated with Proteinase K (PK) to assay for insertion. The PK-digested samples were subjected to immunoprecipitation using antibodies against the substrate's C-terminal opsin tag. This additional step ensured that the recovered product represented correctly inserted SQS. When starting with the CaM-SQS complex but not "SQS only" or "SQS only with CaM", SQS ends up being inserted (as indicated by the appearance of a glycosylated band). Also, predominately in the "SQS only" and "SQS only with CaM" samples there is a higher molecular weight product that represents aggregated SQS. This product is partially resistant to protease digestion. However, the aggregation product is not immunoprecipitated with the opsin antibody, which only recognizes the fully intact C-terminal protected fragment (PF) indicative of successful insertion. The small amount of insertion seen in the "SQS only" and "SQS only with CaM" samples can be explained by the small amount that escapes aggregation during the initial translation reaction



**4.6 SQS insertion in membranes is concomitant with its release from CaM.** CaM-SQS complexes were presented to cRM in the presence or absence of excess  $\text{Ca}^{2+}$  (either 0.2 or 0.5 mM) and analysed either directly (-UV) or after irradiation with UV light to induce crosslinking (+UV). Glycosylation is used to assay insertion (+glyc).

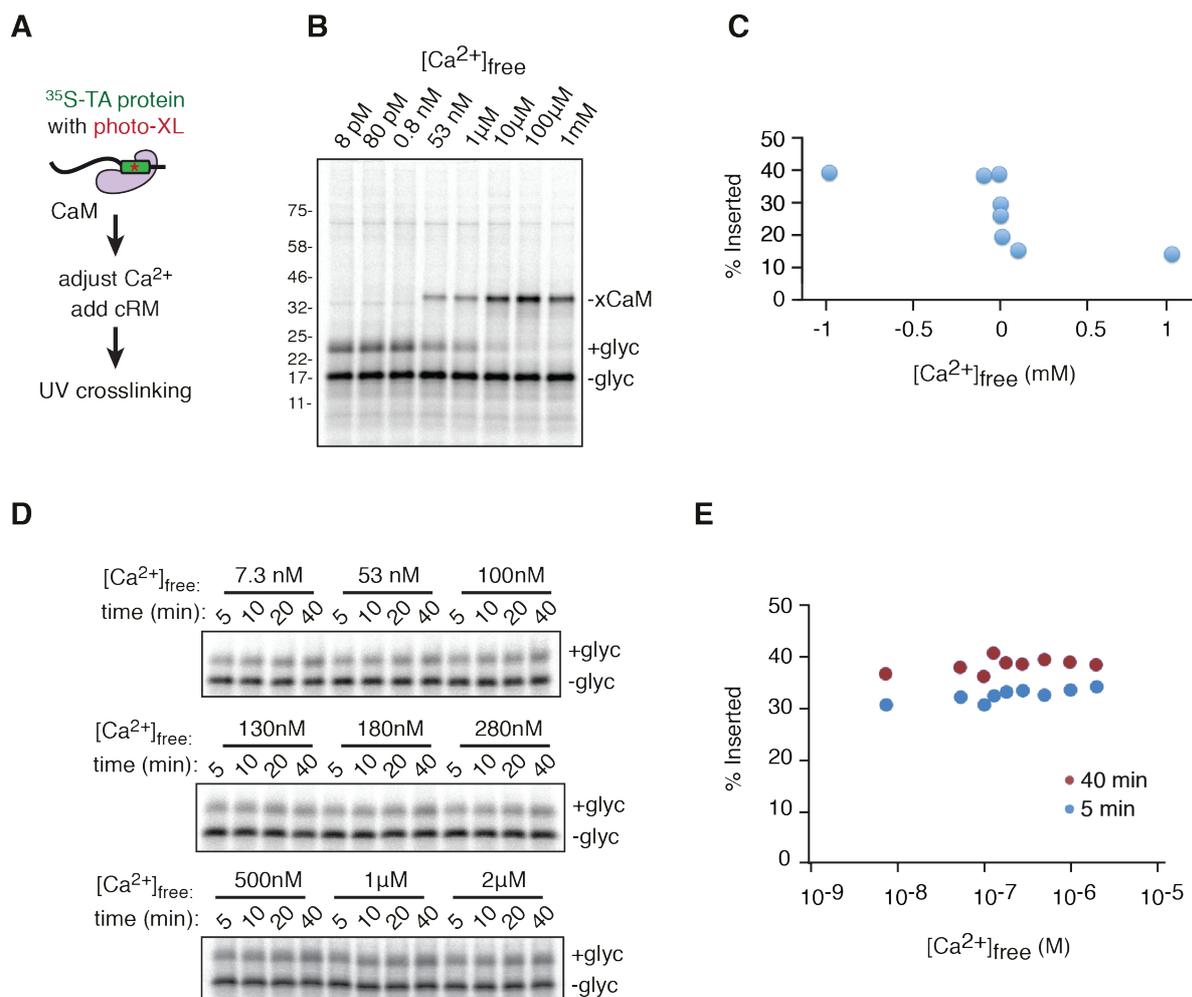
(Figure 4.5A). The insertion efficiency of SQS in the SQS-CaM sample was similar to that observed in total cytosol. These results indicate that the initial capture of SQS by CaM is essential for maintaining its insertion competence. Once aggregated, SQS can no longer engage CaM and therefore cannot insert into the membrane.

To explicitly test whether SQS insertion occurred concomitantly with release from CaM, site-specific photo-crosslinking was used to monitor the insertion reaction (Figure 4.6). Isolated SQS-CaM complexes, prepared in 100 nM  $\text{Ca}^{2+}$ , were incubated with cRM in the absence or presence of excess  $\text{Ca}^{2+}$  (either 0.2 or 0.5 mM) and analysed directly or irradiated with UV light to induce crosslinking before analysis. Excess  $\text{Ca}^{2+}$  is known to prevent substrate release once CaM is bound (Shao, 2011). In the absence of microsomes, SQS remains bound to CaM. However, the addition of microsomes prompted the release of SQS and its insertion into membranes as indicated by glycosylation. In the presence of supra-physiologic  $\text{Ca}^{2+}$ , SQS stays bound to CaM, thereby precluding it from insertion.

#### 4.4 CaM dynamically interacts with substrates at physiologic calcium levels

The loss of SQS crosslinking to CaM upon insertion into membranes is strong evidence that

## Chapter 4: Cytosolic factors maintaining TA protein insertion competence



**4.7 Sensitivity of the SQS-CaM interaction to calcium levels.** (A) Scheme used to test the dynamics of CaM association with SQS across a range of Ca<sup>2+</sup> concentrations. SQS-CaM complexes were prepared with 100 nM Ca<sup>2+</sup> and free Ca<sup>2+</sup> levels were adjusted with either Ca<sup>2+</sup> or EGTA (Schoenmakers et al., 1992). Samples were treated with UV light to induce photo-crosslinking unless otherwise stated. (B) The CaM interaction was assayed across a broad range of Ca<sup>2+</sup>. (C) Quantification of SQS insertion from panel B. (D) SQS-CaM complexes were assayed for insertion within a narrower range of Ca<sup>2+</sup>, with samples collected at the indicated time points. Reactions were not subjected to UV crosslinking. (E) Quantification of SQS insertion from (D) at 5 and 40 min.

this is an important intermediate interaction en route to the ER. However, unlike TRC40, it is unclear whether there is a dedicated CaM receptor at the ER (Erdmann et al., 2011; Shao and Hegde, 2011b). This suggests an insertion model where dynamic substrate release from CaM transiently provides opportunities for ER engagement before re-capture by CaM. If this is the case, it is important to verify that CaM is capable of this mode of interaction across the cell's full physiologic range of Ca<sup>2+</sup> concentration.

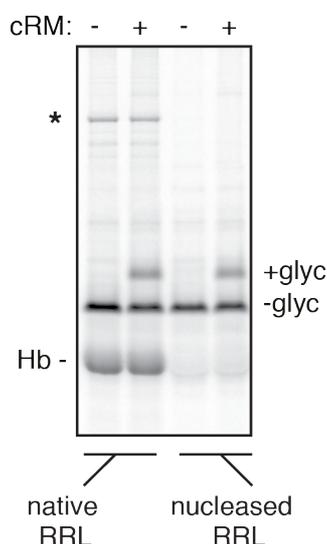
To test this, insertion reactions with the isolated SQS-CaM complex into cRM were done in the presence of a range of Ca<sup>2+</sup> concentrations (Figure 4.7A). The SQS-CaM complexes were prepared in the PURE system in the presence of 100 nM Ca<sup>2+</sup> and the concentration of

## Chapter 4: Cytosolic factors maintaining TA protein insertion competence

$\text{Ca}^{2+}$  in the final reaction was subsequently adjusted with either free  $\text{Ca}^{2+}$  or EGTA (Schoenmakers et al., 1992). Samples were then immediately removed to ice and UV irradiated to allow for visualization of SQS still in complex with CaM. At high  $\text{Ca}^{2+}$ , the majority of SQS stays bound to CaM, and does not insert into cRM (Figure 4.7B). As such, SQS-CaM complexes were assayed for insertion into cRM within a narrower range of  $\text{Ca}^{2+}$ , with samples collected at specific time points (Figure 4.7C). Resting  $\text{Ca}^{2+}$  in the cytosol is thought to be ~100 nM. During physiologic stimuli, cytosolic  $\text{Ca}^{2+}$  is thought to rise to ~1000 nM. Insertion efficiency at the collected time points were quantified (Figure 4.7D). Note that insertion is rapid, being nearly complete at 5 min. Insertion efficiency is therefore comparable within this physiologic range of cytosolic  $\text{Ca}^{2+}$ , indicating that CaM retains its ability to chaperone and dynamically interact with the SQS TMD.

### 4.5 SGTA can mediate SQS insertion in the absence of CaM

Initial experiments investigating SQS insertion showed no difference in insertion efficiency between nuclease treated versus non-nucleated RRL (Figure 4.8). The non-nucleated, native RRL system contains endogenous mRNA (primarily encoding the ~14 kDa globins and a prominent ~70 kDa protein) that contribute to background, but has the advantage of retaining

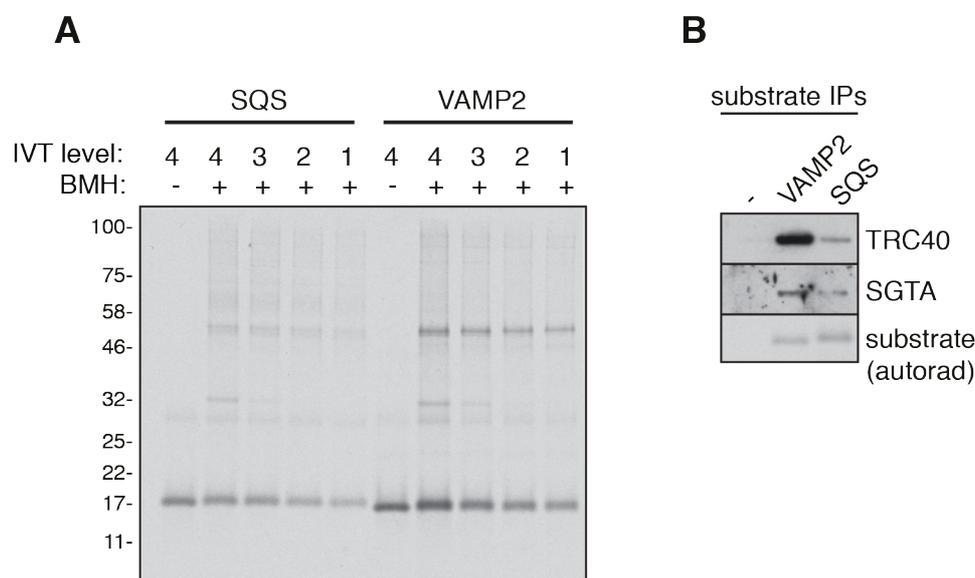


**4.8 Analysis of SQS insertion when CaM is inactivated.** SQS was translated in native RRL or nuclease-treated RRL and incubated with cRM to assay insertion. Nuclease-treated RRL contains high levels of EGTA, which renders the  $\text{Ca}^{2+}$  sensitive CaM incapable of interacting with the SQS TMD (see Figure 4.7B). Note that haemoglobin (Hb) and the 70 kDa product (\*) are not translated in the nuclease-treated RRL.

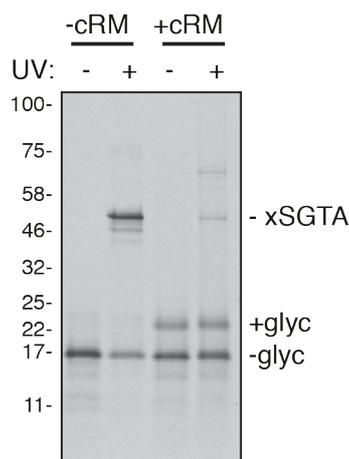
## Chapter 4: Cytosolic factors maintaining TA protein insertion competence

native cytosolic conditions, notably endogenous  $\text{Ca}^{2+}$  levels. However, to reduce background translation, native RRL can be nucleated. For this,  $\text{CaCl}_2$  is added to 1 mM and the  $\text{Ca}^{2+}$  activated nuclease from *S. aureus* is used to digest endogenous mRNA. The nuclease is then inactivated by  $\text{Ca}^{2+}$ -chelation with 2 mM EGTA (Pelham and Jackson, 1976). This manipulation means the nucleated RRL system has an estimated free  $\text{Ca}^{2+}$  level that is sub-physiologic (10 nM), rendering CaM comparatively inactive for substrate binding. As shown in Figure 4.4, CaM inactivation causes SQS to transfer onto a different set of factors, presumably other TMD binding proteins that now have increased access. The fact that insertion efficiency of SQS is largely unchanged implies that these other factors may also be able to retain SQS in an insertion competent state.

To identify these other factors, SQS and VAMP2 were translated at varying levels of expression in nucleated RRL, after which the lysate was subjected to crosslinking with BMH (Figure 4.9A). All samples were immunoprecipitated using the FLAG tag on the substrate and analysed by SDS-PAGE and autoradiography. At limiting amounts of translation, the primary interaction product should be preferentially observed, as there is minimal saturation of any cellular factors by the substrate. For VAMP2, this is TRC40. SQS also interacts with a similar sized product, but this was shown to be SGTA by immunoprecipitation (Figure 4.9B)



**4.9 SQS interactions in nucleated RRL.** (A) SQS and VAMP2 were translated in nucleated RRL at various levels and then subjected to chemical crosslinking with BMH. Samples were immunoprecipitated using the substrate's FLAG tag and analysed by autoradiography. (B) SQS and VAMP2 were translated as in (A) and immunoprecipitated under native conditions. The amount of associated TRC40 and SGTA were determined by blotting.



**4.10 SQS-SGTA insertion into membranes.** SQS-SGTA complexes were incubated either with or without membranes, then analysed directly (-UV) or after irradiation with UV light to induce crosslinking (+UV). Glycosylation is used to assay insertion (+glyc).

and mass spectrometry of a large-scale translation reaction (data not shown). This interaction was less prominent in native RRL, where CaM is the primary interaction partner (Figures 4.1, 4.3, 4.4). Thus, in the absence of CaM, a major interaction partner for SQS is SGTA. It is worth noting that SQS is also seen to interact with Bag6 and Ubiquilins at higher translation levels, presumably for the purpose of degradation in case of failed insertion (Hessa et al., 2011; Itakura et al., 2016). Furthermore, when interactions are assessed by photo-crosslinking in undiluted translation reactions (Figure 4.9A), some TRC40 interactions is also observed. This interaction is apparently dynamic and easily lost upon dilution or immunoprecipitation (Figure 3.12), explaining why it does not effectively mediate targeting and insertion of SQS (Figure 3.13). Therefore, in the absence of CaM, the general TMD binding factor SGTA can interact with SQS in nucleated RRL.

As is the case for CaM, we wanted to test whether the SGTA interaction is meaningful for targeting in a purified system (Figure 4.10). For this, the SQS-SGTA complex was prepared in the PURE system as previously described (Figure 3.9) and incubated with or without cRM. After incubation, the samples were divided and one half was subjected to UV irradiation to detect interactions with the SQS TMD. Insertion was observed as indicated by glycosylation (+glyc) and was similar in efficiency as for SQS translated in native and nucleated RRL (Figure 4.8). Furthermore, insertion was accompanied by release from SGTA. Note that, in contrast to CaM,  $Ca^{2+}$  had no effect on insertion mediated by SGTA. As such, in both complete cytosol and in a purified system, SGTA can support the insertion of SQS.

### 4.6 Discussion

From our work, it is clear that the targeting model for TA proteins with low-to-medium hydrophobicity TMDs differs substantially from the TRC pathway used by high hydrophobicity substrates. The TRC pathway is a clear example of a route that involves a dedicated targeting complex. However, the pathway described here for substrates such as SQS is a more indirect one, in which a chaperone seems predominately important for preventing aggregation and keep the substrate in a translocation competent state. The substrate is then ‘targeted’ to the membrane upon dissociation. As such, any general TMD binding factor that does not commit the substrate to a particular fate, such as degradation, can be used to fulfil this role.

The use of generalist binding factors to maintain translocation and insertion competence, but not necessarily to mediate targeting, has many biological precedents across species. For example, in bacteria substrates inserted by YidC are not delivered by a specific chaperone (Wang and Dalbey, 2011). In yeast, multiple cytosolic factors such as Hsp70 and Tcp1 associate with substrates as they are post-translationally targeted to the Sec channel. These factors keep the substrate competent for translocation but do not specifically interact with the signal sequence (Plath and Rapoport, 2000; Plath et al., 1998). In mammals, short secretory proteins (Shao and Hegde, 2011b) and many mitochondrial proteins (Wiedemann and Pfanner, 2017) also make use of general TMD binding factors. Our data indicates that the delivery of TRC-independent TA proteins exploit a similar set of interactions in the cytosol.

To ensure that a dedicated targeting factor was not missed using traditional biochemical methods, a novel photo-crosslinking system was developed that allowed us to account for all the interactions made by the SQS TMD. This permitted a full accounting of all major cytosolic physical interactions made by the TMD without relying on fractionation or other manipulations that might perturb weak interactors. In accordance with other biochemical analysis, CaM was shown to be the primary interaction partner. There were other, less prominent interactions with the quality control factor Ubiquilin1 and the known TMD chaperones SGTA and TRC40. While it is difficult to prove that no other factor may exist, careful quantification of the photo-crosslinking indicates that any such factor would be engaged with at most ~3% of substrate molecules, which would be difficult to reconcile with the observed ~30-50% insertion efficiency that occurs within 5 min (Figure 4.7). TRC40 cannot function effectively to mediate SQS targeting, and Ubiquilin is not specific for ER proteins and is likely operating as a quality

## Chapter 4: Cytosolic factors maintaining TA protein insertion competence

control factor (Itakura et al., 2016).

As such, we focused on characterizing the mode by which CaM and SGTA interact with substrates by investigating the TA-CaM or TA-SGTA complex in the PURE complex. Both CaM and SGTA can prevent SQS aggregation. At physiologic  $\text{Ca}^{2+}$ , CaM is capable of dynamically engaging with TMDs, just as it has been shown to do for signal sequences (Martoglio et al., 1997; Shao and Hegde, 2011b). Given the high abundance of CaM in the cytosol and its ability to bind hydrophobic elements reversibly, it is reasonable that it would be exploited by the cell as a buffer against signal sequence and TMD aggregation. Conversely, the fate of a substrate after SGTA binding seems to be dictated by the properties of its TMD. Should a TMD be sufficiently hydrophobic, it will be transferred onto TRC40 and targeted to the membrane this way. However, it seems that certain TA substrates such as SQS can instead engage SGTA in a dynamic, non-committal mode as part of an ‘alternative’ pathway to the ER. This TRC-independent route is apparently not accessible to TRC substrates like VAMP2 given that the VAMP2-SGTA complex is insertion incompetent into ER microsomes unless complemented with TRC40 and the Bag6 complex (Shao et al., 2017). Of note, it has previously been shown that, in concordance with our work, insertion of both TRC-dependent and independent substrates is inhibited at very high  $\text{Ca}^{2+}$  concentrations (Haßdenteufel et al., 2011). However, it is worth noting that these observations by Haßdenteufel et al. (2011) only occurred at extremely high exogenous  $\text{Ca}^{2+}$  concentrations, making it difficult to parse between consequent direct and indirect effects on insertion.

Given that these chaperones are capable of interacting with a range of substrates, they do not seem to distinguish a specificity element (i.e. the signal sequence or TMD). Therefore, substrate discrimination must occur at the membrane, presumably by the machinery which also mediates the insertion reaction. However, seeing as all these cytosolic chaperones do not have a specific ER-resident receptor, it remains unclear how these TA substrates actually get inserted into the lipid bilayer. Identifying the insertion machinery responsible for this insertion process is the subject of the following two chapters.

### 4.7. Materials and methods

#### *Constructs and standard translation methods*

Constructs for expression in the rabbit reticulocyte lysate (RRL) and PURE system were as

## Chapter 4: Cytosolic factors maintaining TA protein insertion competence

described in Chapters 2 and 3. The only additional construct used for the purification of a recombinant protein is the His-tagged calmodulin in the pRSETA vector (Shao and Hegde, 2011b). His-tagged calmodulin was expressed and purified from BL21(DE3) cells using standard methods (Shao et al., 2017).

The mammalian *in vitro* expression system using either RRL or nucleated RRL were as explained in the text or Chapter 2. Unless stated, the standard *in vitro* post-translational insertion assay described in Chapter 3 was used to probe for TA substrate insertion. Incorporation of benzoyl-phenylalanine (BpA) into RRL translation products was achieved by the addition of 250 nM BpA and 0.1 mg/ml purified recombinant BpA-RS (Chin et al., 2002) to the translation reaction. BpA-RS is a Tyr-tRNA synthetase from *Methanocaldococcus janaschii* containing mutations that permit its use of BpA instead of Tyr. Because the identity elements of eukaryotic Tyr-tRNA are recognized by the archaeal synthetase (Kobayashi et al., 2003) a subset of Tyr-tRNA in the RRL translation reaction will be charged with BpA instead of Tyr. This allows stochastic incorporation of BpA at Tyr codons. Immediately after the translation reaction, the samples were irradiated on ice with UV light from a UVP B-100 series lamp (UVP LLC) for 15 min ~8 cm from the light source. After crosslinking, the samples (typically 20  $\mu$ l) were layered onto 180  $\mu$ l of 20% sucrose in PSB and centrifuged at 100,000 rpm for 30 min in a TLA120.1 rotor (Beckman) to remove any incomplete ribosome-associated products and aggregates. The supernatant was then subject to immunoprecipitation of the substrate before SDS-PAGE.

Translations in the PURE system used the modified amber-suppression competent homemade system (Chapter 3). Purified CaM was added to translation reactions at 12  $\mu$ M in the presence of 100 nM  $\text{Ca}^{2+}$ . The SQS-CaM complex was isolated using the same method as that for TA-TRC40 or TA-SGTA complexes, but instead pooling fractions 2-4 for downstream reactions (Chapter 3).

### ***PURE system insertion reactions***

Standard insertion reactions using isolated PURE complexes were 5  $\mu$ l and included 3.5  $\mu$ l of the preformed complex, 1  $\mu$ l of membranes, and 0.5  $\mu$ l of  $\text{Ca}^{2+}$  or EGTA at the indicated concentration. To ensure that substrate release from CaM only occurred in the presence of membranes, the complex was always the last component added to the reaction following careful mixing of the membranes and adjusted  $\text{Ca}^{2+}$ . Reactions were then incubated at 32  $^{\circ}\text{C}$  for 30 min unless otherwise indicated in the text and then removed to ice. UV irradiation for

## Chapter 4: Cytosolic factors maintaining TA protein insertion competence

insertion reactions were done on ice ~10 cm away from a UVP B-100 series lamp for 15 min, after which protein sample buffer was added directly.

Protease protection assays were carried out by the addition of 0.5 mg/ml Proteinase K (PK) on ice for 1 hour. Reactions were inactivated with 1-2 mM PMSF on ice for 5 min and then rapidly denatured in 1% SDS, 0.1M Tris pH 7.5 with 1  $\mu$ l used per 10  $\mu$ l of reaction. The samples were then boiled for an additional 5 min to ensure complete PK inactivation and protein denaturation. If indicated in the text, samples were immunoprecipitated with anti-opsin antibody against the C-terminal epitope tag to ensure that protease-protected material corresponded with the correctly inserted domain. Immunoprecipitations (IPs) were performed with Protein A beads at 4 °C for 2 hours with end-over-end mixing. Samples were washed three times in IP buffer (1% Triton X-100, 50 mM HEPES pH 7.4, 100 mM NaCl) and re-suspended in 2X sample buffer. Samples were denatured at 95 °C for 3-5 min and then analysed by SDS-PAGE and autoradiography.

### ***Fractionation, chemical crosslinking, and affinity purification***

The methods for fractionation and chemical crosslinking closely followed earlier published protocols (Hessa et al., 2011; Mariappan et al., 2010; Stefanovic and Hegde, 2007). In short, 200  $\mu$ L RRL translation reactions of TA proteins (done either in the presence or absence of exogenous EGTA) were layered onto a 2 ml 5-25% sucrose gradient in physiologic salt buffer and centrifuged for 5 hours in a TLS-55 rotor. Eleven 200  $\mu$ l fractions were collected and put on ice. Aliquots of individual fractions were treated with 250  $\mu$ M bis-maleimido-hexane (BMH) for 30 min on ice for sulfhydryl-mediated crosslinking, or 250  $\mu$ M di-succinimidyl-suberate (DSS) for 30 min at 22 °C for amine-mediated crosslinking. After crosslinking, protein sample buffer was added directly for SDS-PAGE analysis. In some experiments, the sucrose gradient step was omitted, and the total translation reaction was diluted 10-fold in PSB and subjected to crosslinking as above. Affinity purification under native conditions were as previously described (Chapter 3).

## **Chapter 5: The ER membrane complex (EMC) is essential for tail-anchored protein insertion**

It is clear that a significant subset of tail-anchored (TA) proteins do not have access to the post-translational TRC pathway. These TA substrates are those which have transmembrane domains (TMDs) of low-to-moderate hydrophobicity. They are instead shielded in the cytosol by general TMD binding factors, primarily calmodulin (CaM). CaM can engage TMDs in a dynamic manner, allowing substrates multiple rounds of release and re-capture until they are inserted into the endoplasmic reticulum (ER). Despite differences in TMD properties, these and more strongly hydrophobic TMD containing TA proteins insert with equal efficiency into the ER (Chapter 4). Therefore, there must exist an equally effective means of inserting TRC-independent substrates into the ER membrane. This other insertion process must be capable of correctly recognizing less hydrophobic TMDs and partitioning them into the lipid bilayer.

Determining how this insertion process occurs is the subject of this chapter. Based on extensive previous characterization, SQS was used as a model substrate to represent low hydrophobicity TA proteins. First, we established that the insertion event is protein mediated, as SQS insertion is impaired into ER microsomes pre-treated with trypsin. This implies that there is some ER-resident machinery responsible for SQS insertion. Known translocation factors, namely the Sec61 channel and the Sec62-Sec63 complex were shown to not be involved in this process. Taking a candidate approach, we postulated that the insertion factor would be highly conserved across eukaryotes, abundant, and cause pleiotropic membrane-associated phenotypes when deleted. One such factor is the ER-membrane protein complex (EMC), a widely conserved eight to ten subunit complex of previously unknown function. Whether the EMC could act as a TMD insertase was investigated using a combination of *in vivo* and *in vitro* techniques.

Insertion assays into ER derived from EMC knockout cells showed a selective reduction in SQS insertion. This was seen when SQS was prepared both in crude cytosol or provided as a defined complex with one of its possible chaperones (CaM or SGTA). The insertion defect was specific to SQS, implying that the EMC is important for this particular class of TA proteins. The insertion of VAMP2, which uses the dedicated TRC pathway insertase WRB/CAML remained unimpaired. The behaviour of SQS *in vivo* was then investigated in EMC knockout cells. In the absence of EMC, SQS mis-localizes to the cytosol, forming punctate aggregates in high-expressing cells. This phenotype was rescued with the re-

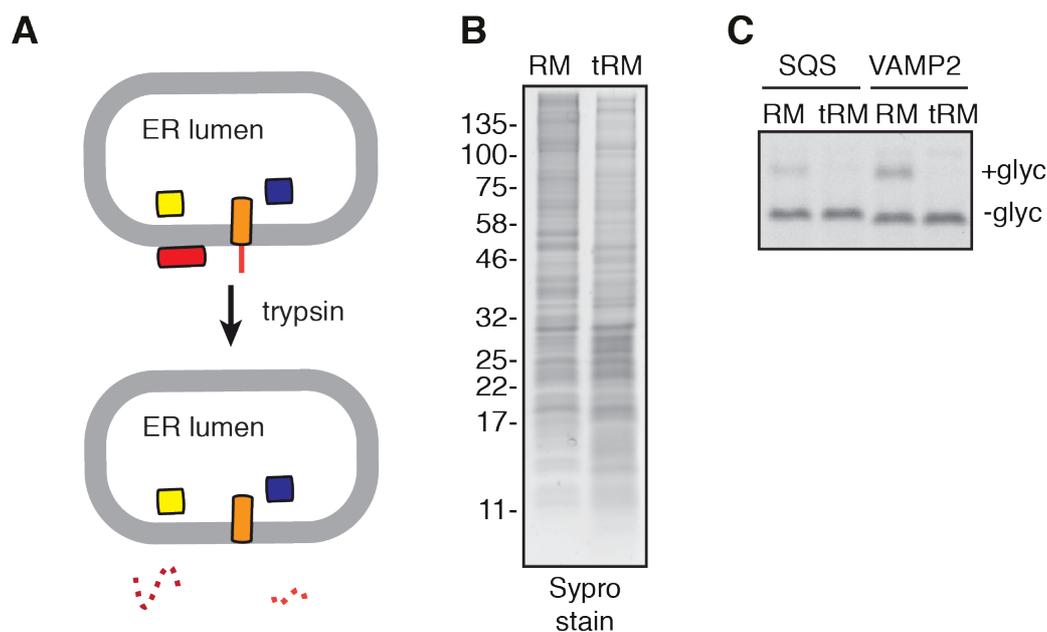
## Chapter 5: The EMC is essential for TA protein insertion

expression of the full EMC. VAMP2 localization remained unchanged. Additional flow cytometry (FACS) analysis supported the *in vivo* necessity of EMC for SQS insertion. A previously characterized set of endogenous TA proteins and SQS hydrophobicity mutants were assayed for EMC dependency *in vitro*. The insertion of all TRC-independent substrates was nearly abolished in the absence of EMC. However, all documented TRC-dependent substrates remained essentially unaffected. Therefore, both *in vitro* and *in vivo* the EMC is shown to be an essential membrane component mediating the insertion of non-TRC TA proteins.

### 5.1 SQS insertion is a protein mediated process

The possibility of alternative means of insertion to the TRC pathway for certain TA substrates is not a novel idea. Investigation into putative non TRC-dependent routes was initially prompted by the observation that depletion of TRC factors still supports the insertion of certain TA proteins (Casson et al., 2017; Jonikas et al., 2009; Rivera-Monroy et al., 2016; Schuldiner et al., 2008; Shao et al., 2017). An extensively studied proposed mode of insertion is an ‘unassisted’ one in which the hydrophobic TMD can essentially spontaneously intercalate into the ER lipid bilayer (Borgese and Fasana, 2011). This model does not implicate the need for additional ER-resident machinery. Rather, insertion specificity is imparted by differences in the cholesterol level of the lipid bilayer between the various membrane bound organelles (Brambillasca et al., 2005). The most thoroughly investigated substrate which has been demonstrated as capable of unassisted insertion *in vitro* is the mammalian TA protein cytochrome b5 (Cb5). The hydrophobicity of Cb5’s TMD is remarkably low for a TA protein, in the same range as SQS (Chapter 4). Indeed, Cb5 translated in rabbit reticulocyte lysate (RRL) and allowed to insert into lipid bilayers can translocate its C-terminus as rapidly and as efficiently as into microsomal membranes (Brambillasca et al., 2005). While Cb5 can interact with cytosolic factors, these are not necessary for mediating the insertion reaction, but are thought to act mainly to prevent Cb5 from aggregation, thereby keeping it competent for insertion (Colombo et al., 2009).

The strong similarity between the biophysical properties of the SQS and Cb5 TMD (Figure 3.3), as well as the obvious parallels comparing substrate interactions in the cytosol (Figure 3.4) made it necessary to consider this ‘unassisted’ pathway as a viable option for SQS insertion. For this, SQS insertion was tested *in vitro* in the presence of compromised ER



**5.1 SQS insertion depends on intact membranes.** (A) Effect of trypsin treatment on ER membrane proteins. Luminal proteins (yellow, blue) and membrane embedded (orange) fragments are protected from trypsin cleavage while peripheral proteins and cytosolically exposed fragments (red) are digested. (B) Sypro stained gel showing the protein profile of hRMs that were untreated (RM) or digested with trypsin (tRM). (C) Analysis of SQS and VAMP2 insertion using trypsin treated microsomes from (B).

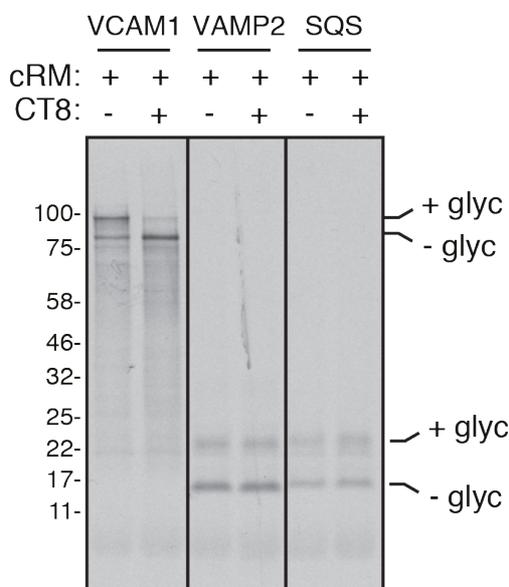
membranes. To mitigate the possibility that any observed effects were a consequence of using highly artificial membranes, HEK293 cells were used as the source of ER. Using standard methods (Zhang et al., 2013), microsomes were made from adherent cells using a series of fractionation steps. These will be referred to as hRM (human rough microsomes). To disturb the protein profile of the membranes while keeping the lipid bilayer intact, hRM were pre-treated with trypsin (Figure 5.1A). Trypsin is a serine protease that promiscuously cleaves peptide chains, mainly at lysine and arginine residues. Trypsin treatment should result in the cleavage of peripherally associated proteins and the cytosolic domains of integral membrane proteins. Clipping of these fragments should render most proteins non-functional, due either to the disruption of a critical catalytic site, interaction domain, or the triggering of misfolding. Luminal proteins and those entirely embedded in the membrane should remain intact. Trypsin treated hRM still have their glycosylation machinery fully functional as it is entirely luminal (Braunger et al., 2018; Wild et al., 2018), allowing this to be used as an assay for insertion. The protein profiles of pre-and post-trypsin treated hRM indicate that a significant number of proteins are cleaved, while the same amount of membranes are recovered based on the equal amounts of certain protected proteins (Figure 5.1B).

## Chapter 5: The EMC is essential for TA protein insertion

These microsomes were then used in the *in vitro* post-translational insertion assay with SQS and VAMP2 expressed in the standard TA cassette (Figure 3.2A). Trypsin treatment inactivates both VAMP2 and SQS insertion (Figure 5.1C). This result was independently verified using protease protection as an assay for insertion (data not shown). Insertion of VAMP2 depends on critical interactions between TRC40 and the cytosolic fragments of the WRB/CAML complex (Mariappan et al., 2011; Wang et al., 2014). Trypsin treatment destroys these domains, thereby preventing VAMP2 insertion. However, compromising the ER also abolishes the ability of SQS to insert into the membrane. This is in contrast to the predictions of the ‘unassisted’ insertion model, where any disruptions to ER proteins should not affect the ability of SQS to interact with and insert into the unperturbed lipid bilayer. This important result implies that this critical insertion step is a protein mediated process.

### 5.2 Known Sec systems are not involved in SQS insertion

Since SQS insertion is a protein mediated process, we decided to test whether ER proteins with known roles in protein translocation may be responsible. Sec61, Sec62 and Sec63 were compromised using either chemical inhibition or siRNA gene knockdown. Though primarily used by co-translational substrates, the Sec translocon can also mediate the post-translational

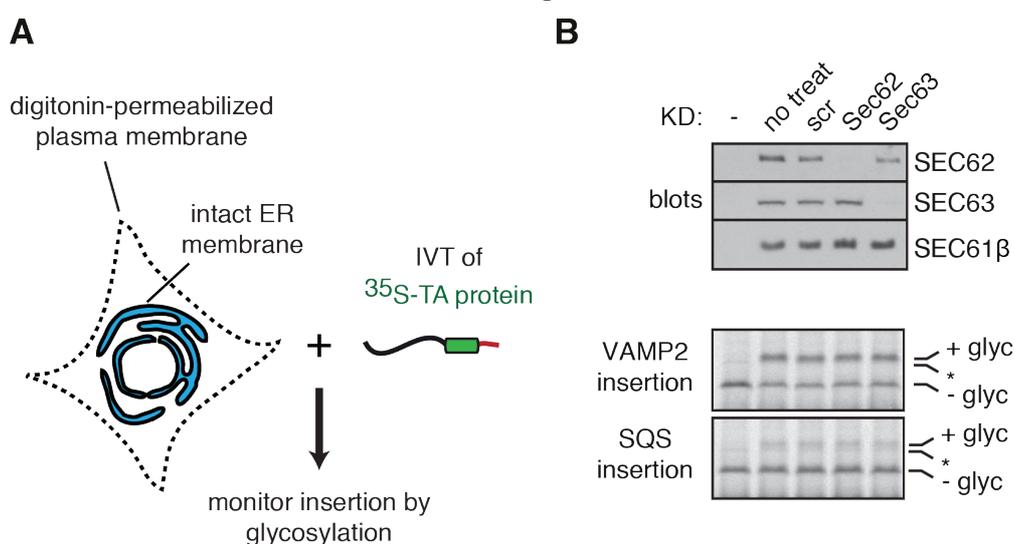


**5.2 Sec61 is not involved in SQS insertion.** VCAM1, VAMP2 and SQS were analysed for translocation or insertion (as indicated by glycosylation) into cRM in the absence or presence of 1  $\mu$ M cotransin (CT8), an inhibitor of the Sec61 translocon (Garrison et al., 2005; MacKinnon et al., 2007). VAMP2 and SQS were tested post-translationally, while VCAM1 was tested co-translationally.

insertion of soluble proteins (Rapoport, 2007). Earlier crosslinking analysis of TA proteins to SRP suggested a role for the Sec61 translocon during TA protein insertion (Abell et al., 2003). However, both biochemical reconstitution experiments, analysis of conditional mutants, and gene knockdown experiments suggests that the translocon is likely not involved in mediating the insertion reaction (Kutay et al., 1995; Lang et al., 2012; Steel et al., 2002; Yabal et al., 2003).

The possibility of Sec61 involvement in TA insertion was tested in our system (Figure 5.2). Cotransin (CT8), a specific chemical inhibitor of the Sec61 translocon was used to remove this as a viable route into the membrane (Garrison et al., 2005; MacKinnon et al., 2007). VCAM1, VAMP2 and SQS were analysed for translocation or insertion into canine rough microsomes (cRMs) in the presence or absence of 1 $\mu$ M CT8. VCAM1 is a known co-translational substrate whose insertion should be severely reduced upon Sec61 channel inhibition. VCAM1 was tested co-translationally, while VAMP2 and SQS were tested post-translationally. In each case, nucleated RRL was used for translation, and translocation was assayed by glycosylation. VCAM1 translocation is clearly impaired in the presence of CT8 as marked by a decrease in the glycosylation, indicating that the activity of the Sec61 channel is compromised. Despite this, VAMP2 and SQS insertion is unimpaired in the presence of CT8, suggesting that they do not use the Sec61 channel for access to the lipid bilayer.

Another complex which has been implicated in translocation is the abundant ER homodimer Sec62-Sec63. The Sec62-Sec63 complex was first identified in close association



**5.3 The Sec62/Sec63 complex is not involved in SQS insertion.** (A) Diagram of insertion assay using semi-permeabilized cells. (B) HEK293 cells were knocked down with SEC62 or SEC63 siRNA for 72 hours. Knockdown was verified by immunoblotting. Cells were then semi-permeabilized and used to test insertion of SQS and VAMP2. A non-specific product (\*) is observed just below the glycosylation product in all lanes.

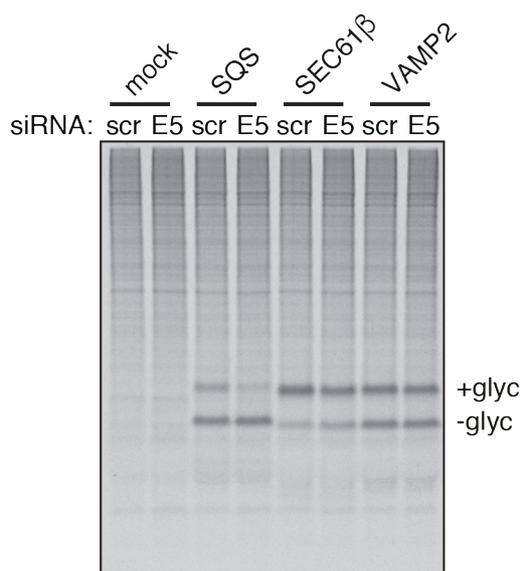
## Chapter 5: The EMC is essential for TA protein insertion

with the Sec61 translocon. The Sec61-Sec62-Sec63 ‘supra-complex’ was observed in a ribosome-free state. Though purely speculative, this was taken to infer that this complex may have a role in post-translational insertion (Meyer et al., 2000). Indeed, later studies showed that the Sec62-Sec63 complex is important for mediating the translocation of small secretory proteins. Though not directly responsible for the insertion reaction, it is thought that Sec62 somehow initially associates with substrates before passing them off to the Sec61 translocon which is responsible for their insertion (Lakkaraju et al., 2012; Lang et al., 2012). As the precise role of the Sec62-Sec63 complex is unclear, we decided to see whether it may have any involvement in TA insertion.

As there are no known chemical inhibitors of Sec62 or Sec63, semi-permeabilized cells which were subjected to siRNA knockdown treatment were used as the source of ER in the standard *in vitro* post-translational assay (Figure 5.3). Semi-permeabilization with a mild concentration of the detergent digitonin permeabilizes the plasma membrane while leaving the ER membrane intact (Setoguchi et al., 2006) (Figure 5.3A). The permeabilization of the plasma membrane grants access to translated proteins for targeting to the ER. The glycosylation machinery remains fully functional, allowing this to be used as a readout for successful insertion. Prior to semi-permeabilization HEK293 cells were treated with control, SEC62, or SEC63 siRNA for 72 hours to knockdown the respective targeted protein (Figure 5.3B). Effective knockdown of Sec62 and Sec63 were verified by immunoblotting. SQS and VAMP2 were then translated in RRL and allowed to insert into siRNA treated cells. Neither VAMP2 nor SQS showed any insertion impairment in Sec62 or Sec63 knockdown backgrounds. Therefore, it is likely that the Sec62-Sec63 complex has no meaningful role in the insertion of TRC-independent TA proteins.

### **5.3 The EMC is essential for SQS insertion *in vitro***

The insertion of low-to-moderate hydrophobicity TA proteins is protein mediated. Yet, none of the ER-resident proteins reported to have roles in insertion seem to be involved in this process. We therefore had to take an exploratory approach and choose certain promising candidates for further investigation. Since the insertion of TA proteins is an essential process in all organisms, we reasoned that an insertion factor would have a similar set of characteristics as other insertases, such as the Sec61 translocon and the WRB/CAML complex. Primarily, this putative factor should be widely conserved across eukaryotes and expressed at relatively high

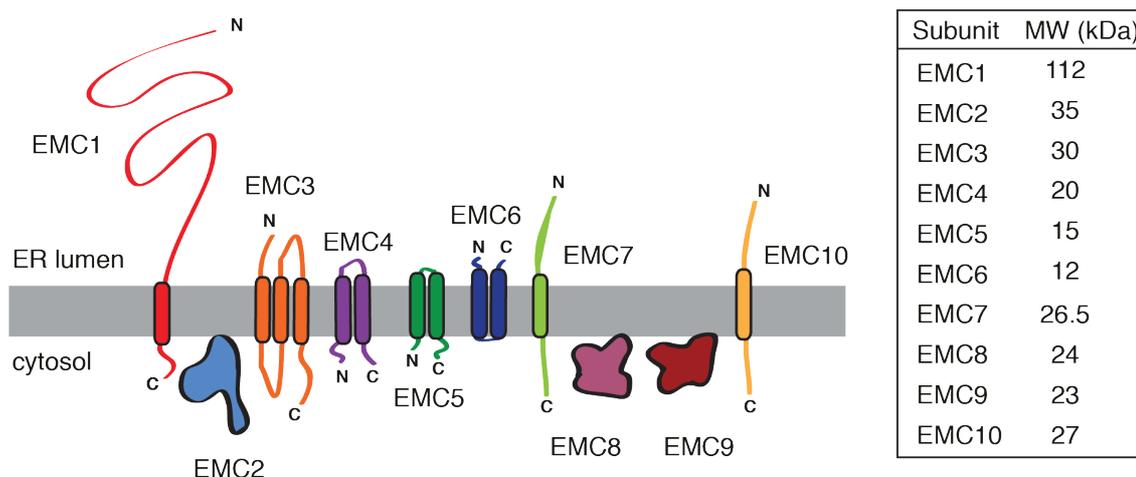


**5.4 SQS insertion is impaired in the absence of EMC5.** HEK293 cells were treated with EMC5 siRNA for 72 hours and used in an insertion assay for SQS, Sec61 $\beta$ , and VAMP2. Note that EMC5 is essential for the stability of the whole EMC complex (Figure 5.5). In this particular experiment, the translation reaction was not treated with puromycin before the addition of semi-permeabilized cells, resulting in some translation of endogenous cellular mRNAs, which is observed in all of the lanes.

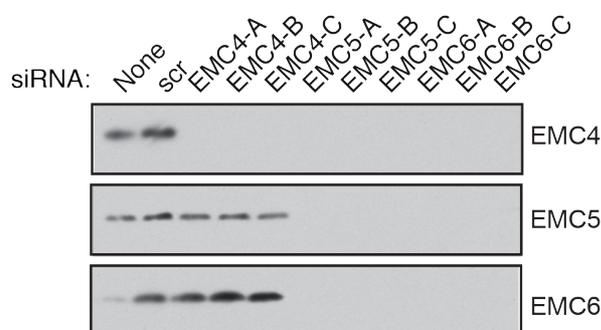
levels in the ER. Its deletion should result in pleiotropic phenotypes, as the insertion of a large number of TA proteins with various functions would presumably be impaired. One factor that meets all these pre-requisites is the ER membrane protein complex (EMC). The EMC is a widely conserved eight to ten subunit complex of unknown function (Christianson et al., 2012; Jonikas et al., 2009; Wideman, 2015). The EMC has been genetically implicated in a number of seemingly unrelated membrane associated processes such as quality control, trafficking, protein maturation, and lipid homeostasis (Christianson et al., 2012; Jonikas et al., 2009; Lahiri et al., 2014; Louie et al., 2012; Richard et al., 2013; Satoh et al., 2015). More importantly, the EMC was a notable hit in our *in vivo* screen looking at the biogenesis of multi-pass IMPs (discussed in Chapter 2). However, its biochemical function remains elusive.

The possible role of EMC in SQS insertion was explored *in vitro*. SQS and VAMP2 were translated in RRL and post-translationally targeted to semi-permeabilized cells which had been treated with EMC5 siRNA. EMC5 (MMGT1) is one of the ten EMC subunits and has previously been implicated in numerous membrane related phenotypes (Christianson et al., 2012; Jonikas et al., 2009; Lahiri et al., 2014; Louie et al., 2012; Satoh et al., 2015). HEK293 cells were treated with control or EMC5 siRNA for 72 hours. Near-complete knockdown of EMC5 was verified by immunoblotting (Figure 5.5B). Insertion of SQS was noticeably

**A**



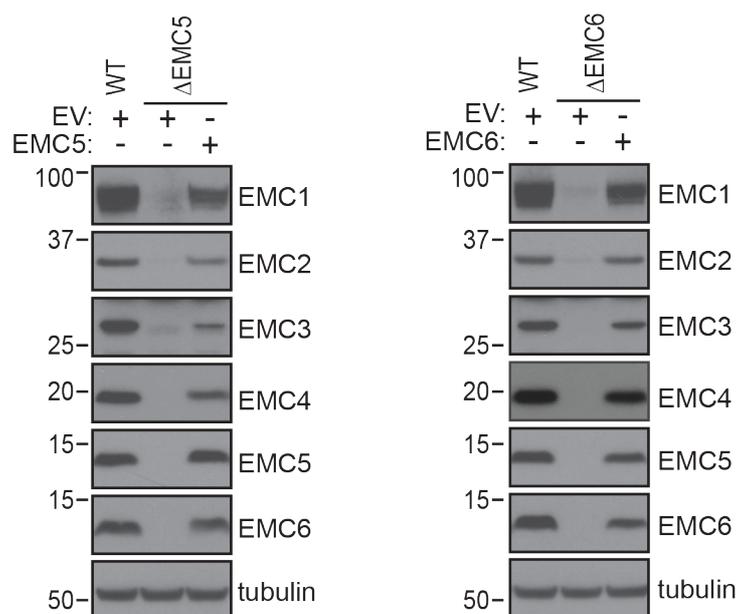
**B**



**5.5 The EMC subunits. (A)** Schematic diagram of the ten subunits of the mammalian EMC and their approximate molecular weights. Seven of the subunits are predicted to be integral membrane proteins, while the other three are likely peripherally associated. **(B)** HEK293 cells were knocked down with EMC4, EMC5 or EMC6 siRNA for 72 hours and the remaining levels of the indicated subunits were probed by immunoblotting. Knockdown of EMC5 and EMC6 leads to the reduction of the other two subunits, suggesting that they are essential for the structural integrity of the entire complex.

impaired in the absence of EMC5, while little or no effect was observed for either Sec61 $\beta$  or VAMP2 (Figure 5.4). This initial promising result suggests that this subunit of the EMC is important for the insertion of SQS. However, as the insertion defect does not extend to VAMP2 or other substrates, the EMC may be specifically responsible for mediating the insertion of low-to-moderate TMD TA proteins.

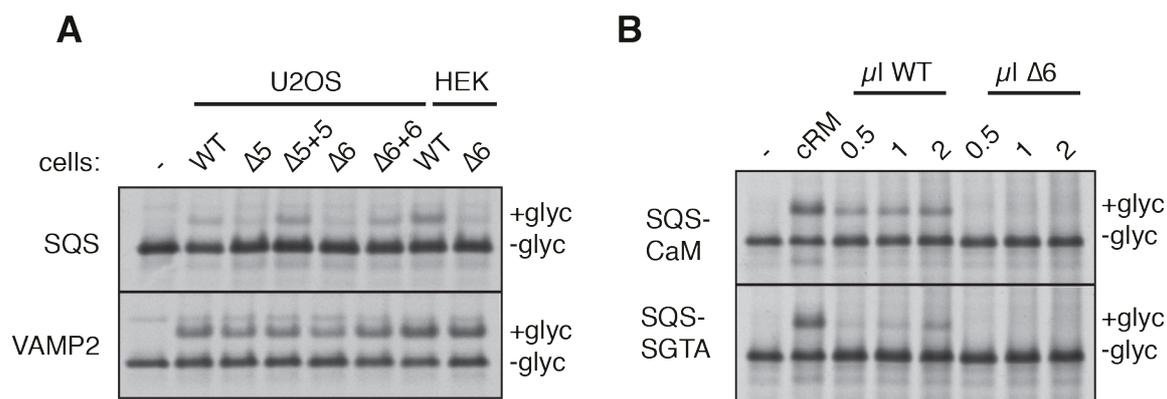
As the EMC5 knockdown may not be absolutely complete, we sought to ablate EMC5 by generating knockout cell lines. With this aim, U2OS cells were genetically modified to generate various EMC subunit knockout and rescue cell lines. The mammalian EMC is composed of ten subunits with varying topologies (Figure 5.5A). Seven of the subunits are predicted to be integral membrane proteins, while the remaining three are peripherally associated. A cursory analysis to investigate the relationship amongst the subunits using siRNA



**5.6 Characterization of EMC KO and rescue cell lines.** Flp-in Tet-on U2OS cell lines were subjected to CRISPR/Cas9 gene editing to disrupt EMC5 or EMC6. The resulting cell lines were rescued by re-introducing doxycycline-inducible expression constructs for EMC5 or EMC5 in the FRT site of the appropriate cell line. As a control, an empty vector (EV) was also introduced in each knockout cell line. Shown is the immunoblot of several of the EMC subunits and tubulin in the parental, knockout, and rescue cell lines. Cell lines and blots were generated by Norbert Volkmar.

knockdown revealed that EMC5 and EMC6 are integral to the stability of the entire complex (Figure 5.5B). As such, these were selected as the best subunits for genetic deletion. The Flp-in Tet-on U2OS cell line was subjected to CRISPR/Cas9-mediated gene editing to disrupt either EMC5 ( $\Delta$ EMC5) or EMC6 ( $\Delta$ EMC6) (Fig 5.6). The resulting cell lines were verified as knockouts by immunoblotting. As expected, deletion of either EMC5 or EMC6 resulted in the loss of all ten subunits. Cells were rescued by introducing doxycycline-inducible expression constructs for EMC5 or EMC6 into the FRT site (referred to as “ $\Delta$ EMC5 rescue” or “ $\Delta$ EMC6 rescue”). Induction of the rescue in the respective cell line restores expression of all EMC subunits to almost normal levels. To verify that any effects on protein insertion were not specific to the U2OS cell line, EMC6 knockouts were also generated in a HEK293 background using the same methods. Microsomes were generated from HEK  $\Delta$ EMC6 and wild type cells for use in insertion assays where indicated (hRM).

SQS and VAMP2 were then assayed for post-translational insertion in the absence of EMC. Substrates were translated in RRL and targeted to semi-permeabilized cells, with glycosylation used as a proxy for successful insertion. Insertion of SQS, not VAMP2, was impaired in the  $\Delta$ EMC5 and  $\Delta$ EMC6 background in both U2OS and HEK cells (Figure 5.7A).



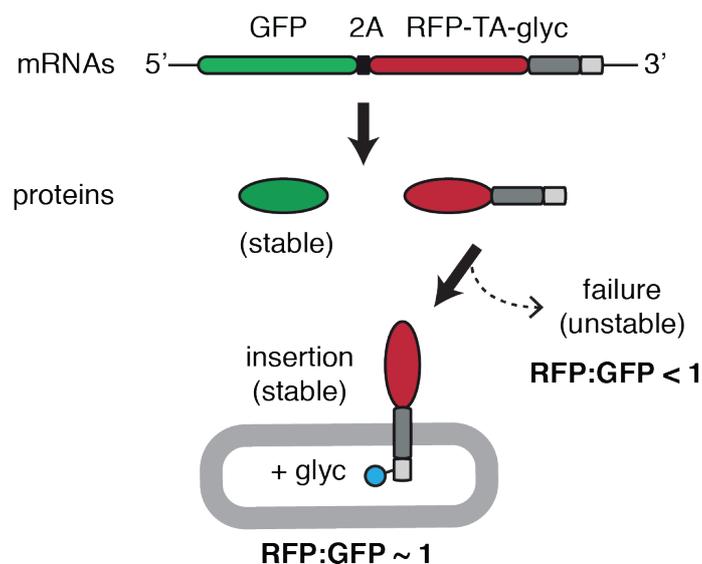
**5.7 SQS cannot insert in the absence of an intact EMC *in vitro*.** (A) Semi-permeabilized cells from the indicated cell lines (see Figure 5.6) were tested for insertion of SQS and VAMP2 using glycosylation as the readout. “-” indicates a control reaction lacking membranes. (B) The isolated SQS-CaM or SGTA-CaM complex was tested for insertion into either cRM or different amounts of hRM from wild type (WT) or  $\Delta$ EMC6 ( $\Delta$ 6) HEK293 cell lines.

This specific deficiency was rescued by re-expressing EMC5 or EMC6 in the respective knockout cell line. To ensure that the EMC has a general role in insertion, SQS was provided as a defined complex with either CaM or SGTA (Figure 5.7B). These complexes were prepared in the PURE system as previously described (Chapters 4 and 5). SQS-chaperone complexes were presented to increasing amounts of  $\Delta$ EMC6 or wild type (WT) hRM. Both SQS-CaM and SQS-SGTA complexes showed EMC-dependent insertion. Note that pancreas derived microsomes (cRM) typically show higher overall efficiencies of protein translocation and insertion than hRM, presumably because pancreatic ER is more highly enriched in biosynthetic machinery and results in a purer ER preparation. EMC-dependence is therefore evident both when SQS is prepared in crude cytosol or provided as a defined complex. These results suggest that EMC is necessary for SQS insertion *in vitro*.

#### 5.4 *In vivo* consequences of EMC deletion on TA insertion

As SQS cannot insert into the ER in the absence of the EMC in our *in vitro* system, it should similarly be impaired *in vivo*. To test this, a combination of fluorescence-activated cell sorting (FACS) and microscopy was used to investigate the role of the EMC in live cells. We exploited the fact that non-inserted TA proteins are typically degraded (Hessa et al., 2011; Itakura et al., 2016) to analyse SQS insertion in cells. A dual-colour fluorescence vector was designed to distinguish between TA protein insertion versus degradation (Figure 5.8). For this, an RFP-tagged TA protein construct was modified to contain the TMD of either SQS or VAMP2. To

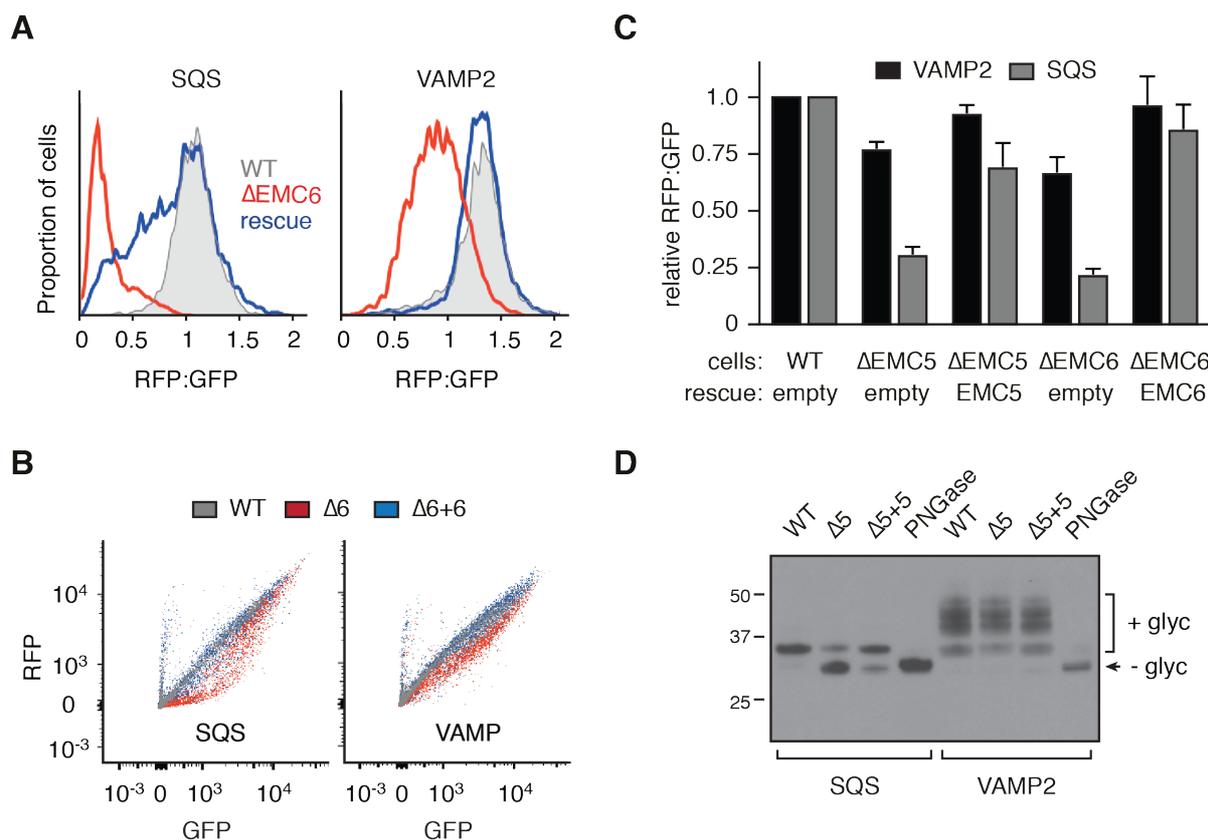
## Chapter 5: The EMC is essential for TA protein insertion



### 5.8 Visualizing TA protein fate *in vivo*. Diagram depicting construct and scheme for detecting TA insertion in cells.

monitor insertion, a C-terminal opsin tag with a glycosylation consensus sequences was appended after the TA TMD. As an internal control for translation, GFP was added at the N-terminus and separated from the RFP by a viral P2A sequence. The P2A sequence results in the skipping of a peptide bond during translation, resulting in two separate translation products (Sharma et al., 2010). In this case, the GFP serves as an expression control against which the level of RFP-TA can be normalized. If the RFP-TA is successfully inserted into the ER, it is relatively stable with a long half-life compared to the situation where insertion fails (Hessa et al., 2011; Itakura et al., 2016). Therefore, the ratio of RFP to GFP can be used as an indicator of TA fate. Should the RFP:GFP ratio equal  $\sim 1$ , this would mean that the majority of the translated TA protein has been inserted into the membrane and therefore avoided degradation. However, a lowering of the RFP:GFP ratio would indicate destabilization of the TA protein, which likely represents failed insertion.

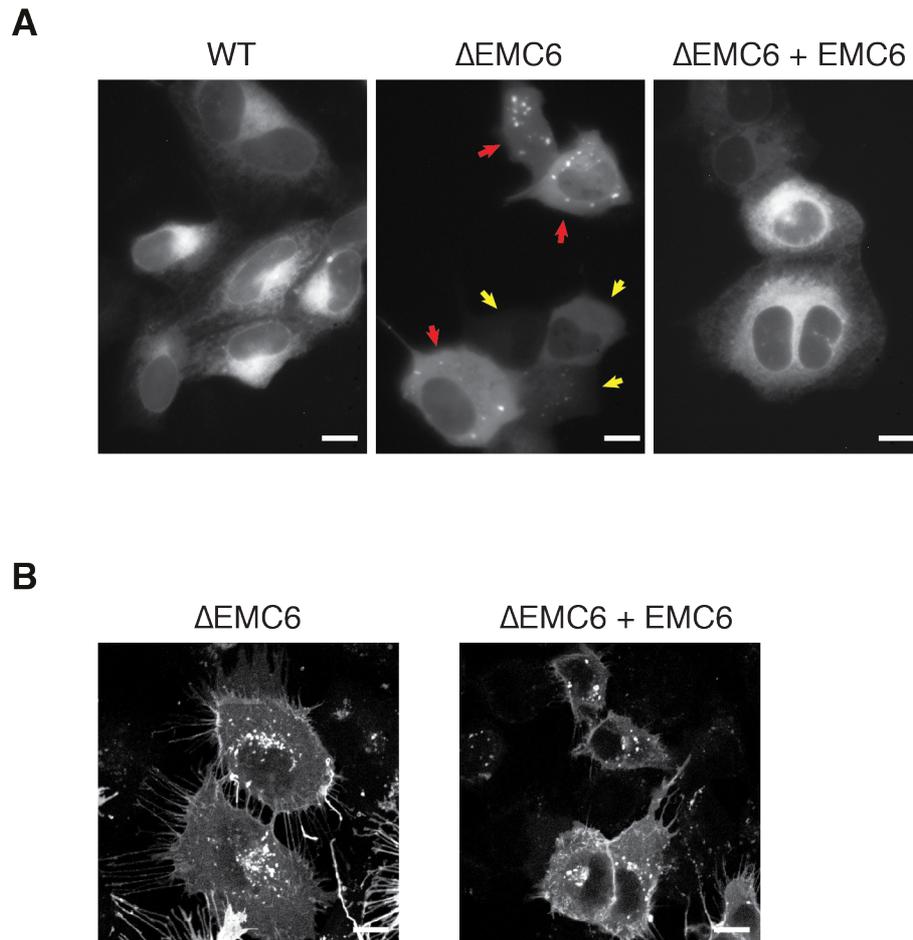
Using this assay, SQS and VAMP2 were transiently expressed in WT,  $\Delta$ EMC6 and  $\Delta$ EMC6 rescue U2OS cells (Figures 5.9A, 5.9B). After 24 hours, transfected cells were analysed for fluorescence by FACS. The GFP:RFP ratio was calculated on a cell by cell basis. In the absence of EMC, there is a significant decrease in the levels of SQS-RFP. Relative to WT, the peak of the RFP:GFP ratio is  $\sim 5$ -fold lower in  $\Delta$ EMC6 cells (Figure 5.9A). This effect is largely rescued by re-expression of EMC6. In the case of VAMP2, the effects are more modest, with the RFP:GFP ratio decreasing by less than 2-fold in  $\Delta$ EMC6 cells. Given that there is no detectable deficiency in VAMP2 insertion in our *in vitro* assays, the small effect in



**5.9 SQS cannot insert in the absence of an intact EMC *in vivo*.** (A) Histograms of flow cytometry analysis of RFP-SQS and RFP-VAMP2 in wild type (WT, grey),  $\Delta$ EMC6 (red), or  $\Delta$ EMC6+EMC6 (rescue, blue) cell lines. (B) Scatter plots of (A). (C) Tabulated mean of RFP:GFP ratios for SQS (grey bars) and VAMP2 (black bars) in the indicated cell lines. Results for each construct were normalized to the value in wild type cells and depict mean  $\pm$  SD from three independent experiments. (D) Immunoblot for SQS-RFP and VAMP-RFP in the indicated cell lines, with loading normalized to GFP expression as determined by flow cytometry. An aliquot of WT sample digested with the glycosidase PNGase shows the non-glycosylated substrate. Raw data collected by Norbert Volkmar and analysed with help from Szymon Juskiewicz.

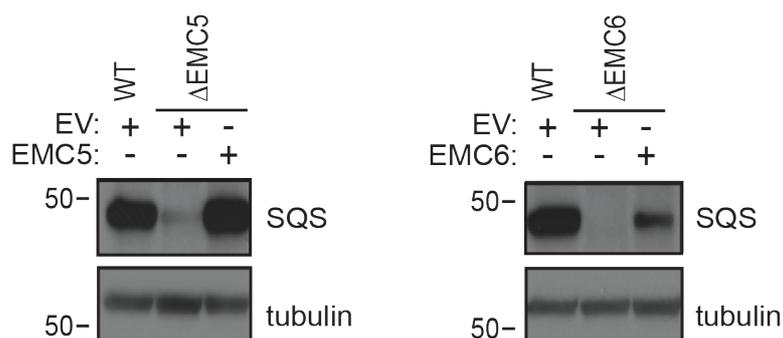
cells may be an indirect one. This is most likely as a consequence of TRC pathway saturation when the EMC pathway is unavailable but could also represent failures in VAMP2 trafficking or post-translational turnover.

To ensure that the selective impairment of SQS insertion in  $\Delta$ EMC6 was not due to clonal differences or the degree of adaptation between cell lines, this analysis was repeated in  $\Delta$ EMC5 cells (Figure 5.9C). Both loss of EMC5 and EMC6 result in the loss of the EMC complex. As such, if the EMC is responsible for SQS insertion,  $\Delta$ EMC5 cells should show the same SQS insertion phenotype. Indeed, the RFP:GFP ratio is significantly lower for SQS-RFP in  $\Delta$ EMC5. There is minimal effect on VAMP2. Three independent experiments in different cell lines were carried out and quantified, confirming that this effect is consistent.



**5.10 SQS mislocalizes and aggregates in the absence of EMC. (A)** Live cell images of GFP-SQS in the indicated cell lines shows altered localization in  $\Delta$ EMC6 cells. In low expressing cells (yellow arrows), SQS is diffusely cytosolic, while punctae in high expressing cells (red arrows) represent aggregates. **(B)** Localization of GFP-VAMP2 showing no change in localization in the absence of the EMC. Images generated by Norbert Volkmar.

Immunoblot and microscopy analysis corroborate that EMC is necessary for SQS insertion. Immunoblots were done to probe for the amounts of SQS-RFP or VAMP2-RFP expressed in various EMC KO cell lines (Figure 5.9D). Loading was normalized to equivalent amounts of GFP expression as determined by flow cytometry. An aliquot of TA substrates presented to WT cells was digested with the glycosidase PNGase, which cleaves off glycans. This control indicates that differences in banding of the TA protein represented glycosylation that occurred as a consequence of insertion. Glycosylation of the ER-resident SQS is limited to the core N-glycan, while VAMP2 acquires complex glycans due to trafficking through the Golgi. The immunoblot analysis nicely mirrors the conclusions from the FACS analysis. SQS is not glycosylated in  $\Delta$ EMC5 cells, but this is rescued upon re-expression of EMC5. This result indicates that decreases in RFP-TA levels by FACS analysis are indeed a consequence



**5.11 Endogenous SQS levels are reduced in the absence of EMC.** Immunoblotting in the indicated cell lines (see Figure 5.6) of endogenous tubulin and SQS. Blots generated by Norbert Volkmar.

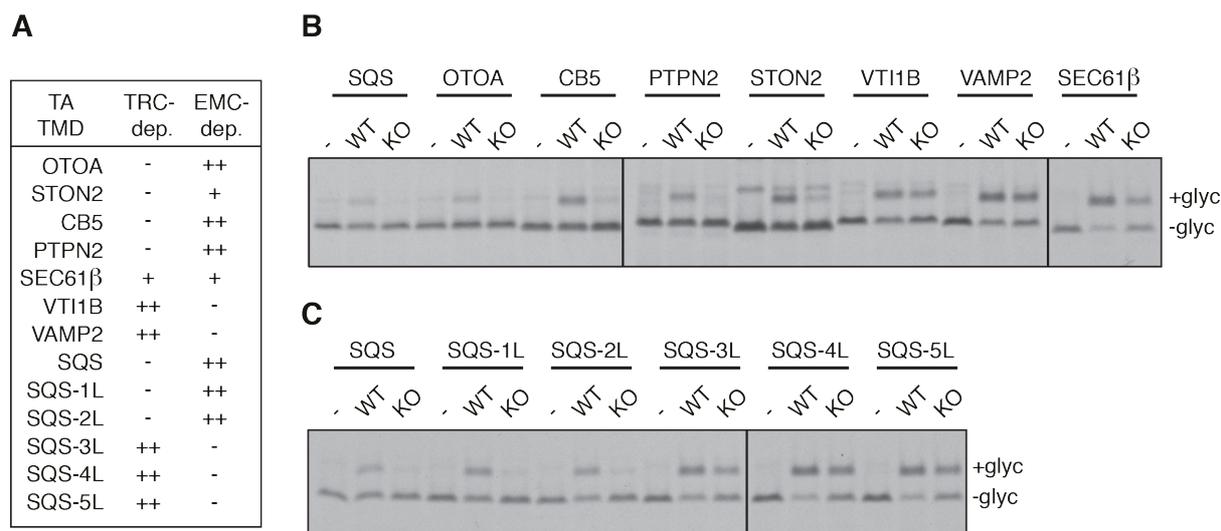
of failed SQS insertion. For VAMP2, the absence of the EMC has no effect on its insertion or subsequent trafficking.

The localization of SQS-GFP and VAMP2-GFP were also visualized in U2OS WT, ΔEMC6 and ΔEMC6 rescue cells after fixation (Figure 5.10). Note that these constructs are different than those used in the FACS and immunoblot analysis. In low expressing ΔEMC6 cells, the localization of SQS-GFP is diffusely cytosolic instead of being localized to the ER as in WT cells (Figure 5.10A). In higher expressing ΔEMC6 cells SQS-GFP is seen to form punctae in the cytosol, presumably representing aggregates. The correct localization of SQS to the ER was restored in ΔEMC6 rescue cells. VAMP2-GFP showed no obvious differences in localization between the cells lines across a wide range of expression levels (Figure 5.10B). In addition, the levels of endogenous SQS in the absence of EMC were also analysed by immunoblotting (Figure 5.11). In accordance with the behaviour of the SQS TMD in our artificial construct, minimal SQS expression was detected in ΔEMC5 and ΔEMC6 cells. This can probably be attributed to the failure of endogenous SQS to insert into the ER, leading to its degradation. Cumulatively, these results show that SQS insertion into the ER membrane is dependent on an intact EMC, the absence of which causes mislocalization, degradation, and aggregation.

**5.5 Less hydrophobic TA proteins need EMC to insert**

The dependency of SQS on the EMC both *in vivo* and *in vitro* is strong evidence that this complex is necessary for mediating the insertion of the SQS TMD into the lipid bilayer. If EMC is a general insertase, this ability should extend to other low-to-moderate hydrophobicity substrates that cannot make use of the TRC pathway. To test this, the full panel of endogenous

## Chapter 5: The EMC is essential for TA protein insertion



**5.12 Analysis of TA proteins for EMC-dependent insertion.** (A) Summary of dependence on either TRC40 (as judged by inhibitory effect of WRB-CC) or EMC for the indicated substrates. (B) Gels of the insertion assays tabulated in (A). Note that in some samples there is a product which migrates slightly slower than the glycosylated product. This represents a mono-ubiquitinated non-inserted product. It is most prominently observed in the case of STON2.

TA substrates (Figure 3.3) were analysed for *in vitro* post-translational insertion into WT or  $\Delta$ EMC6 HEK293 hRM (Figures 5.12A, 5.12B). The proteins capable of interacting with TRC40 and sensitive to TRC40 pathway inhibitors were largely unaffected in the absence of EMC. However, TRC-independent substrates were almost entirely dependent on EMC and were incapable of inserting without it present in membranes. Sec61 $\beta$ , a protein of moderate hydrophobicity, showed only partial dependence on EMC. This is consistent with earlier results showing that Sec61 $\beta$  can use the TRC pathway with some success (Figure 3.6). Note that all of the proteins become ubiquitinated when they are not inserted in the membrane (Hessa et al., 2011; Itakura et al., 2016). Furthermore, the SQS mutants with serial leucine mutations were also analysed for insertion in WT or  $\Delta$ EMC6 HEK293 hRM (Figures 5.12A, 5.12C). As for native TA proteins, those mutants which are capable of using the TRC pathway were unaffected in the  $\Delta$ EMC6 cells. However, TRC-independent substrates were EMC-dependent as evidenced by their inability to insert into  $\Delta$ EMC6 compared to WT cells. These results show that the EMC is essential for inserting all non TRC- dependent TA substrates.

## 5.6 Discussion

The insertion of low-to-moderate hydrophobicity TMD containing TA proteins into the ER membrane is a protein mediated process. The ER membrane complex (EMC), a widely

## Chapter 5: The EMC is essential for TA protein insertion

conserved and highly expressed ten subunit complex of previously uncharacterized biochemical function was shown to be necessary for this insertion reaction in mammalian cells. Both *in vivo* and *in vitro*, the EMC is specifically involved in the insertion of non TRC-dependent TA proteins but seems to have no effect on substrates that have been previously established to use known translocation systems. Indeed, our results suggest that the TRC and EMC-dependent pathways are mostly tuned for TMDs of high and low hydrophobicity, respectively.

Previously, it has been shown in completely reconstituted systems that certain TA proteins such as Cb5 can integrate into the lipid bilayer autonomously, without the assistance of additional protein machinery (Brambillasca et al., 2005). We investigated this as a possible option for the insertion of SQS, a well characterized non TRC-dependent TA protein whose TMD shares some commonalities with that of Cb5. Human derived microsomes treated with trypsin were incapable of SQS insertion. This result suggested that a protein component must be somehow involved in insertion, as trypsin compromises exposed ER proteins but leaves the lipid bilayer completely intact. Note that Cb5 insertion was also later shown to be EMC-dependent (Figure 5.12). This result is in stark contradiction to the prediction of the ‘unassisted’ model (Borgese and Fasana, 2011). Therefore, it is highly likely that previously observed ‘spontaneous’ insertion of Cb5 was a consequence of the methods used to create artificial membranes. This discrepancy is addressed experimentally in the next chapter.

The involvement of a proteinaceous component prompted the search for what this factor may be. Previously implicated translocation machinery such as the Sec61 translocon and the Sec62-Sec63 complex were ruled out. The EMC was considered as a putative insertase based on previous literature genetically implicating it in a wide range of membrane related processes and our siRNA screen with multi-pass IMPs (Christianson et al., 2012; Jonikas et al., 2009; Lahiri et al., 2014; Louie et al., 2012; Satoh et al., 2015). Initial *in vitro* observations using EMC5 knockdown cells suggesting that SQS insertion is specifically impaired in the absence of the EMC were further pursued. Though the WRB/CAML complex was not tested directly, the fact that EMC-null microsomes are inactive for SQS insertion, but competent for VAMP2 insertion, strongly argues that it cannot substitute as an insertase.

Insertion of SQS both *in vitro* and *in vivo* was selectively impaired in  $\Delta$ EMC5 and  $\Delta$ EMC6 cells. Interestingly, deletion of either of these two subunits lead to the disintegration of the entire complex. The same was the case for the cytosolic subunit EMC2 (data not shown). Re-expression of the respective knocked out gene restored the remaining nine subunits.

## Chapter 5: The EMC is essential for TA protein insertion

Clearly, EMC2, EMC5 and EMC6 are integral for the stability of the entire complex. This information may be critical for future biochemical and structural analysis of the EMC. Here, it was exploited as an easy means of disrupting the entire EMC for the purposes of testing TA insertion. Note that, in Fig 5.7A, there is still a noticeable amount of SQS insertion into  $\Delta$ EMC5 and  $\Delta$ EMC6 cells. This residual insertion has three possible explanations: (i) SQS can use the TRC pathway at some level when no other option is available; (ii) the residual levels of other EMC subunits might permit insertion depending on which subunits are the critical ones for the insertion reaction; (iii) a yet unidentified pathway. Of these, the first is most likely as evidenced by the complete lack of insertion in EMC-null microsomes when purified SQS-CaM complexes are used (i.e., when alternate pathways are not an option); see Figure 5.7B.

After establishing the importance of EMC for the insertion of SQS, we analysed all previously tested TA proteins (described in Chapter 3). All TRC-independent TA proteins were EMC-dependent and vice versa. Sec61 $\beta$ , a protein of moderate hydrophobicity, showed partial dependence on both EMC and TRC40, identifying the approximate point of overlap between these two pathways. These results suggest that hydrophobicity is the key parameter governing which post-translational pathway is utilized, though other features such as TMD length such as TMD length or helicity may also influence pathway choice. The lower hydrophobicity of clients for the EMC pathway presumably explains why a dedicated targeting pathway with constant TMD shielding is not needed, instead relying on temporary release from general TMD binding proteins to engage the membrane.

Interestingly, FACS analysis indicated that there was a slight effect on VAMP2 insertion in the absence of EMC, with more of it being degraded relative to wild type cells. This likely has a simple explanation: the absence of the EMC pathway likely results in many abundant dual-pathway substrates (such as Sec61 $\beta$ ) being forced to use the TRC pathway. This increased flux probably results in some degree of TRC pathway saturation, being observed as an indirect but small effect on VAMP2 insertion.

The results in this chapter tentatively suggest that the EMC is an insertase for specific TA substrates. However, the EMC appears to be important for many processes as its absence has been demonstrated to cause pleiotropic effects. As such, it is possible that the loss of EMC indirectly affects insertion, by some function other than acting as an insertase. For example, it could be the chaperone of an insertase, since it has been shown to be important for the stabilization of several membrane proteins (Richard et al., 2013; Satoh et al., 2015). Alternatively, it can affect lipid homeostasis in a way that impacts insertion (Lahiri et al., 2014).

## Chapter 5: The EMC is essential for TA protein insertion

This is one explanation that could reconcile earlier ‘unassisted’ models with the dependency on the EMC. Namely, we have shown that SQS is impaired in the absence of EMC. SQS is responsible for catalysing the first committed step in sterol synthesis. Impairing this enzyme would presumably lead to defects in the production of cholesterol, which would translate to changes in its amount in the ER membrane. This would affect the properties of the lipid bilayer and could thereby debilitate the insertion of low hydrophobicity TA proteins as in artificial membranes (Brambillasca et al., 2005). Therefore, to make a claim for EMC insertase function, it is important to show direct involvement. In the next chapter, we show that EMC is indeed a TMD insertase by demonstrating that it can independently catalyse insertion in a fully purified system.

### 5.7 Materials and methods

#### *Plasmids, antibodies, siRNA, proteins, and reagents*

All TA constructs with the exception of *in vivo* expression constructs used for microscopy and FACS analysis have been described in Chapters 4 and 5. GFP-SQS used for microscopy was generated by PCR amplification of the C-terminal region containing the TMD and flanking residues (aa 355-417 of SQS) and fused in-frame downstream of GFP. A similar approach was taken to generate the GFP-VAMP2 construct. The dual colour reporter for protein degradation was based on constructs described previously (Itakura et al., 2016), but with the TMD of the desired TA protein appended to the end of the RFP coding region instead of to GFP. Note that the mCherry variant of RFP and the mEGFP variant of GFP were used throughout, but the simpler nomenclature of RFP and GFP are used in the text and figures. Human VCAM1 has been previously described (Garrison et al., 2005). All constructs for expression in cultured cells were in the pcDNA5/FRT/TO vector. EMC5 and EMC6 cDNAs for rescuing knock out cell lines were of human origin.

Antibodies against EMC subunits were from the following sources: EMC1 (kind gift of Espreafico lab); EMC2 (Proteintech #25443-1-AP); EMC3 (Abcam #ab175537); EMC4 (Abcam #ab123719); EMC5 (Abcam #ab174366); EMC6 (Abcam #ab84902), and tubulin (Sigma #T5168). Antibodies against Sec62 and Sec63 were as previously described (Garrison, 2005). Pre-designed and validated Silencer Select siRNA from Thermo Fisher were obtained for EMC5 (s41129), SEC62 (s14188), and SEC63 (s22166) knockdowns.

### *Mammalian and PURE system translation*

Translation reactions in RRL and preparations of SQS complexes in the PURE system are as described in Chapters 2, 3 and 4. The use of native versus nucleated RRL is indicated in the text or figures.

### *Semi-permeabilized cells and cell derived microsomes*

Semi-permeabilized cells used for targeting reactions were prepared from untreated or siRNA-treated HEK293 cells. The cells were first washed with PSB and cooled on ice before incubation with 100 ng/ml digitonin in 100 mM KAc, 50 mM Tris pH 8.0, 10 mM MgAc<sub>2</sub> for 5 min on ice. Cells were washed twice with the above buffer lacking digitonin, collected by centrifugation at 12,000 x g for 15 s on a benchtop centrifuge, and re-suspended in 50 mM KAc, 25 mM HEPES, 1 mM MgAc<sub>2</sub> on ice to a concentration of  $\sim 2.5 \times 10^7$  cells per ml. They were kept on ice and used immediately for insertion assays.

Microsomes from HEK293 cells (hRM) were made as previously described (Zhang et al., 2013). Briefly, cells were washed three times in PSB and collected by spinning at 200 x g for 5 min. Cells were re-suspended in three volumes of ice-cold sucrose buffer (10 mM HEPES, 250 mM sucrose, 2 mM MgCl<sub>2</sub>) with EDTA-free protease inhibitor cocktail (Roche). Cells were mechanically lysed via passage through a 26-gauge needle. The cell lysate was spun at 3,800 x g at 4 °C for 30 min and clarified a second time with the same spin conditions. The resulting post-nuclear supernatant was centrifuged at 4 °C at 75,000 x g for 1 hour in a TLA100.3 rotor. The resulting microsome pellet was re-suspended in microsome buffer (10 mM HEPES, 250 mM sucrose, 1 mM MgCl<sub>2</sub>, 0.5 mM DTT) and adjusted to an A<sub>280</sub> value of 75. Insertion assays typically contained 1 µl of hRMs per 10 µl reaction.

Trypsin-inactivation of hRM was with 100 µg/ml trypsin for 15 min at 32 °C. Trypsin inhibitor was added to 500 µg/ml on ice, and the membranes were re-isolated by centrifugation through a 200 µl 20% sucrose cushion at 55,000 rpm in a TLA-55 rotor. Membranes were re-suspended in microsome buffer with 0.1 mg/ml trypsin inhibitor.

### *Cell culture*

U2OS Flp-In TRex cells (kind gift of Mads Gyrd-Hansen, Oxford, UK) and Flp-In TRex 293

## Chapter 5: The EMC is essential for TA protein insertion

cells (Invitrogen) were maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% foetal bovine serum and 2 mM L-glutamine. EMC5 and EMC6 knockout cell lines were generated using the CRISPR/Cas-9 system using previously described protocols (Ran, 2013). Cells were transfected with a pSpCas9(BB)- 2A-Puro (PX459) plasmid (Addgene) containing sgRNAs targeting EMC5 or EMC6. Cells were then grown for 48 hours and treated with puromycin for 72 hours to select for successfully transfected clones. Single-cell clones were isolated by limiting dilution. Disruption to cell expression was confirmed by blotting of whole cell lysates and additionally by deep sequencing for cells generated in a U2OS background. Multiple independent clones generated with different sgRNAs for each gene were tested for successful knockout and shown to behave similarly in the assays for TA protein insertion. The  $\Delta$ EMC5 and  $\Delta$ EMC6 U2OS cells in the figures were made with sgRNAs GCATCATGGCGCCGTCGCTGTGG and GCCGCCTCGCTGATGAACGGCGG, respectively, while the  $\Delta$ EMC6 HEK293 cells used CCGAGGTCCGGCAATAATCCAGG.

Rescue of the knockout cell lines with the respective expression constructs (in the pcDNA5/FRT/TO vector) was achieved by integration into the FRT locus using the Flp recombinase system as per manufacturer's instructions (Invitrogen). In short,  $1 \times 10^6$  cells in a 6 cm tissue culture plate were co-transfected with 3  $\mu$ g of the rescue plasmid and 1  $\mu$ g of the Flp recombinase expression plasmid (pOG44) using Lipofectamine 2000. Cells stably integrating the gene of interest were selected with 250  $\mu$ g/ml hygromycin B (U2OS cells) or 100  $\mu$ g/ml hygromycin B (293 cells) for 7–10 days. Control rescue cells were generated in parallel using the empty pcDNA5/FRT/TO vector and used as the knockout control for comparison in TA protein assays.

### *Flow cytometry analysis*

Cells growing in a 6 cm tissue culture plate were transfected with 250 ng of either GFP-2A-RFP- SQS or GFP-2A-RFP-VAMP2. 24 hours after transfection, cells were detached with trypsin/EDTA, pelleted and resuspended in ice-cold PSB + 3 mM EDTA, and analysed by flow cytometry using a FACS Canto (BD Biosciences, Franklin Lakes, NJ). The flow cytometry and data analysis (using the FlowJo software package) were essentially as previously described (Itakura et al., 2016).

### *Microscopy*

The localization of SQS was visualized in live cells using an expression construct for GFP-SQS in the pcDNA/FRT/TO vector. After transient transfection and induction of expression for 24 h (100 ng/ml doxycycline), the live cells were visualized at an excitation wavelength of 488 nm using an Olympus CKX41 microscope. The localization of VAMP2 was visualized using confocal microscopy to better discriminate cell surface localization from any potential cytosolic population. Cells growing on 12 mm glass coverslips (Nunc) were transfected with the GFP-2A- RFP-VAMP2 construct. After allowing expression for 24 hours, the cells were fixed with 4% paraformaldehyde (PFA) and mounted in Fluoromount G (Southern Biotech). Imaging was performed using an LSM 710 confocal microscope (Zeiss).

## **Chapter 6: Reconstitution of EMC-dependent tail-anchored protein insertion with purified factors**

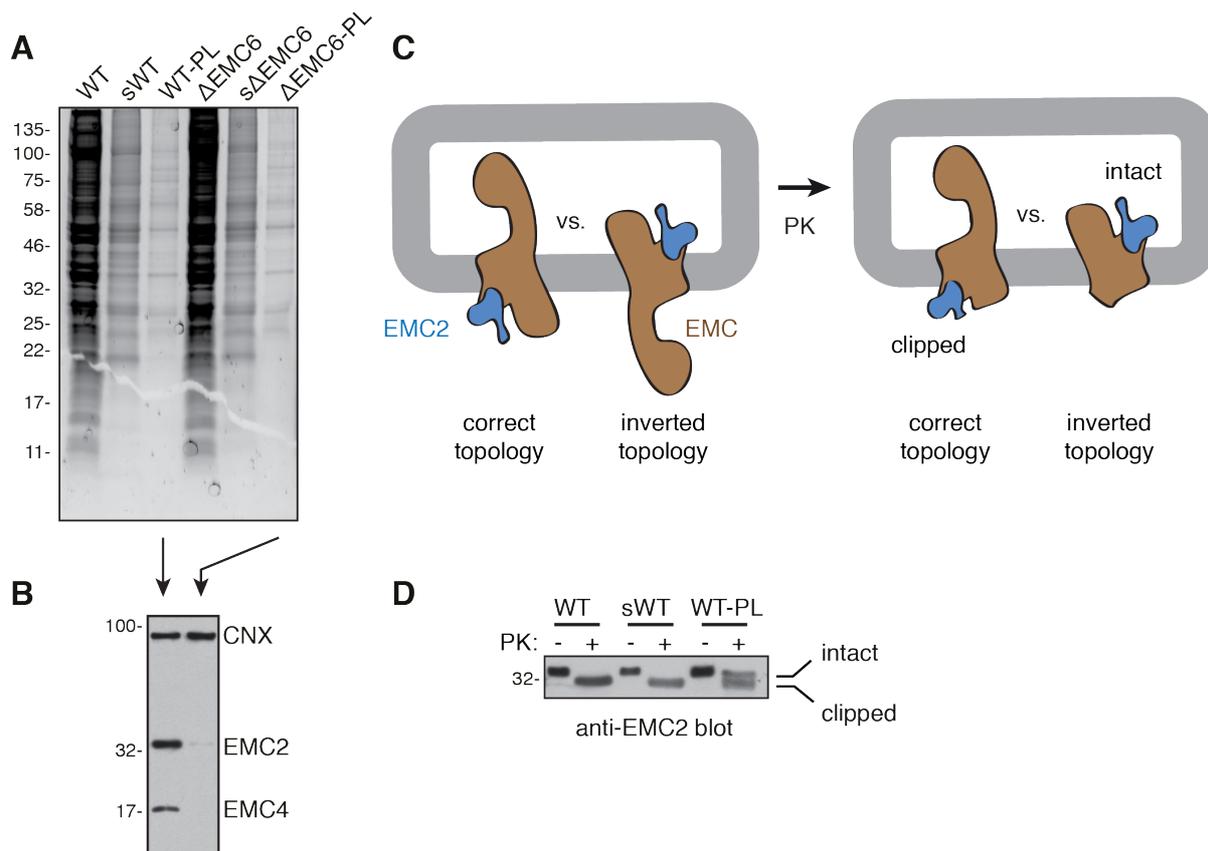
The insertion of low hydrophobicity transmembrane domain (TMD) containing tail anchored (TA) proteins such as squalene synthase (SQS) into the endoplasmic reticulum (ER) depends on the ER membrane protein complex (EMC). The selective inability of non TRC-dependent TA substrates to insert in the absence of EMC implies that this factor is somehow involved in mediating their insertion into the lipid bilayer. However, the loss of EMC is known to cause pleiotropic effects, leaving open the possibility that any observed defects in insertion are indirect. For example, loss of EMC could impact the composition of the lipid bilayer such that insertion becomes unfavourable or stabilize another factor which indeed acts as a TMD ‘insertase’. It is therefore important to show direct evidence that EMC is involved in the insertion of TRC-independent substrates such as SQS.

The sufficiency of EMC in post-translationally inserting SQS is provided in this chapter through a series of reconstitution experiments with purified factors. First, we show that EMC dependency is maintained after total proteins from WT and EMC knockout ( $\Delta$ EMC) cell derived microsomes are reconstituted into liposomes. To demonstrate EMC sufficiency, the intact 10-protein complex was purified from mammalian cells and reconstituted into liposomes. The reconstituted EMC remains fully intact, with approximately one-third of the complex oriented correctly. EMC was shown to strongly stimulate the insertion of SQS synthesized in native rabbit reticulocyte lysate (RRL) into liposomes but had minimal activity toward the TRC-dependent substrate VAMP2. Similarly, EMC proteoliposomes supported the insertion of the recombinant SQS-CaM complex to near native levels at matched levels of correctly oriented EMC. As in native microsomes, trypsin treatment of EMC proteoliposomes drastically reduces SQS insertion into the membrane. These experiments cumulatively show that SQS insertion is a protein mediated process that is directly dependent on an intact EMC.

### **6.1 EMC dependency is maintained after reconstitution**

To determine whether EMC is sufficient for TA protein insertion, we found conditions for its reconstitution into liposomes. The incorporation of proteins into synthetic lipids removes the possibility that any observed differences between WT and  $\Delta$ EMC6 HEK293 cell derived rough

## Chapter 6: Reconstitution of EMC-dependent TA protein insertion with purified factors



**6.1 Characterization of EMC in reconstituted proteoliposomes with total protein.** (A) Protein profiles at different stages of reconstitution of total protein derived from wild type and  $\Delta$ EMC6 microsomes. The starting microsomes (WT,  $\Delta$ EMC6), solubilized proteins (sWT, s $\Delta$ EMC6), and reconstituted proteins (WT-PL,  $\Delta$ EMC6-PL) are shown. (B) Immunoblot of reconstituted WT-PL and  $\Delta$ EMC6-PL to show recovery of specific membrane proteins. (C) Schematic of protease (PK) protection assay used to probe the topology of the reconstituted EMC. (D) PK of intact hRMs (WT), solubilized hRMs (sWT), and reconstituted proteoliposomes (WT-PL) analysed for EMC2.

microsomes (hRM) could be due to changes in the properties of the membrane in the absence of EMC. This is important given previous observations that loss of EMC can affect lipid composition (Lahiri et al., 2014) and that TA proteins with low TMD hydrophobicity, such as Cb5, have been shown to insert into artificial membranes contingent on an unperturbed lipid bilayer (Brambillasca et al., 2005). SQS, a well characterized TRC-independent, EMC-dependent substrate was used as a model TA protein of low TMD hydrophobicity. Its behaviour was compared to that of the TRC-dependent, EMC-independent TA substrate VAMP2.

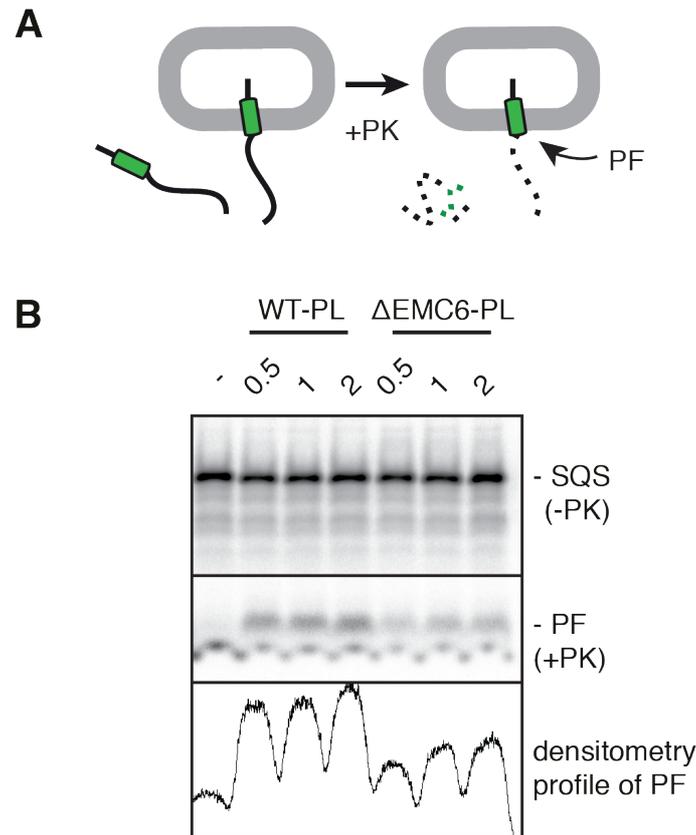
To make reconstituted vesicles, membrane proteins were first solubilized from either WT or  $\Delta$ EMC6 hRM using conditions optimized to recover maximal amounts of EMC. Following reconstitution, protein recovery was assessed by Sypro staining and immunoblotting (Figures 6.1A, 6.1B). The same general protein profile was solubilized and recovered for both WT

## Chapter 6: Reconstitution of EMC-dependent TA protein insertion with purified factors

proteoliposomes (WT-PL) and  $\Delta$ EMC6 proteoliposomes ( $\Delta$ EMC6-PL) as judged by Sypro staining. Immunoblot analysis of WT-PL and  $\Delta$ EMC6-PL shows equal recovery of calnexin (CNX), a control ER-resident protein, and near-absence of EMC subunits in the knockout sample. Upon reconstitution, membrane proteins have no incentive to reintegrate in the membrane in the correct orientation. Therefore, despite recovery of the EMC into synthetic lipids, it is important to establish what proportion of the complex is appropriately oriented relative to the inside of re-formed vesicles. A Proteinase K (PK) protease protection assay was used to probe the topology of the reconstituted EMC (Figure 6.1C). Intact hRM, solubilized hRM (shRM) and reconstituted proteoliposomes (WT-PL) were analysed for EMC2 (Figure 6.1D). EMC2 is ordinarily a peripheral membrane protein entirely on the cytosolic side of the ER. Thus, it is accessible to cytosolic PK, resulting in digestion to a stable core. The same digestion pattern is seen in shRM. After reconstitution, approximately half of EMC2 is now fully protected from protease, indicating EMC incorporation in the opposite orientation. The other half is PK accessible as in hRMs, suggesting integration in the correct orientation. The strategy of comparing levels of intact versus clipped EMC2 after PK digest was repeatedly used to indicate the proportion of correctly oriented EMC following reconstitution.

Reconstituted proteoliposomes do not have functional glycosylation machinery, removing this as a viable option for assessing TA substrate insertion. The post-translational insertion of SQS into WT-PL and  $\Delta$ EMC6-PL was therefore probed using a protease protection assay (Figure 6.2A). For this, both TA constructs had a 3F4 tag appended to their C-terminus, a slight modification to the standard TA glycosylation cassette previously described (Figure 3.2A). The 3F4 tag contains two methionine residues, allowing the inserted, protected fragment to be visualized by autoradiography following protease digestion. To this end, SQS-CaM complexes were prepared in the PURE system (as described in detail in Chapter 3) and presented to reconstituted vesicles. SQS release from CaM was triggered by the addition of the calcium chelator EGTA.

SQS insertion was seen to be consistently lower in  $\Delta$ EMC6-PL compared to WT-PL, indicating that EMC-dependent stimulation of SQS can be maintained after solubilisation and reconstitution (Figure 6.2B). It is worth noting that after reconstitution from a detergent-lipid mixture, insertion into vesicles lacking EMC is higher than seen in hRMs that were never solubilized (i.e. there is still appreciable insertion in lanes 5 through 7 in Figure 6.2B compared

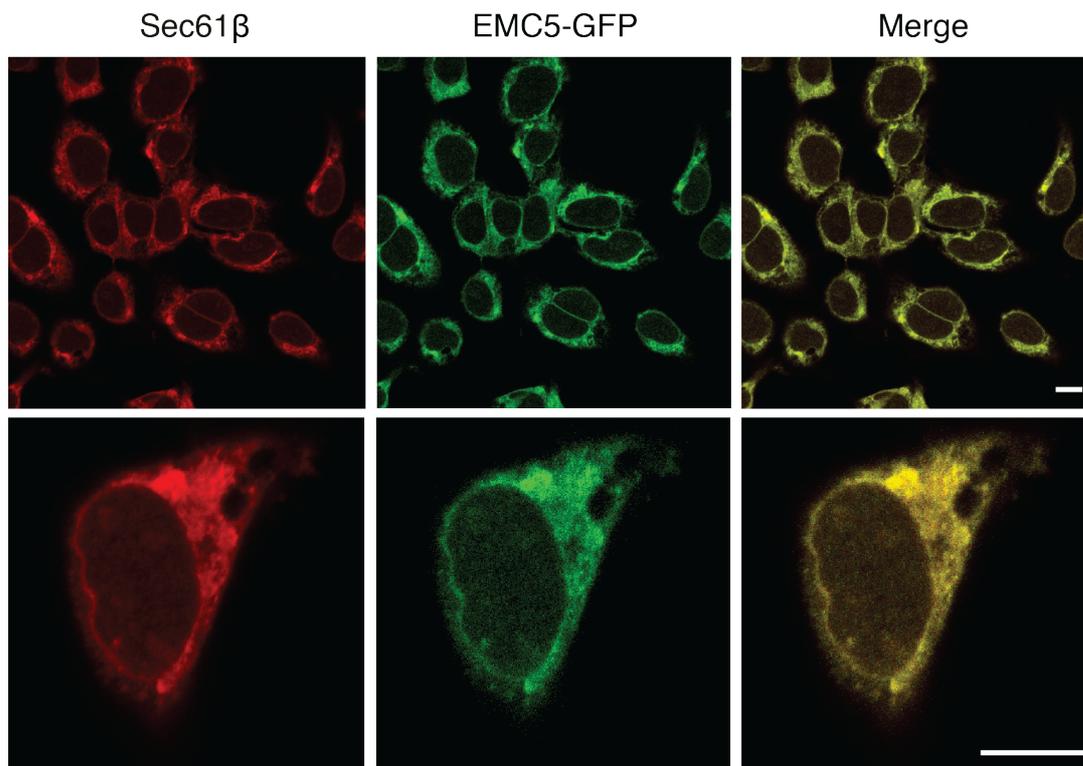


**6.2 SQS insertion into  $\Delta$ EMC6-PL** (A) Schematic of protease protection (PK) assay used to analyse SQS insertion into proteoliposomes. Correct insertion into the membrane should result in a protected fragment (PF). (B) SQS insertion into proteoliposomes reconstituted from total wild type (WT-PL) and  $\Delta$ EMC6 hRMs ( $\Delta$ EMC6-PL).

to lanes 6 through 9 in Figure 5.7B). There are multiple reasons why this may be the case. Primarily, the physical properties of the artificial liposomal membrane (composed only of PC and PE in a 3:1 ratio) may be more conducive to spontaneous insertion. Secondly, the isolation of hRM is not pure in terms of selectively separating the ER from other internal membranes such as mitochondria or the nuclear envelope. Thus, the resident proteins of these other membranes are also solubilized and likely also reintegrate into synthetic lipids upon reconstitution, making it difficult to accurately recapitulate the protein profile of the native ER. Although this makes the stimulatory effect of EMC more difficult to observe above background, the effect remains a substantial  $\sim$ 3-fold difference between WT-PL and  $\Delta$ EMC6-PL. Nevertheless, we conclude that EMC function can be retained after its solubilisation and reconstitution into liposomes, providing a way to test its sufficiency for TA protein insertion.

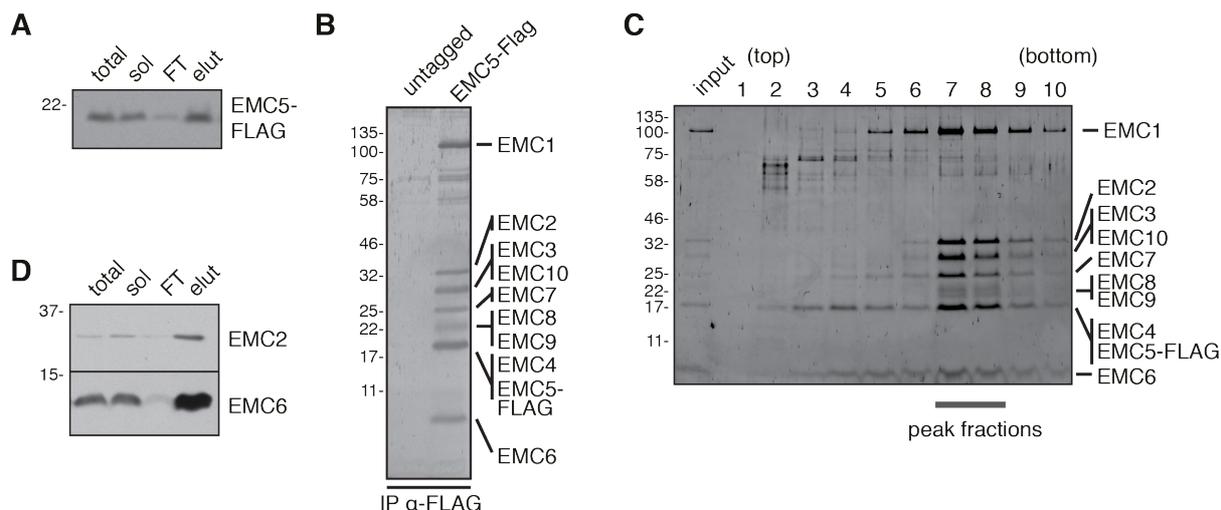
## 6.2 Purification and reconstitution of the mammalian EMC

The mammalian EMC is a ten protein subunit complex composed of three cytosolic subunits and seven integral membrane proteins (Figure 5.5A). The original identification of the EMC as a complex in yeast also demonstrated that it can be isolated by affinity purification, indicating that the subunits are relatively tightly associated (Jonikas et al., 2009). Previous work in yeast also suggested that tagging individual EMC subunits is supported (Christianson et al., 2012; Jonikas et al., 2009; Lahiri et al., 2014). As such, an affinity purification approach was taken to isolate the full ten-subunit EMC from mammalian cells. A 3X-FLAG tag was appended onto the C-terminal end of the EMC5 subunit (EMC5-FLAG). This was integrated into a Flp-In TRex 293 cell line under an inducible promoter. A cell line with a GFP appended onto the C-terminus of EMC5 was made simultaneously. Analysis by fluorescence microscopy of the induced EMC5-GFP cell line indicated that C-terminal tagging did not impact the localization of the EMC at the ER (Figure 6.3).



**6.3 Localization of EMC5-GFP.** GFP was conjugated to the C-terminus of EMC5 (EMC5-GFP) to allow for visualization of the EMC complex. Flp-In TRex HEK293 were modified to contain EMC5-GFP at the FRT site under an inducible promoter. Cells were induced with doxycycline and EMC5-GFP localization was compared to the ER membrane marker Sec61 $\beta$  in fixed cells. Scale bar represents 15  $\mu$ M.

## Chapter 6: Reconstitution of EMC-dependent TA protein insertion with purified factors

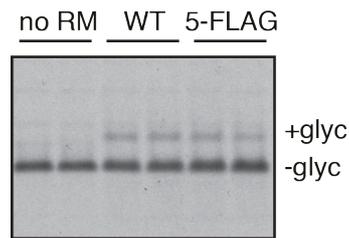


**6.4 Purification of EMC from mammalian cells.** (A) Cells stably expressing EMC5-FLAG were solubilized (sol) in digitonin and subjected to affinity purification using anti-FLAG resin. The flow-through (FT) and elution (elut) fractions show efficient depletion and recovery of EMC5-FLAG, respectively. (B) Larger scale elution of (B) from HEK293 cells expressing untagged or FLAG-tagged EMC5 stained with Sypro. (C) Sample from (B) subjected to analysis by sucrose gradient. All of the subunits of the EMC were verified by mass spectrometry. Keratin contamination is seen in many of the lanes in the ~60-70 kDa region. (D) Immunoblotting for EMC2 and EMC6 from an EMC5-FLAG purification.

To isolate the EMC, cells stably over-expressing EMC5-FLAG for at least three generations were solubilized in digitonin, subjected to affinity purification using anti-FLAG resin followed by elution using a FLAG peptide. The EMC is fully solubilized by digitonin, mostly depleted by the anti-FLAG resin, and recovered in the peptide elution (Figure 6.4A). This initial promising observation prompted a large-scale attempt at affinity purifying the entire complex (Figure 6.4B). The resulting elution was subjected to analysis by sucrose gradient followed by SDS-PAGE and Sypro staining (Figure 6.4C). All the subunits of the EMC (verified by mass spectrometry and immunoblotting, data not shown) were seen to co-sediment in the same fractions, and very little free EMC5 is observed. This indicates that the complex remains intact throughout the purification process, with all the subunits present at seemingly stoichiometric levels. Additionally, immunoblotting for EMC2 and EMC6 from a EMC5-FLAG purification shows that these subunits are largely depleted by removal of EMC5-FLAG (Figure 6.4D). This suggests that the majority of EMC in these cells contains the over-expressed EMC5-FLAG, having displaced the untagged endogenous EMC5.

In addition to confirming that C-terminal tagging of EMC5 did not alter the localization of the EMC, we wanted to ensure that the complex remained fully functional. For this, hRM

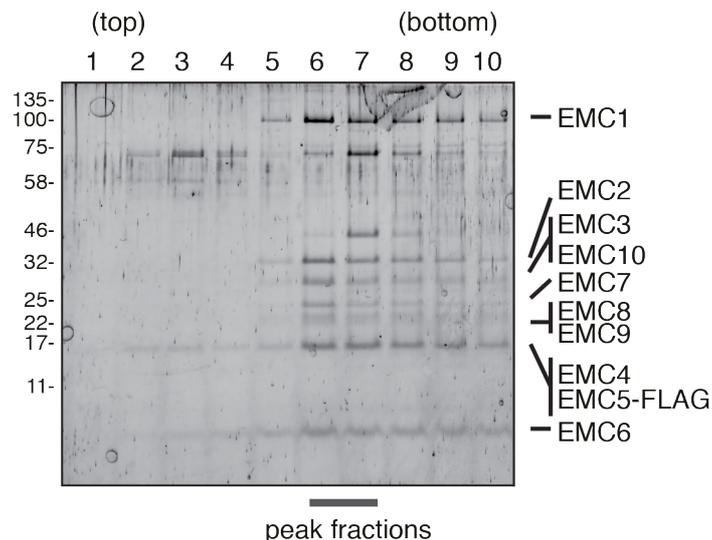
Chapter 6: Reconstitution of EMC-dependent TA protein insertion with purified factors



**Figure 6.5 SQS insertion is intact in EMC5-FLAG cells.** SQS was subjected to insertion assays into microsomes derived from either wild type HEK293 cells (WT) or EMC5-FLAG overexpressing cells (5-FLAG). The duplicate lanes for each condition represent samples without or with anti-FLAG antibody, which proved to have no effect on insertion. Glycosylation is used as a readout for insertion.

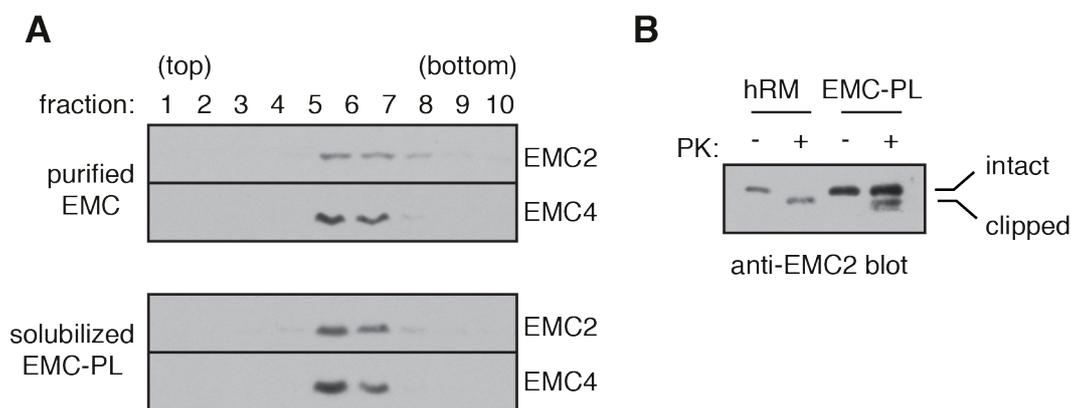
were derived from EMC5-FLAG cells. SQS insertion into EMC5-FLAG hRM was then monitored using the standard post-translational *in vitro* glycosylation-based assay (Figure 3.2B). Relative to WT hRM, EMC-FLAG hRM are unimpaired for SQS insertion, despite nearly all EMC containing the EMC5-FLAG (Figure 6.5). This suggests that the tagged complex remains is functional for TA insertion.

Having determined that the EMC can be purified intact and should be functional in its EMC5-tagged form, we then reintegrated it into synthetic lipids to make EMC proteoliposomes (EMC-PL). EMC-PL can be used to test whether the EMC is sufficient for TA insertion, as any stimulation in insertion relative to empty liposomes (PL) must be attributed to the presence of the complex. For reconstitution into synthetic lipids, EMC was instead purified in the detergent



**6.6 EMC can be purified in DBC.** Purification of the EMC complex from cells stably expressing EMC5-FLAG as in Figure 6.4, instead using the detergent deoxyBigChap (DBC).

## Chapter 6: Reconstitution of EMC-dependent TA protein insertion with purified factors



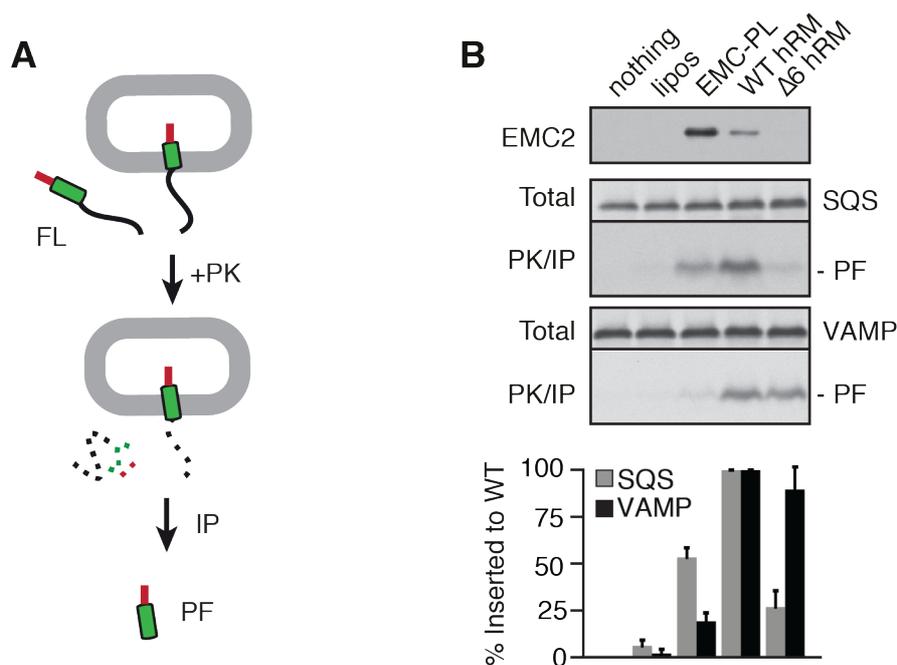
**6.7 Characterization of EMC in purified proteoliposomes.** (A) Purified EMC in DBC was reconstituted in proteoliposomes (EMC-PL). An aliquot of the starting purified EMC and DBC-resolubilised proteoliposomes were separated on a sucrose gradient and immunoblotted for EMC2 and EMC4. (B) Rough microsomes from wild type cells (hWT) and EMC-PL were subjected to the protease protection assay described in Figure 6.1 to determine which proportion of the EMC complex re-integrates into membranes in the correct orientation.

deoxyBigChap (DBC). Solubilisation in DBC still permitted the recovery of the intact EMC but caused the complex to migrate in lighter sucrose gradient fractions due to a different detergent micelle size (Figure 6.6). To ensure that reconstitution didn't perturb the complex, an aliquot of the starting purified EMC resolubilized proteoliposomes were re-solubilized in DBC and separated on a sucrose gradient (Figure 6.7A). The individual subunits were immunoblotted for EMC2 and EMC4. Both subunits co-sediment in the same fractions before and after reconstitution, indicating that the complex in EMC-PL is intact. To gauge what proportion of the EMC reintegrated in the correct orientation, the EMC-PL were PK digested followed by probing of the EMC2 subunit (Figure 6.7B). PK digests on EMC-PL compared to hRMs indicate that only one-third of the EMC is correctly oriented in reconstituted vesicles.

### 6.3 EMC is sufficient for SQS insertion

The ability of SQS to insert into EMC-PL could then be tested to determine whether EMC is sufficient for catalysing TMD insertion. For this, VAMP2 and SQS with a C-terminal 3F4 tag were translated in RRL and post-translationally targeted to EMC-PL as well as WT and  $\Delta$ EMC6 hRM. Following insertion, reactions were treated with PK to determine the amount of TA protein successfully inserted into membranes (Figure 6.8A). To ensure that the resulting protected product was not due to some interaction of the TA protein with the membrane, we added an additional immunoprecipitation step against the C-terminal 3F4 epitope. The 3F4

## Chapter 6: Reconstitution of EMC-dependent TA protein insertion with purified factors

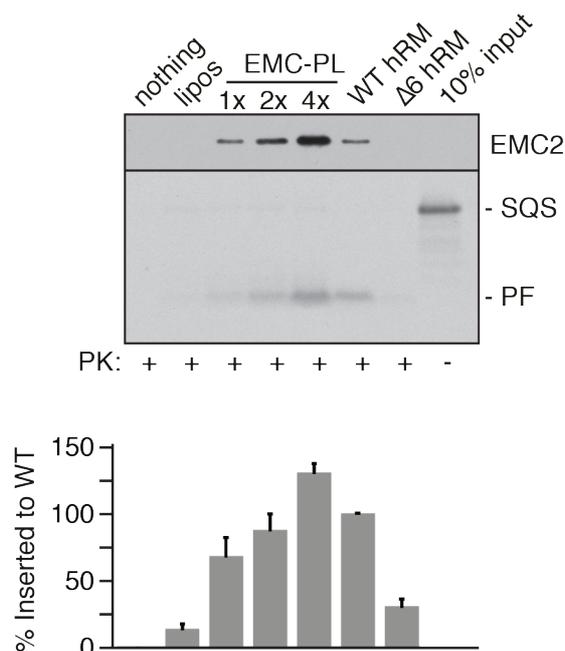


**6.8 EMC is sufficient for SQS, but not VAMP2 insertion into membranes.** (A) Diagram of the protease protection (PK) assay for TA protein insertion using a C-terminal epitope tag (red) to selectively recover the projected fragment (PF), which indicates successful insertion. (B) SQS or VAMP2 were synthesized in native RRL and subjected to insertion assays into liposomes reconstituted with or without purified EMC. Native ER microsomes (hRM) from WT or  $\Delta$ EMC6 HEK293 cells were tested in parallel. Relative levels of EMC are indicated by an EMC2 blot. The graph represents cumulative results from four experiments (mean  $\pm$  SD), normalized to insertion in WT hRM.

antibody requires an intact C-terminus, making it very unlikely that protection from the highly active PK is by any means other than true insertion. This assay has been well documented to be highly specific and indicative of genuine insertion (Mariappan et al., 2010; Stefanovic and Hegde, 2007). Using this assay, SQS was shown to insert into EMC-PL with approximately half the efficiency of that observed in native ER microsomes (Figure 6.8B). SQS insertion into empty liposomes or  $\Delta$ EMC6 hRM was markedly lower. By contrast, insertion of VAMP2 was equally efficient into ER microsomes from both cell types, but poor in EMC-PL. Therefore, EMC-PL are insertion competent for SQS, and display substrate specificity as observed in native microsomes and cells.

EMC-PL also supported the insertion of the recombinant SQS-CaM complex prepared in the PURE system (Figure 6.9). For this, EMC-PL were made with increasing titrations of EMC. These vesicles showed an EMC concentration dependence for insertion of SQS, with near-native levels of insertion relative to hRM observed where the amounts of correctly oriented EMC were matched. As only one third of the EMC is in the correct orientation after

## Chapter 6: Reconstitution of EMC-dependent TA protein insertion with purified factors

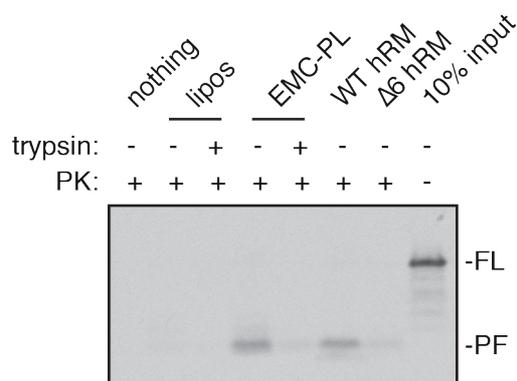


### 6.9 EMC is sufficient to insert the isolated SQS-CaM complex to near-native levels.

Liposomes were reconstituted with varying amounts of purified EMC with a constant amount of lipids (EMC-PL). Relative EMC levels are visualized by immunoblotting for EMC2. The isolated SQS-CaM complex, an aliquot of which is shown in the last lane, was subjected to insertion assays by protease protection in EMC-PL and native ER microsomes from WT (WT hRM) or  $\Delta$ EMC6 HEK293 cells ( $\Delta 6$  hRM). The graph represents cumulative results from four experiments (mean  $\pm$  SD), normalized to insertion in WT hRM.

reconstitution, this means that insertion efficiency is matched when there are between two to four times the amount of total EMC in EMC-PL relative to hRM.

We then wanted to ensure that any observed effects were not due to some artefact of reconstitution. That is, it is possible that SQS insertion is a consequence of perturbations in the lipid bilayer caused by the presence of the EMC, and not necessarily representative of a

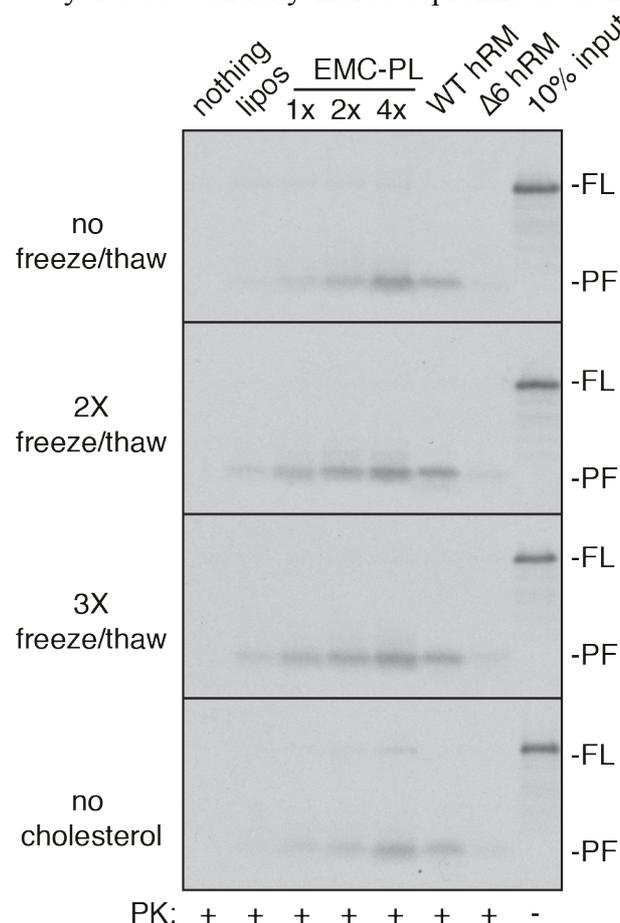


**6.10 SQS insertion depends on an intact EMC.** Empty liposomes or EMC proteoliposomes (EMC-PL) were pre-treated with trypsin and re-isolated before the SQS-CaM complex was assayed for insertion using the protease protection assay. Microsomes from (WT hRM) or  $\Delta$ EMC6 HEK293 cells ( $\Delta 6$  hRM) were used as controls.

## Chapter 6: Reconstitution of EMC-dependent TA protein insertion with purified factors

catalytic activity. To confirm that this is not the case, EMC-PL were treated with trypsin to show that the EMC-stimulated insertion is sensitive to protease digestion just like in biological microsomes (Figure 6.10). As expected, trypsin treatment sharply reduced the insertion of SQS, demonstrating that the insertion activity is indeed protein mediated and contingent on an intact EMC. Thus, the ability of the purified EMC to support SQS insertion into liposomes at near-native efficiencies establishes a direct role for the EMC as an insertase for moderately hydrophobic TMDs into the ER membrane. These observations were repeated with multiple EMC purification preps and reconstitutions.

Note that, in contrast to previous observations (Brambillasca et al., 2005), we did not observe any insertion of SQS into empty liposomes. In early experiments, we found that insertion efficiencies into liposomes and proteoliposomes were variable, although EMC-PLs were consistently more active across experiments. Subsequent characterization led to the finding that freeze/thaw cycles substantially increase promiscuous insertion into liposomes,



**6.11 SQS insertion into empty liposomes is susceptible to freeze/thaw cycles.** Insertion assays as in Figure 6.9 with liposomes and EMC proteoliposomes subjected to multiple freeze/thaw cycles or made without supplemental cholesterol.

## Chapter 6: Reconstitution of EMC-dependent TA protein insertion with purified factors

but not microsomes (Figure 6.11). This suggested that the membrane bilayer of liposomes and PLs were more fragile than the native membrane of microsomes. This fragility could be reduced partially (but not eliminated) by inclusion of cholesterol (at 1.2 mg per 10 mg total phospholipids) in the lipid mixture during reconstitution. Thus, when used immediately after preparation, cholesterol at this concentration had no noticeable impact on insertion activity into either liposomes and PLs; however, the cholesterol-containing vesicles were less variable after a freeze/thaw cycle. Therefore, most assays were either performed on freshly prepared vesicles, or once freeze-thawed vesicles that contained cholesterol. This observation may partially explain why TA substrates such as Cb5 have been shown to integrate into the lipid bilayer ‘unassisted’ (Borgese and Fasana, 2011).

### 6.4 Discussion

In this chapter, the EMC is rigorously established as an insertase for moderately hydrophobic TMDs into the ER membrane. This was accomplished using fully purified components to show that EMC stimulates insertion of SQS into otherwise inactive empty liposomes. Additionally, the purification of the entire ten subunit EMC complex from mammalian cells provides further opportunities to dissect the exact molecular mechanisms by which EMC catalyses TMD insertion. The presentation of the defined SQS-CaM complex, followed by SQS release from CaM and its subsequent insertion into the lipid bilayer by the EMC demonstrates the reconstitution of this TRC-independent pathway in a fully defined system.

Our work does contradict previous observations that TA proteins with low hydrophobicity such as Cb5 can spontaneously integrate into liposomes (Brambillasca et al., 2005, 2006; Colombo et al., 2009). Like SQS, Cb5 is a TRC-independent, EMC-dependent TA protein as shown in Chapters 3 and 5. While Cb5 and SQS both have low hydrophobicity TMDs, there are other notable features that must be considered when comparing any differences in their ability to integrate into empty liposomes. Primarily, SQS has been shown to aggregate in the cytosol in the absence of EMC, and when expressed in the PURE system in the absence of additional chaperones (see Figures 5.10A and Figure 4.5A, respectively). Cb5 and the TA protein PTP1B, which have both been shown as capable of ‘unassisted’ insertion, both remain soluble in a chaperone free system despite sharing a similar TMD hydrophobicity with SQS (Brambillasca et al., 2006). This is likely due to other TMD properties of Cb5,

## Chapter 6: Reconstitution of EMC-dependent TA protein insertion with purified factors

namely a proline in the middle of its TMD, resulting in a ‘kinked’ conformation that may prevent its subsequent aggregation. Though not completely clear as to why Cb5 behaves differently than SQS in totally purified systems despite its clear *in vivo* requirement on EMC for insertion, it is likely that these differences in biophysical properties somehow contribute.

Despite these discrepancies, the key finding is that the EMC is a bona fide TMD insertase. The dependency on EMC is clearly demonstrated in the EMC titration experiment, showing increased SQS incorporation in the membrane in an EMC concentration dependent manner. Clearly, it is this factor in the membrane which facilitates the insertion reaction, rather than stabilizing or supporting some other ER-resident protein that performs this function. The fact that EMC-PL operate with the same efficiency as native microsomes at matched levels of correctly oriented EMC suggests that no additional factors are needed to support the EMC. However, it is difficult at present to formally exclude the involvement of other factors.

Finally, these observations lead to a number of outstanding questions. First, it remains ambiguous as to how EMC actually mediates the insertion reaction at a mechanistic level. Also, some form of TA substrate specificity must be encoded in the complex itself, as selection is not determined at the cytosolic chaperone level (i.e. SQS can be released from either CaM or SGTA). The handing over of substrates from CaM to EMC is also not entirely clear. Seeing as no CaM was seen to interact with the ER membrane (data not shown), it seems unlikely that there is a dedicated docking site for CaM at the EMC. However, the fact that trypsin shaving of EMC-PL essentially abolishes insertion may be due to an essential role of one of the EMC cytosolic subunits in either mediating substrate selection and subsequent transfer to the integral membrane complex components, which presumably mediate the insertion reaction into the lipid bilayer. Finally, in the totally purified system it appears that, in contrast to the TRC pathway, there appears to be no energy requirement for EMC-mediated insertion. Future work directed at addressing these issues will be indispensable for a full understanding of the role of the EMC in TA protein insertion.

## 6.5 Materials and methods

### *Plasmids and antibodies*

Constructs for TA expression in the RRL and PURE systems were as described in Chapters 2 and 3. TA constructs used for protease protection assays were modified to contain a C-terminal

## Chapter 6: Reconstitution of EMC-dependent TA protein insertion with purified factors

3F4 tag which permitted increased radiolabel incorporation and the option to immunoprecipitate the protected fragment to verify its identity (Stefanovic and Hegde, 2007). All constructs for expression in cultured cells were in the pcDNA5/FRT/TO vector. The 3X-FLAG or GFP in the EMC5-FLAG or EMC5-GFP constructs were appended to the EMC5 cDNA of human origin. Antibodies for EMC subunits were as described in Chapter 5.

### *Mammalian and PURE system translation, and cell derived microsomes*

Translations in RRL and preparations of the SQS-CaM complex in the PURE system are as described in Chapters 2 and 3. The preparation of microsomes from HEK293 cells expressing the EMC-FLAG construct were as described in Chapter 3. Prior to harvesting, cells were induced for 72 hours with 100 ng/ml doxycycline.

### *Affinity purification*

The EMC was purified from EMC5-FLAG expressing T-Rex HEK293 cells that had been induced to express EMC5-FLAG with 100 ng/ $\mu$ l doxycycline for at least three cell divisions. Cells were washed in PSB, collected by centrifugation, and placed on ice. The cell pellet was solubilized on ice for 30 min with 1% digitonin in 200 mM NaCl, 50 mM HEPES, 2 mM MgAc<sub>2</sub>. Cell lysate was centrifuged at 4 °C for 10 min at maximum speed in a tabletop microcentrifuge. The supernatant was carefully removed and incubated with FLAG-M2 agarose (Sigma) at 4 °C for 1.5 hours. The resin was washed five times with 0.1% digitonin, 200 mM NaCl, 50 mM HEPES. Elution was with 0.2 mg/ml 3X FLAG peptide in 0.25% digitonin, 150 mM NaCl, 50 mM HEPES for 30 min at 25 °C with mixing. For reconstitution experiments, the EMC was purified in the identical manner, but using deoxy-BigChap (DBC) rather than digitonin. To assess complex integrity, 20  $\mu$ l of the eluted product was layered onto a 200  $\mu$ l 5-25% sucrose gradient with 100 mM NaCl, 50 mM HEPES, 2 mM MgAc<sub>2</sub> and either 0.25% DBC or 0.25% digitonin and spun for 100 min at 55,000 rpm in a TLS-55 rotor with slow acceleration and deceleration. Eleven 20  $\mu$ l fractions were collected and analyzed by SDS-PAGE and Sypro staining or immunoblotting.

## Chapter 6: Reconstitution of EMC-dependent TA protein insertion with purified factors

### *Proteoliposome reconstitutions and insertion assay*

Reconstitution of proteins into liposomes followed minor variations of earlier methods (Mariappan et al., 2011). Phospholipids were obtained from Avanti Polar Lipids and included phosphatidyl-choline (PC) and phosphatidyl-ethanolamine (PE) from bovine liver, and synthetic 1,2-dioleoyl-*sn*-glycero-3- phosphoethanolamine-N-lissamine rhodamine B (Rh-PE). The standard liposome mixture contained PC:PE:Rh-PE at a mass ratio of 8:1.9:0.1. Rh-PE was used for quantification and to monitor recovery throughout the reconstitution procedure. Lipids were mixed in the above ratios as chloroform stocks, adjusted to 10 mM DTT and dried by centrifugation under vacuum for 16 hours (SpeedVac, Eppendorf). Lipid films were rehydrated to a final concentration of 20 mg/ml in lipid buffer (15% glycerol, 50 mM HEPES pH 7.4) and mixed end over end for 8 hours at 25 °C with occasional vortexing until the mixture was homogeneous. The lipids were diluted with additional lipid buffer and supplemented with deoxy-BigCHAP (DBC) to produce a lipid/DBC mixture containing 2% DBC and 10 mg/ml lipids.

BioBeads-SM2 (BioRad) were prepared by first wetting them with methanol, then washing extensively with distilled water. After all traces of methanol were removed, the beads were adjusted with water so that the settled beads occupied 50% of the total volume. For reconstitutions, the BioBeads were dispensed from this 50% slurry in the desired amount, with the excess liquid removed by aspiration just before use. All indicated volumes of BioBeads refer to the packed volume of beads. Reconstitutions used purified EMC in 0.25% DBC obtained as described above.

Proteoliposome reconstitutions with total proteins extracted from RMs required optimization to ensure EMC solubilisation and protein re-integration into synthetic lipids. hRMs prepared from WT or  $\Delta$ EMC6 cells were solubilized on ice for 15 min in 0.8% DBC, 350 mM KAc, 10 mM Tris pH 7.5, 2 mM MgCl<sub>2</sub>. This was then centrifuged at 100,000 rpm for 20 min at 4 °C in a TLA120.1 rotor to remove insoluble material, and the supernatant containing the extracted proteins was reserved. For a standard reaction, 150  $\mu$ l of the RM extract was mixed with 12.75  $\mu$ l lipid/DBC mixture and added to 260  $\mu$ l BioBeads. The slurry was gently mixed for at least 18-19 hours at 4 °C. The fluid phase was then removed and incubated with the same amount of fresh BioBeads for an additional 2 hours at 25 °C with mixing. The fluid phase was separated and diluted with five volumes of ice-cold water. The proteoliposomes were then sedimented in a TLA100.3 rotor at 70,000 rpm for 30 min and

## Chapter 6: Reconstitution of EMC-dependent TA protein insertion with purified factors

resuspended in 15  $\mu$ l liposome resuspension buffer (100 mM KAc, 50 mM HEPES pH 7.4, 2 mM MgAc<sub>2</sub>, 250 mM sucrose, 1 mM DTT).

In initial experiments, we determined the relative concentration of purified EMC compared to the amount in native hRMs from HEK293 cells. This relative concentration was used to gauge the amount needed in the reconstitution reaction to achieve a range of final EMC levels in liposomes spanning the level found in hRMs. Serial dilutions of purified EMC were mixed with a constant amount of lipids and were adjusted so the final buffer concentration was 100 mM NaCl, 25 mM HEPES, 2 mM MgCl<sub>2</sub>, 0.8% DBC. Control reactions contained the same buffer and detergent conditions but lacked protein. A standard 100  $\mu$ l reaction contained 10-40  $\mu$ l purified EMC, 30  $\mu$ l of the 10 mg/ml lipid/DBC mixture, and the remaining volume made up with buffer, salts, and detergent. This protein/lipid/detergent mixture was added to 40  $\mu$ l BioBeads in round bottom 2 ml tubes. The slurry was gently mixed in a thermomixer for 18 hours at 4 °C. The fluid phase was then removed separated and diluted with five volumes of ice-cold water. The proteoliposomes were then sedimented in a TLA120.2 rotor at 70,000 rpm for 30 min and resuspended in 25  $\mu$ l liposome resuspension buffer (100 mM KAc, 50 mM HEPES pH 7.4, 2 mM MgAc<sub>2</sub>, 250 mM sucrose, 1 mM DTT).

Insertion assays contained 5.5  $\mu$ l of purified CaM-SQS complex, 2  $\mu$ l of hRM, 4  $\mu$ l liposomes, or 4  $\mu$ l proteoliposomes in a final volume to 10  $\mu$ l. The reaction was initiated by addition of EGTA to 0.5 mM to chelate excess Ca<sup>2+</sup> and trigger SQS release from CaM. The insertion reaction proceeded at 32 °C for 20 min unless otherwise stated. The samples were transferred to ice and treated with 0.5 mg/ml Proteinase K for 1 hour. The digestion was terminated with 5 mM PMSF for 5 min on ice before being transferred to a 5-fold volume of 1% SDS, 0.1 M Tris pH 8 pre-heated to 100 °C. An additional affinity purification step for the protected TA fragment was performed immediately after. Samples were diluted 10-fold in IP buffer (100 mM NaCl, 50 mM HEPES, pH 7.6, 1% Triton X-100), chilled on ice, and incubated with 2.5  $\mu$ l 3F4 rabbit polyclonal antiserum and 10  $\mu$ l Protein A resin for 2 hours at 4 °C. The resin was washed with IP buffer three times and eluted with SDS-PAGE sample buffer before analysis by SDS-PAGE and autoradiography. The same protease protection conditions were used to assess EMC orientation using immunoblotting against EMC2.

## Chapter 6: Reconstitution of EMC-dependent TA protein insertion with purified factors

### *Cell culture*

The cell line expressing EMC5-FLAG was grown and prepared in the same way as the rescue U2OS cells described in Chapter 5. Briefly, this was done by integration of the EMC5-FLAG construct into the FRT locus using the Flp recombinase system into Flp-In TRex 293 cells as per manufacturer's instructions followed by selection (Invitrogen).

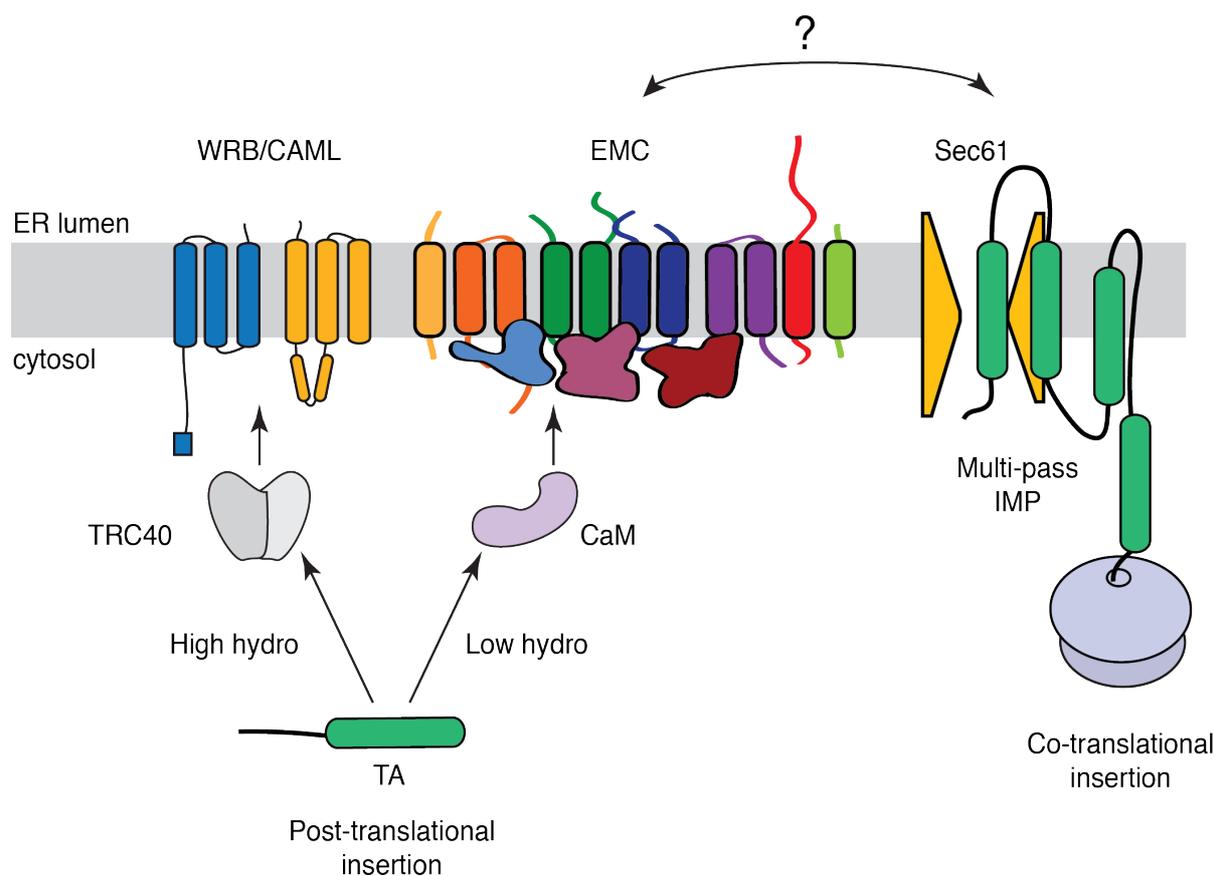
### *Microscopy*

The localization of EMC5 was visualized in Flp-In TRex 293 cells stably expressing EMC5-FLAG. Cells were grown on 12-mm (Nunc) polylysine coated cover-slips and induced with 10 ng/ $\mu$ l doxycycline for 72 hours prior to being fixed in 3.6% formaldehyde. Sec61 $\beta$  was used as a colocalization marker for the ER. For this, fixed cells were permeabilized in 0.1% Triton-X and incubated with rabbit Sec61 antibody at 1:100. Following extensive washing, cells were then incubated for one hour with the far-red dye Alexa 564 at 1:500 and then mounted in SlowFade Gold (Thermo Fisher Scientific). Imaging was performed using a LSM 780 Confocal microscope (Zeiss).

## Chapter 7: Perspectives and future work

This work has demonstrated that the ER membrane complex (EMC) is a transmembrane domain (TMD) insertase. The biochemical function of EMC was revealed by investigating a subset of tail-anchored (TA) proteins of low-to-moderate TMD hydrophobicity which, by virtue of their inability to stably interact with TRC40/Get3, have limited access to the TRC/GET pathway. Instead, these TA proteins are shielded in the cytosol by calmodulin (CaM) and are subsequently inserted into the lipid bilayer by the EMC, which is both necessary and sufficient for this process. This biochemical insight may prove to be essential in determining the role of EMC in the biosynthesis of multi-pass integral membrane proteins (IMPs). Indeed, in light of EMC's insertase function, it is likely that TMDs within multi-pass IMPs require it in addition to the Sec61 translocon for their correct insertion and assembly at the ER membrane (Figure 7.1). The mechanistic details of how EMC is involved in the biogenesis of the beta-1 adrenergic receptor ( $\beta$ 1-AR) and that of other multi-pass IMPs remains the subject of future work.

The EMC is therefore the third widely conserved membrane protein insertase at the eukaryotic ER (Wideman, 2015), joining the Sec61 translocon which is primarily used for co-translational insertion (Heinrich and Rapoport, 2003), and the WRB/CAML complex used by post-translational substrates delivered by the TRC pathway (Wang et al., 2014). The use of multiple insertase systems likely reflects the topologic and biophysical diversity of IMPs and the varying features of their individual TMDs. The biological consequences of this diversity are evident when considering the succession of observations that led to the discovery of insertion mechanisms used by TA proteins. It has long been known that the co-translational pathway is not a viable route for many substrates such as TA and small secretory proteins as their late emergence from the ribosomal exit tunnel precludes their ability to be recognized by SRP (Kutay et al., 1993, 1995). This observation led to the discovery of the post-translational TRC system (Stefanovic and Hegde, 2007). However, the work presented in this thesis demonstrates that the TRC pathway is constrained by the hydrophobicity threshold needed for successful interaction with TRC40. This limitation is compensated by the EMC. The EMC seems capable of inserting low hydrophobicity TMD containing TA substrates as they are transiently released from general cytosolic binding proteins. A considerable number of TA



**7.1 The role of EMC in post- and co-translational insertion.** Multiple routes exist for the delivery and insertion of IMPs into the ER membrane. For TA proteins, hydrophobicity determines whether substrates use the TRC or EMC pathway. Low hydrophobicity substrates are recognized in the cytosol by TRC40 which targets them to the ER via interactions with the WRB/CAML complex, which also facilitates insertion. Low hydrophobicity substrates are kept soluble in the cytosol by calmodulin (CaM), a general TMD binding factor. The ER membrane complex mediates insertion into the lipid bilayer as TA proteins are dynamically released from CaM. EMC may also have a role in the biosynthesis of multi-pass IMPs, perhaps working in conjunction with the Sec61 translocon to mediate their insertion or folding.

proteins therefore depend on the EMC, with certain TA proteins of intermediate hydrophobicity likely capable of using both TRC and EMC-dependent routes into the ER membrane.

The essential role of the EMC for the insertion of roughly half of TA proteins likely explains a number of its previously reported phenotypes. It is easy to rationalize why deficiencies in the biogenesis of EMC TA clients might indirectly cause ER stress (Jonikas et al., 2009), aberrant intracellular trafficking or degradation of membrane proteins (Christianson et al., 2012; Louie et al., 2012; Richard et al., 2013; Satoh et al., 2015), problems with lipid homeostasis (Lahiri et al., 2014) or altered viral replication (Savidis et al., 2016). The molecular explanation of these phenotypes should now be amenable to dissection by finding

## Chapter 7: Perspectives and future work

the particular EMC client(s) necessary for a respective process. Likewise, it is possible that these phenotypic outcomes are a result of the EMC's involvement in the biogenesis of multi-pass IMPs. Determining the underlying cause of EMC's reported phenotypes in cells and organisms will depend on future *in vitro* and *in vivo* work.

The observations presented in this thesis also call into question of the idea of 'spontaneous insertion', whereby certain TA proteins of low hydrophobicity such as cytochrome b5 (Cb5) seem to insert into the lipid bilayer without the need of a proteinaceous factor (Brambillasca et al., 2005; Ridder et al., 2002). In light of Cb5's *in vivo* insertion dependency on EMC, it is likely that these previous findings are a consequence of the exclusive use of highly artificial *in vitro* systems to interrogate the insertion process. However, there does remain the possibility that in the absence of the EMC, the physico-chemical properties of the ER membrane are changed such that Cb5's 'spontaneous' insertion is no longer permitted. In the future, it will be essential to explicitly demonstrate whether the EMC directly mediates the insertion of the model substrate Cb5.

As mentioned above, the insertase activity of the EMC could also be exploited for co-translationally delivered substrates. Specifically, its ability to handle and insert sub-optimal TMDs could make it particularly important for the insertion of multi-pass IMPs. Our work with  $\beta$ 1-AR, in addition to the exiting body of literature, shows that numerous multi-pass IMPs have poorly hydrophobic TMDs, some of which might be integrated post-translationally (Lu et al., 2000). Therefore, it is possible that the EMC may have dual functions: independently inserting post-translationally delivered TA proteins or acting as an insertase in conjunction with the Sec61 translocon for co-translational substrates.

If true, an insertase that can act independently as well as in collaboration with Sec61 would be a novel discovery at the eukaryotic ER. However, a complex that functions in this capacity has already been described in the bacterial plasma membrane. Indeed, YidC has been rigorously shown to act as an insertase in isolation (Serek et al., 2004), and as a chaperone in conjunction with the SecYEG translocon (Zhu et al., 2013). To perform this function, YidC has five TMDs arranged to form a partially hydrophilic groove which is open towards both the lipid bilayer and the cytosol (Kumazaki et al., 2014). Crosslinking studies suggest that this groove acts as a binding site for TMDs, which can subsequently be released into the lipid bilayer (Klenner et al., 2008). Furthermore, YidC is thought to use this TMD-binding activity when collaborating with the SecYEG translocon (Klenner et al., 2008; Nagamori et al., 2004; Serdiuk et al., 2016; Zhu et al., 2013). The absence of a membrane-spanning channel within YidC is consistent with its inability to handle membrane proteins that require soluble domain

## Chapter 7: Perspectives and future work

translocation across the lipid bilayer. Of note, most TA proteins which utilize EMC also have relatively short soluble luminal domains.

YidC is part of a highly evolutionarily related family of non-Sec insertases that includes Oxa1 and Alb3. Oxa1 resides in the topologically equivalent mitochondrial inner membrane (Hell et al., 2001) while Alb3 is found in the chloroplast membrane (Moore et al., 2000). Recent analysis demonstrates that the YidC/Oxa1/Alb3 family may actually be highly related, and possibly share a common ancestor, with eukaryotic ER membrane machinery implicated in various aspects of insertion, including a subunit of the EMC (Anghel et al., 2017). These predicted homologs of YidC include the Get1/2 complex, EMC3 and the putative insertase TMCO1. While both Get1/2 and EMC have been shown to facilitate TMD insertion in purified systems, TMCO1's biochemical function remains a matter of speculation. However, TMCO1 has been observed in photo-crosslinking assays to interact with a nascent TMD-containing protein (Anghel et al., 2017).

Further arguing for potential similarities between insertion factors, homology searches of Get1 using HHpred identifies the EMC3 subunit with high confidence (Söding et al., 2005). Both seem to be members of the DUF106 family. The crystal structure of an archaeal DUF106 family member (Mj0480) revealed a three-TMD protein that structurally aligns well with three of the five TMDs of YidC (Borowska et al., 2015). Collectively, these findings argue for an ancient evolutionary relationship between the YidC family of insertases and the ER insertases Get1/2, EMC, and potentially TMCO1, suggesting that they might share mechanisms for TMD insertion. Ultimately, structural information about the EMC, as well as Get1/Get2 and TMCO1 are needed to determine whether these factors operate similarly to YidC.

A number of outstanding questions remain about the role of EMC in TA protein insertion. Since general cytosolic TMD binding partners bind rather indiscriminately, one can reason that it is EMC which establishes specificity and selects appropriate substrates for ER membrane insertion. For this, it must somehow be able to distinguish between highly similar ER and mitochondrially destined TA proteins. Since trypsin treatment, which digests cytosolically exposed fragments of the EMC, inactivates its insertase activity, initial TMD engagement might occur with a cytosolic subunit. This interaction may subsequently provide access to the membrane embedded subunits that catalyse the actual TMD insertion reaction. Of note, the cytosolic ECM2 has been shown to contain three tetracopeptide repeat domains, which are known to mediate protein-protein interactions. Speculatively, EMC2 may be responsible for mediating interactions with cytosolic chaperones delivering substrates, or directly interacting with the substrates themselves. The exact details of the route taken by TA

## Chapter 7: Perspectives and future work

TMDs into the lipid bilayer via EMC awaits future biochemical analysis, hopefully guided by structural information.

Secondly, our results from the *in vivo* mammalian screen, in addition to EMC's homology to YidC strongly suggests that it may act in conjunction with Sec61 for the insertion of multi-pass IMPs. The exact biochemical role of the ECM in this capacity awaits further investigation. At this point, one can only speculate as to the various roles it may have.

It may be important for the actual insertion of poor TMDs that fail recognition by Sec61, be involved in determining topology, or simply 'hold' TMDs until the entire polypeptide can be inserted and start folding. To distinguish between these possibilities, the key insight that EMC can act as a TMD insertase, along with the ability to purify and reconstitute it in liposomes can be combined with the *in vitro*  $\beta$ 1-AR system to establish the role of EMC in co-translational insertion.

## References

- Abell, B.M., Jung, M., Oliver, J.D., Knight, B.C., Tyedmers, J., Zimmermann, R., and High, S. (2003). Tail-anchored and signal-anchored proteins utilize overlapping pathways during membrane insertion. *J. Biol. Chem.* *278*, 5669–5678.
- Abell, B.M., Pool, M.R., Schlenker, O., Sinning, I., and High, S. (2004). Signal recognition particle mediates post-translational targeting in eukaryotes. *EMBO J.* *23*, 2755–2764.
- Abell, B.M., Rabu, C., Leznicki, P., Young, J.C., and High, S. (2007). Post-translational integration of tail-anchored proteins is facilitated by defined molecular chaperones. *J. Cell Sci.* *120*, 1743–1751.
- Akopian, D., Shen, K., Zhang, X., and Shan, S. (2013). Signal recognition particle: an essential protein-targeting machine. *Annu. Rev. Biochem.* *82*, 693–721.
- Anghel, S.A., McGilvray, P.T., Hegde, R.S., and Keenan, R.J. (2017). Identification of Oxal1 Homologs Operating in the Eukaryotic Endoplasmic Reticulum. *Cell Rep.* *21*, 3708–3716.
- Ast, T., Cohen, G., and Schuldiner, M. (2013). A network of cytosolic factors targets SRP-independent proteins to the endoplasmic reticulum. *Cell* *152*, 1134–1145.
- Audigier, Y., Friedlander, M., and Blobel, G. (1987). Multiple topogenic sequences in bovine opsin. *Proc. Natl. Acad. Sci. U. S. A.* *84*, 5783–5787.
- Aviram, N., Ast, T., Costa, E.A., Arakel, E.C., Chuartzman, S.G., Jan, C.H., Haßdenteufel, S., Dudek, J., Jung, M., Schorr, S., et al. (2016). The SND proteins constitute an alternative targeting route to the endoplasmic reticulum. *Nature* *540*, 134–138.
- Azzam, M.E., and Algranati, I.D. (1973). Mechanism of puromycin action: fate of ribosomes after release of nascent protein chains from polysomes. *Proc. Natl. Acad. Sci. U. S. A.* *70*, 3866–3869.
- Bacher, G., Lütcke, H., Jungnickel, B., Rapoport, T.A., and Dobberstein, B. (1996). Regulation by the ribosome of the GTPase of the signal-recognition particle during protein targeting. *Nature* *381*, 248–251.
- Baldridge, R.D., and Rapoport, T.A. (2016). Autoubiquitination of the Hrd1 Ligase Triggers Protein Retrotranslocation in ERAD. *Cell* *166*, 394–407.
- Bays, N.W., Gardner, R.G., Seelig, L.P., Joazeiro, C.A., and Hampton, R.Y. (2001). Hrd1p/Der3p is a membrane-anchored ubiquitin ligase required for ER-associated degradation. *Nat. Cell Biol.* *3*, 24–29.
- Beckmann, R., Bubeck, D., Grassucci, R., Penczek, P., Verschoor, A., Blobel, G., and Frank, J. (1997). Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex. *Science* *278*, 2123–2126.
- Beilharz, T., Egan, B., Silver, P.A., Hofmann, K., and Lithgow, T. (2003). Bipartite signals mediate subcellular targeting of tail-anchored membrane proteins in *Saccharomyces*

## References

- cerevisiae. *J. Biol. Chem.* 278, 8219–8223.
- Van den Berg, B., Clemons, W.M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C., and Rapoport, T.A. (2004). X-ray structure of a protein-conducting channel. *Nature* 427, 36–44.
- Berndt, U., Oellerer, S., Zhang, Y., Johnson, A.E., and Rospert, S. (2009). A signal-anchor sequence stimulates signal recognition particle binding to ribosomes from inside the exit tunnel. *Proc. Natl. Acad. Sci. U. S. A.* 106, 1398–1403.
- Bircham, P.W., Maass, D.R., Roberts, C.A., Kiew, P.Y., Low, Y.S., Yegambaram, M., Matthews, J., Jack, C.A., and Atkinson, P.H. (2011). Secretory pathway genes assessed by high-throughput microscopy and synthetic genetic array analysis. *Mol. Biosyst.* 7, 2589–2598.
- Blobel, G. (1980). Intracellular protein topogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 77, 1496–1500.
- Borgese, N., and Fasana, E. (2011). Targeting pathways of C-tail-anchored proteins. *Biochim. Biophys. Acta - Biomembr.* 1808, 937–946.
- Borowska, M.T., Dominik, P.K., Anghel, S.A., Kosiakoff, A.A., and Keenan, R.J. (2015). A YidC-like Protein in the Archaeal Plasma Membrane. *Structure* 23, 1715–1724.
- Bozkurt, G., Stjepanovic, G., Vilardi, F., Amlacher, S., Wild, K., Bange, G., Favalaro, V., Rippe, K., Hurt, E., Dobberstein, B., et al. (2009). Structural insights into tail-anchored protein binding and membrane insertion by Get3. *Proc. Natl. Acad. Sci. U. S. A.* 106, 21131–21136.
- Brambillasca, S., Yabal, M., Soffientini, P., Stefanovic, S., Makarow, M., Hegde, R.S., and Borgese, N. (2005). Transmembrane topogenesis of a tail-anchored protein is modulated by membrane lipid composition. *EMBO J.* 24, 2533–2542.
- Brambillasca, S., Yabal, M., Makarow, M., and Borgese, N. (2006). Unassisted translocation of large polypeptide domains across phospholipid bilayers. *J. Cell Biol.* 175, 767–777.
- Braunger, K., Pfeffer, S., Shrimal, S., Gilmore, R., Berninghausen, O., Mandon, E.C., Becker, T., Förster, F., and Beckmann, R. (2018). Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. *Science* (80-. ). 7899, eaar7899
- Brkljacic, J., Zhao, Q., and Meier, I. (2009). WPP-domain proteins mimic the activity of the HSC70-1 chaperone in preventing mistargeting of RanGAP1-anchoring protein WIT1. *Plant Physiol.* 151, 142–154.
- Buentzel, J., and Thoms, S. (2017). The Use of Glycosylation Tags as Reporters for Protein Entry into the Endoplasmic Reticulum in Yeast and Mammalian Cells. *Methods Mol. Biol.* 1595, 221–232.
- Casson, J., McKenna, M., Haßdenteufel, S., Aviram, N., Zimmerman, R., and High, S. (2017). Multiple pathways facilitate the biogenesis of mammalian tail-anchored proteins. *J.*

## References

Cell Sci. *130*, 3851–3861.

Chartron, J.W., Hunt, K.C.L., and Frydman, J. (2016). Cotranslational signal-independent SRP preloading during membrane targeting. *Nature* *536*, 224–228.

Chin, J.W., Martin, A.B., King, D.S., Wang, L., and Schultz, P.G. (2002). Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 11020–11024.

Chio, U.S., Cho, H., and Shan, S. (2017). Mechanisms of Tail-Anchored Membrane Protein Targeting and Insertion. *Annu. Rev. Cell Dev. Biol.* *33*, 417–438.

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., and Kopito, R.R. (2012). Defining human ERAD networks through an integrative mapping strategy. *Nat. Cell Biol.* *14*, 93–105.

Claessen, J.H.L., and Ploegh, H.L. (2011). BAT3 guides misfolded glycoproteins out of the endoplasmic reticulum. *PLoS One* *6*, e28542.

Claessen, J.H.L., Sanyal, S., and Ploegh, H.L. (2014). The chaperone BAG6 captures dislocated glycoproteins in the cytosol. *PLoS One* *9*, e90204.

Colombo, S.F., Longhi, R., and Borgese, N. (2009). The role of cytosolic proteins in the insertion of tail-anchored proteins into phospholipid bilayers. *J. Cell Sci.* *122*, 2383–2392.

Connolly, T., and Gilmore, R. (1989). The signal recognition particle receptor mediates the GTP-dependent displacement of SRP from the signal sequence of the nascent polypeptide. *Cell* *57*, 599–610.

Connolly, T., Rapiejko, P.J., and Gilmore, R. (1991). Requirement of GTP hydrolysis for dissociation of the signal recognition particle from its receptor. *Science* *252*, 1171–1173.

Conti, B.J., Devaraneni, P.K., Yang, Z., David, L.L., and Skach, W.R. (2015).

Cotranslational stabilization of Sec62/63 within the ER Sec61 translocon is controlled by distinct substrate-driven translocation events. *Mol. Cell* *58*, 269–283.

Costa, E.A., Subramanian, K., Nunnari, J., and Weissman, J.S. (2018). Defining the physiological role of SRP in protein-targeting efficiency and specificity. *Science* *359*, 689–692.

Cymer, F., and von Heijne, G. (2013). Cotranslational folding of membrane proteins probed by arrest-peptide-mediated force measurements. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 14640–14645.

Cymer, F., von Heijne, G., and White, S.H. (2015). Mechanisms of integral membrane protein insertion and folding. *J. Mol. Biol.* *427*, 999–1022.

Dejgaard, K., Theberge, J.-F., Heath-Engel, H., Chevet, E., Tremblay, M.L., and Thomas, D.Y. (2010). Organization of the Sec61 translocon, studied by high resolution native electrophoresis. *J. Proteome Res.* *9*, 1763–1771.

## References

- Devaraneni, P.K., Conti, B., Matsumura, Y., Yang, Z., Johnson, A.E., and Skach, W.R. (2011). Stepwise insertion and inversion of a type II signal anchor sequence in the ribosome-Sec61 translocon complex. *Cell* 146, 134–147.
- Do, H., Falcone, D., Lin, J., Andrews, D.W., and Johnson, A.E. (1996). The cotranslational integration of membrane proteins into the phospholipid bilayer is a multistep process. *Cell* 85, 369–378.
- Ellgaard, L., and Helenius, A. (2003). Quality control in the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* 4, 181–191.
- Elofsson, A., and Heijne, G. von (2007). Membrane Protein Structure: Prediction versus Reality. *Annu. Rev. Biochem.* 76, 125–140.
- Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* 2, 953–971.
- Enquist, K., Fransson, M., Boekel, C., Bengtsson, I., Geiger, K., Lang, L., Pettersson, A., Johansson, S., von Heijne, G., and Nilsson, I. (2009). Membrane-integration characteristics of two ABC transporters, CFTR and P-glycoprotein. *J. Mol. Biol.* 387, 1153–1164.
- Erdmann, F., Schäuble, N., Lang, S., Jung, M., Honigsmann, A., Ahmad, M., Dudek, J., Benedix, J., Harsman, A., Kopp, A., et al. (2011). Interaction of calmodulin with Sec61 $\alpha$  limits Ca<sup>2+</sup> leakage from the endoplasmic reticulum. *EMBO J.* 30, 17–31.
- Favaloro, V., Spasic, M., Schwappach, B., and Dobberstein, B. (2008). Distinct targeting pathways for the membrane insertion of tail-anchored (TA) proteins. *J. Cell Sci.* 121, 1832–1840.
- Feige, M.J., Behnke, J., Mittag, T., and Hendershot, L.M. (2015). Dimerization-dependent folding underlies assembly control of the clonotypic  $\alpha\beta$ T cell receptor chains. *J. Biol. Chem.* 290, 26821–26831.
- Finn, B.E., Evenäs, J., Drakenberg, T., Waltho, J.P., Thulin, E., and Forsén, S. (1995). Calcium-induced structural changes and domain autonomy in calmodulin. *Nat. Struct. Biol.* 2, 777–783.
- Fons, R.D., Bogert, B.A., and Hegde, R.S. (2003). Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. *J. Cell Biol.* 160, 529–539.
- Freeman, M. (2014). The rhomboid-like superfamily: molecular mechanisms and biological roles. *Annu. Rev. Cell Dev. Biol.* 30, 235–254.
- Garrison, J.L., Kunkel, E.J., Hegde, R.S., and Taunton, J. (2005). A substrate-specific inhibitor of protein translocation into the endoplasmic reticulum. *Nature* 436, 285–289.
- Gilmore, R., Blobel, G., and Walter, P. (1982a). Protein translocation across the endoplasmic reticulum. I. Detection in the microsomal membrane of a receptor for the signal recognition particle. *J. Cell Biol.* 95, 463–469.

## References

- Gilmore, R., Walter, P., and Blobel, G. (1982b). Protein translocation across the endoplasmic reticulum. II. Isolation and characterization of the signal recognition particle receptor. *J. Cell Biol.* *95*, 470–477.
- Goder, V., Bieri, C., and Spiess, M. (1999). Glycosylation can influence topogenesis of membrane proteins and reveals dynamic reorientation of nascent polypeptides within the translocon. *J. Cell Biol.* *147*, 257–266.
- Gogala, M., Becker, T., Beatrix, B., Armache, J.P., Barrio-Garcia, C., Berninghausen, O., and Beckmann, R. (2014). Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion. *Nature* *506*, 107–110.
- Görllich, D., and Rapoport, T.A. (1993). Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. *Cell* *75*, 615–630.
- Görllich, D., Hartmann, E., Prehn, S., and Rapoport, T.A. (1992a). A protein of the endoplasmic reticulum involved early in polypeptide translocation. *Nature* *357*, 47–52.
- Görllich, D., Prehn, S., Hartmann, E., Kalies, K.U., and Rapoport, T.A. (1992b). A mammalian homolog of SEC61p and SECYp is associated with ribosomes and nascent polypeptides during translocation. *Cell* *71*, 489–503.
- Greenblatt, E.J., Olzmann, J.A., and Kopito, R.R. (2011). Derlin-1 is a rhomboid pseudoprotease required for the dislocation of mutant  $\alpha$ -1 antitrypsin from the endoplasmic reticulum. *Nat. Struct. Mol. Biol.* *18*, 1147–1152.
- Gristick, H.B., Rao, M., Chartron, J.W., Rome, M.E., Shan, S.-O., and Clemons, W.M. (2014). Crystal structure of ATP-bound Get3-Get4-Get5 complex reveals regulation of Get3 by Get4. *Nat. Struct. Mol. Biol.* *21*, 437–442.
- Grudnik, P., Bange, G., and Sinning, I. (2009). Protein targeting by the signal recognition particle. *Biol. Chem.* *390*, 775–782.
- Hainzl, T., and Sauer-Eriksson, A.E. (2015). Signal-sequence induced conformational changes in the signal recognition particle. *Nat. Commun.* *6*, 7163.
- Halic, M., and Beckmann, R. (2005). The signal recognition particle and its interactions during protein targeting. *Curr. Opin. Struct. Biol.* *15*, 116–125.
- Halic, M., Becker, T., Pool, M.R., Spahn, C.M.T., Grassucci, R.A., Frank, J., and Beckmann, R. (2004). Structure of the signal recognition particle interacting with the elongation-arrested ribosome. *Nature* *427*, 808–814.
- Haßdenteufel, S., Schäuble, N., Cassella, P., Leznicki, P., Müller, A., High, S., Jung, M., and Zimmermann, R. (2011). Ca<sup>2+</sup>-calmodulin inhibits tail-anchored protein insertion into the mammalian endoplasmic reticulum membrane. *FEBS Lett.* *585*, 3485–3490.
- Haßdenteufel, S., Sicking, M., Schorr, S., Aviram, N., Fecher-Trost, C., Schuldiner, M., Jung,

## References

- M., Zimmermann, R., and Lang, S. (2017). hSnd2 protein represents an alternative targeting factor to the endoplasmic reticulum in human cells. *FEBS Lett.* *38*, 42–49.
- Hauser, A.S., Attwood, M.M., Rask-Andersen, M., Schiöth, H.B., and Gloriam, D.E. (2017). Trends in GPCR drug discovery: new agents, targets and indications. *Nat. Rev. Drug Discov.* *16*, 829–842.
- Hegde, R.S., and Keenan, R.J. (2011). Tail-anchored membrane protein insertion into the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* *12*, 787–798.
- Heinrich, S.U., and Rapoport, T.A. (2003). Cooperation of transmembrane segments during the integration of a double-spanning protein into the ER membrane. *EMBO J.* *22*, 3654–3663.
- Heinrich, S.U., Mothes, W., Brunner, J., and Rapoport, T.A. (2000). The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell* *102*, 233–244.
- Hell, K., Neupert, W., and Stuart, R.A. (2001). Oxa1p acts as a general membrane insertion machinery for proteins encoded by mitochondrial DNA. *EMBO J.* *20*, 1281–1288.
- Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S.H., and von Heijne, G. (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* *433*, 377–381.
- Hessa, T., Meindl-Beinker, N.M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S.H., and von Heijne, G. (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* *450*, 1026–1030.
- Hessa, T., Sharma, A., Mariappan, M., Eshleman, H.D., Gutierrez, E., and Hegde, R.S. (2011). Protein targeting and degradation are coupled for elimination of mislocalized proteins. *Nature* *475*, 394–397.
- High, S., Martoglio, B., Görlich, D., Andersen, S.S., Ashford, A.J., Giner, A., Hartmann, E., Prehn, S., Rapoport, T.A., and Dobberstein, B. (1993). Site-specific photocross-linking reveals that Sec61p and TRAM contact different regions of a membrane-inserted signal sequence. *J. Biol. Chem.* *268*, 26745–26751.
- Horie, C., Suzuki, H., Sakaguchi, M., and Mihara, K. (2002). Characterization of signal that directs C-tail-anchored proteins to mammalian mitochondrial outer membrane. *Mol. Biol. Cell* *13*, 1615–1625.
- Hu, J., Li, J., Qian, X., Denic, V., and Sha, B. (2009). The Crystal Structures of Yeast Get3 Suggest a Mechanism for Tail-Anchored Protein Membrane Insertion. *PLoS One* *4*, 1–7.
- Ingolia, N.T., Ghaemmighami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* *324*, 218–223.
- Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*

## References

147, 789–802.

Itakura, E., Zavodszky, E., Shao, S., Wohlever, M.L., Keenan, R.J., and Hegde, R.S. (2016). Ubiquilins Chaperone and Triage Mitochondrial Membrane Proteins for Degradation. *Mol. Cell* 63, 21–33.

Jan, C.H., Williams, C.C., and Weissman, J.S. (2014). Principles of ER cotranslational translocation revealed by proximity- specific ribosome profiling. *Science* (80-. ). 346, 748–751.

Janda, C.Y., Li, J., Oubridge, C., Hernández, H., Robinson, C. V., and Nagai, K. (2010). Recognition of a signal peptide by the signal recognition particle. *Nature* 465, 507–510.

Jonikas, M.C., Collins, S.R., Denic, V., Oh, E., Quan, E.M., Schmid, V., Weibezahn, J., Schwappach, B., Walter, P., Weissman, J.S., et al. (2009). Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* 323, 1693–1697.

Jung, S., Kim, J.E.H., Reithinger, J.H., and Kim, H. (2014). The Sec62-Sec63 translocon facilitates translocation of the C-terminus of membrane proteins. *J. Cell Sci.* 127, 4270–4278.

Jungnickel, B., and Rapoport, T.A. (1995). A posttargeting signal sequence recognition event in the endoplasmic reticulum membrane. *Cell* 82, 261–270.

Kalbfleisch, T., Cambon, A., and Wattenberg, B.W. (2007). A bioinformatics approach to identifying tail-anchored proteins in the human genome. *Traffic* 8, 1687–1694.

Kalies, K.U., Rapoport, T.A., and Hartmann, E. (1998). The beta subunit of the Sec61 complex facilitates cotranslational protein transport and interacts with the signal peptidase during translocation. *J. Cell Biol.* 141, 887–894.

Kang, S.W., Rane, N.S., Kim, S.J., Garrison, J.L., Taunton, J., and Hegde, R.S. (2006). Substrate-Specific Translocational Attenuation during ER Stress Defines a Pre-Emptive Quality Control Pathway. *Cell* 127, 999–1013.

Kauko, A., Hedin, L.E., Thebaud, E., Cristobal, S., Elofsson, A., and von Heijne, G. (2010). Repositioning of transmembrane alpha-helices during membrane protein folding. *J. Mol. Biol.* 397, 190–201.

Keenan, R.J., Freymann, D.M., Walter, P., and Stroud, R.M. (1998). Crystal structure of the signal sequence binding subunit of the signal recognition particle. *Cell* 94, 181–191.

Keenan, R.J., Freymann, D.M., Stroud, R.M., and Walter, P. (2001). The Signal Recognition Particle. *Annu. Rev. Biochem.* 70, 755–775.

Klenner, C., Yuan, J., Dalbey, R.E., and Kuhn, A. (2008). The Pf3 coat protein contacts TM1 and TM3 of YidC during membrane biogenesis. *FEBS Lett.* 582, 3967–3972.

Kobayashi, T., Nureki, O., Ishitani, R., Yaremchuk, A., Tukalo, M., Cusack, S., Sakamoto, K., and Yokoyama, S. (2003). Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat. Struct. Biol.* 10, 425–432.

Kriechbaumer, V., Shaw, R., Mukherjee, J., Bowsher, C.G., Harrison, A.-M., and Abell, B.M. (2009). Subcellular distribution of tail-anchored proteins in Arabidopsis. *Traffic* 10,

## References

1753–1764.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L.. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. Edited by F. Cohen. *J. Mol. Biol.* *305*, 567–580.

Kumazaki, K., Chiba, S., Takemoto, M., Furukawa, A., Nishiyama, K., Sugano, Y., Mori, T., Dohmae, N., Hirata, K., Nakada-Nakura, Y., et al. (2014). Structural basis of Sec-independent membrane protein insertion by YidC. *Nature* *509*, 516–520.

Kuroda, R., Ikenoue, T., Honsho, M., Tsujimoto, S., Mitoma, J.Y., and Ito, A. (1998). Charged amino acids at the carboxyl-terminal portions determine the intracellular locations of two isoforms of cytochrome b5. *J. Biol. Chem.* *273*, 31097–31102.

Kutay, U., Hartmann, E., and Rapoport, T.A. (1993). A class of membrane proteins with a C-terminal anchor. *Trends Cell Biol.* *3*, 72–75.

Kutay, U., Ahnert-Hilger, G., Hartmann, E., Wiedenmann, B., and Rapoport, T.A. (1995). Transport route for synaptobrevin via a novel pathway of insertion into the endoplasmic reticulum membrane. *EMBO J.* *14*, 217–223.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* *157*, 105–132.

Lahiri, S., Chao, J.T., Tavassoli, S., Wong, A.K.O., Choudhary, V., Young, B.P., Loewen, C.J.R., and Prinz, W.A. (2014). A conserved endoplasmic reticulum membrane protein complex (EMC) facilitates phospholipid transfer from the ER to mitochondria. *PLoS Biol.* *12*, e1001969.

Lakkaraju, A.K.K., Thankappan, R., Mary, C., Garrison, J.L., Taunton, J., and Strub, K. (2012). Efficient secretion of small proteins in mammalian cells relies on Sec62-dependent posttranslational translocation. *Mol. Biol. Cell* *23*, 2712–2722.

Lang, S., Benedix, J., Fedeles, S. V., Schorr, S., Schirra, C., Schäuble, N., Jalal, C., Greiner, M., Hassdenteufel, S., Tatzelt, J., et al. (2012). Different effects of Sec61 $\alpha$ , Sec62 and Sec63 depletion on transport of polypeptides into the endoplasmic reticulum of mammalian cells. *J. Cell Sci.* *125*, 1958–1969.

Lareau, L.F., Hite, D.H., Hogan, G.J., and Brown, P.O. (2014). Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *Elife* *3*, e01257.

Lee, J., Kim, D.H., and Hwang, I. (2014). Specific targeting of proteins to outer envelope membranes of endosymbiotic organelles, chloroplasts, and mitochondria. *Front. Plant Sci.* *5*, 173.

Li, L., Park, E., Ling, J., Ingram, J., Ploegh, H., and Rapoport, T.A. (2016). Crystal structure of a substrate-engaged SecY protein-translocation channel. *Nature* *531*, 395–399.

Lilley, B.N., and Ploegh, H.L. (2004). A membrane protein required for dislocation of misfolded proteins from the ER. *Nature* *429*, 834–840.

## References

- Louie, R.J., Guo, J., Rodgers, J.W., White, R., Shah, N., Pagant, S., Kim, P., Livstone, M., Dolinski, K., McKinney, B.A., et al. (2012). A yeast phenomic model for the gene interaction network modulating CFTR- $\Delta$ F508 protein biogenesis. *Genome Med.* *4*, 103.
- Lu, Y., Xiong, X., Helm, A., Kimani, K., Bragin, A., and Skach, W.R. (1998). Co- and posttranslational translocation mechanisms direct cystic fibrosis transmembrane conductance regulator N terminus transmembrane assembly. *J. Biol. Chem.* *273*, 568–576.
- Lu, Y., Turnbull, I.R., Bragin, A., Carveth, K., Verkman, A.S., and Skach, W.R. (2000). Reorientation of aquaporin-1 topology during maturation in the endoplasmic reticulum. *Mol. Biol. Cell* *11*, 2973–2985.
- MacKinnon, A.L., Garrison, J.L., Hegde, R.S., and Taunton, J. (2007). Photo-leucine incorporation reveals the target of a cyclodepsipeptide inhibitor of cotranslational translocation. *J. Am. Chem. Soc.* *129*, 14560–14561.
- MacKinnon, A.L., Paavilainen, V.O., Sharma, A., Hegde, R.S., and Taunton, J. (2014). An allosteric Sec61 inhibitor traps nascent transmembrane helices at the lateral gate. *Elife* *2014*, 1–23.
- Madamanchi, A. (2007). Beta-adrenergic receptor signaling in cardiac function and heart failure. *McGill J. Med.* *10*, 99–104.
- Mariappan, M., Li, X., Stefanovic, S., Sharma, A., Mateja, A., Keenan, R.J., and Hegde, R.S. (2010). A ribosome-associating factor chaperones tail-anchored membrane proteins. *Nature* *466*, 1120–1124.
- Mariappan, M., Mateja, A., Dobosz, M., Bove, E., Hegde, R.S., and Keenan, R.J. (2011). The mechanism of membrane-associated steps in tail-anchored protein insertion. *Nature* *477*, 61–66.
- Martoglio, B., Hofmann, M.W., Brunner, J., and Dobberstein, B. (1995). The protein-conducting channel in the membrane of the endoplasmic reticulum is open laterally toward the lipid bilayer. *Cell* *81*, 207–214.
- Martoglio, B., Graf, R., and Dobberstein, B. (1997). Signal peptide fragments of preprolactin and HIV-1 p-gp160 interact with calmodulin. *EMBO J.* *16*, 6636–6645.
- Marty, N.J., Teresinski, H.J., Hwang, Y.T., Clendening, E.A., Gidda, S.K., Sliwiska, E., Zhang, D., Miernyk, J.A., Brito, G.C., Andrews, D.W., et al. (2014). New insights into the targeting of a subset of tail-anchored proteins to the outer mitochondrial membrane. *Front. Plant Sci.* *5*, 426.
- Mateja, A., Szlachcic, A., Downing, M.E., Dobosz, M., Mariappan, M., Hegde, R.S., and Keenan, R.J. (2009). The structural basis of tail-anchored membrane protein recognition by Get3. *Nature* *461*, 361–366.
- Mateja, A., Paduch, M., Chang, H.-Y., Szydlowska, A., Kossiakoff, A.A., Hegde, R.S., and Keenan, R.J. (2015). Structure of the Get3 targeting factor in complex with its membrane protein cargo. *Science* (80-. ). *347*, 1152–1155.

## References

- McCormick, P.J., Miao, Y., Shao, Y., Lin, J., and Johnson, A.E. (2003). Cotranslational protein integration into the ER membrane is mediated by the binding of nascent chains to translocon proteins. *Mol. Cell* *12*, 329–341.
- Meacock, S.L., Lecomte, F.J.L., Crawshaw, S.G., and High, S. (2002). Different transmembrane domains associate with distinct endoplasmic reticulum components during membrane integration of a polytopic protein. *Mol. Biol. Cell* *13*, 4114–4129.
- Meador, W.E., Means, A.R., and Quioco, F.A. (1992). Target enzyme recognition by calmodulin: 2.4 Å structure of a calmodulin-peptide complex. *Science* *257*, 1251–1255.
- Means, A.R., and Dedman, J.R. (1980). Calmodulin--an intracellular calcium receptor. *Nature* *285*, 73–77.
- Means, A.R., VanBerkum, M.F., Bagchi, I., Lu, K.P., and Rasmussen, C.D. (1991). Regulatory functions of calmodulin. *Pharmacol. Ther.* *50*, 255–270.
- van Meer, G. (1989). Lipid traffic in animal cells. *Annu. Rev. Cell Biol.* *5*, 247–275.
- Mellman, I., and Warren, G. (2000). The road taken: past and future foundations of membrane traffic. *Cell* *100*, 99–112.
- Ménétret, J.-F., Hegde, R.S., Aguiar, M., Gygi, S.P., Park, E., Rapoport, T.A., and Akey, C.W. (2008). Single copies of Sec61 and TRAP associate with a nontranslating mammalian ribosome. *Structure* *16*, 1126–1137.
- Ménétret, J.F., Hegde, R.S., Heinrich, S.U., Chandramouli, P., Ludtke, S.J., Rapoport, T.A., and Akey, C.W. (2005). Architecture of the ribosome-channel complex derived from native membranes. *J. Mol. Biol.* *348*, 445–457.
- Meyer, D.I., Krause, E., and Dobberstein, B. (1982). Secretory protein translocation across membranes--the role of the "docking protein". *Nature* *297*, 647–650.
- Meyer, H.A., Grau, H., Kraft, R., Kostka, S., Prehn, S., Kalies, K.U., and Hartmann, E. (2000). Mammalian Sec61 is associated with Sec62 and Sec63. *J. Biol. Chem.* *275*, 14550–14557.
- Miller, J.D., Tajima, S., Lauffer, L., and Walter, P. (1995). The beta subunit of the signal recognition particle receptor is a transmembrane GTPase that anchors the alpha subunit, a peripheral membrane GTPase, to the endoplasmic reticulum membrane. *J. Cell Biol.* *128*, 273–282.
- Mock, J.-Y., Chartron, J.W., Zaslaver, M., Xu, Y., Ye, Y., and Clemons, W.M. (2015). Bag6 complex contains a minimal tail-anchor-targeting module and a mock BAG domain. *Proc. Natl. Acad. Sci. U. S. A.* *112*, 106–111.
- Moore, M., Harrison, M.S., Peterson, E.C., and Henry, R. (2000). Chloroplast Oxa1p homolog albino3 is required for post-translational integration of the light harvesting chlorophyll-binding protein into thylakoid membranes. *J. Biol. Chem.* *275*, 1529–1532.
- Moss, K., Helm, A., Lu, Y., Bragin, A., and Skach, W.R. (1998). Coupled translocation

## References

events generate topological heterogeneity at the endoplasmic reticulum membrane. *Mol. Biol. Cell* 9, 2681–2697.

Mothes, W., Prehn, S., and Rapoport, T.A. (1994). Systematic probing of the environment of a translocating secretory protein during translocation through the ER membrane. *EMBO J.* 13, 3973–3982.

Mothes, W., Jungnickel, B., Brunner, J., and Rapoport, T.A. (1998). Signal sequence recognition in cotranslational translocation by protein components of the endoplasmic reticulum membrane. *J. Cell Biol.* 142, 355–364.

Nagamori, S., Smirnova, I.N., and Kaback, H.R. (2004). Role of YidC in folding of polytopic membrane proteins. *J. Cell Biol.* 165, 53–62.

Ng, D.T., Brown, J.D., and Walter, P. (1996). Signal sequences specify the targeting route to the endoplasmic reticulum membrane. *J. Cell Biol.* 134, 269–278.

Nilsson, I.M., and von Heijne, G. (1993). Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *J. Biol. Chem.* 268, 5798–5801.

O’Neil, K.T., and DeGrado, W.F. (1990). How calmodulin binds its targets: sequence independent recognition of amphiphilic alpha-helices. *Trends Biochem. Sci.* 15, 59–64.

Oliver, J., Jungnickel, B., Görlich, D., Rapoport, T., and High, S. (1995). The Sec61 complex is essential for the insertion of proteins into the membrane of the endoplasmic reticulum. *FEBS Lett.* 362, 126–130.

Panzner, S., Dreier, L., Hartmann, E., Kostka, S., and Rapoport, T.A. (1995). Posttranslational Protein Transport in Yeast Reconstituted with a Purified Complex of Set Proteins and Kar2p. *61*, 561–570.

Papaloukas, C., Granseth, E., Viklund, H., and Elofsson, A. (2008). Estimating the length of transmembrane helices using Z-coordinate predictions. *Protein Sci.* 17, 271–278.

Pelham, H.R., and Jackson, R.J. (1976). An efficient mRNA-dependent translation system from reticulocyte lysates. *Eur. J. Biochem.* 67, 247–256.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786.

Pfeffer, S., Dudek, J., Gogala, M., Schorr, S., Linxweiler, J., Lang, S., Becker, T., Beckmann, R., Zimmermann, R., and Förster, F. (2014). Structure of the mammalian oligosaccharyl-transferase complex in the native ER protein translocon. *Nat. Commun.* 5, 3072.

Pfeffer, S., Burbaum, L., Unverdorben, P., Pech, M., Chen, Y., Zimmermann, R., Beckmann, R., and Förster, F. (2015). Structure of the native Sec61 protein-conducting channel. *Nat. Commun.* 6, 8403.

Pfeffer, S., Dudek, J., Zimmermann, R., and Förster, F. (2016). Organization of the native

## References

ribosome-translocon complex at the mammalian endoplasmic reticulum membrane. *Biochim. Biophys. Acta* *1860*, 2122–2129.

Pitonzo, D., and Skach, W.R. (2006). Molecular mechanisms of aquaporin biogenesis by the endoplasmic reticulum Sec61 translocon. *Biochim. Biophys. Acta* *1758*, 976–988.

Pitonzo, D., Yang, Z., Matsumura, Y., Johnson, A.E., and Skach, W.R. (2009). Sequence-specific retention and regulated integration of a nascent membrane protein by the endoplasmic reticulum Sec61 translocon. *Mol. Biol. Cell* *20*, 685–698.

Plath, K., and Rapoport, T.A. (2000). Spontaneous release of cytosolic proteins from posttranslational substrates before their transport into the endoplasmic reticulum. *J. Cell Biol.* *151*, 167–178.

Plath, K., Mothes, W., Wilkinson, B.M., Stirling, C.J., and Rapoport, T.A. (1998). Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. *Cell* *94*, 795–807.

Rao, M., Okreglak, V., Chio, U.S., Cho, H., Walter, P., and Shan, S.-O. (2016). Multiple selection filters ensure accurate tail-anchored membrane protein targeting. *Elife* *5*.

Rapiejko, P.J., and Gilmore, R. (1997). Empty site forms of the SRP54 and SR alpha GTPases mediate targeting of ribosome-nascent chain complexes to the endoplasmic reticulum. *Cell* *89*, 703–713.

Rapoport, T.A. (2007). Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* *450*, 663–669.

Rapoport, T.A., Goder, V., Heinrich, S.U., and Matlack, K.E.S. (2004). Membrane-protein integration and the role of the translocation channel. *Trends Cell Biol.* *14*, 568–575.

Rask-Andersen, M., Masuram, S., and Schiöth, H.B. (2014). The druggable genome: Evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication. *Annu. Rev. Pharmacol. Toxicol.* *54*, 9–26.

Richard, M., Boulin, T., Robert, V.J.P., Richmond, J.E., and Bessereau, J.-L. (2013). Biosynthesis of ionotropic acetylcholine receptors requires the evolutionarily conserved ER membrane complex. *Proc. Natl. Acad. Sci. U. S. A.* *110*, E1055-63.

Ridder, A.N.J.A., van de Hoef, W., Stam, J., Kuhn, A., de Kruijff, B., and Killian, J.A. (2002). Importance of hydrophobic matching for spontaneous insertion of a single-spanning membrane protein. *Biochemistry* *41*, 4946–4952.

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., et al. (2016). Mice lacking WRB reveal differential biogenesis requirements of tail-anchored proteins in vivo. *Sci. Rep.* *6*, 39464.

Rodrigo-Brenni, M.C., Gutierrez, E., and Hegde, R.S. (2014). Cytosolic Quality Control of Mislocalized Proteins Requires RNF126 Recruitment to Bag6. *Mol. Cell* *55*, 227–237.

## References

- Ruggiano, A., Foresti, O., and Carvalho, P. (2014). Quality control: ER-associated degradation: protein quality control and beyond. *J. Cell Biol.* *204*, 869–879.
- Sadlish, H., Pitonzo, D., Johnson, A.E., and Skach, W.R. (2005). Sequential triage of transmembrane segments by Sec61alpha during biogenesis of a native multispanning membrane protein. *Nat. Struct. Mol. Biol.* *12*, 870–878.
- Santos, R., Ursu, O., Gaulton, A., Bento, A.P., Donadi, R.S., Bologa, C.G., Karlsson, A., Al-Lazikani, B., Hersey, A., Oprea, T.I., et al. (2017). A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* *16*, 19–34.
- Satoh, T., Ohba, A., Liu, Z., Inagaki, T., and Satoh, A.K. (2015). dPob/EMC is essential for biosynthesis of rhodopsin and other multi-pass membrane proteins in *Drosophila* photoreceptors. *Elife* *4*.
- Savidis, G., McDougall, W.M., Meraner, P., Perreira, J.M., Portmann, J.M., Trincucci, G., John, S.P., Aker, A.M., Renzette, N., Robbins, D.R., et al. (2016). Identification of Zika Virus and Dengue Virus Dependency Factors using Functional Genomics. *Cell Rep.* *16*, 232–246.
- Schaffitzel, C., Oswald, M., Berger, I., Ishikawa, T., Abrahams, J.P., Koerten, H.K., Koning, R.I., and Ban, N. (2006). Structure of the *E. coli* signal recognition particle bound to a translating ribosome. *Nature* *444*, 503–506.
- Schibich, D., Gloge, F., Pöhner, I., Björkholm, P., Wade, R.C., von Heijne, G., Bukau, B., and Kramer, G. (2016). Global profiling of SRP interaction with nascent polypeptides. *Nature* *536*, 219–223.
- Schoebel, S., Mi, W., Stein, A., Ovchinnikov, S., Pavlovicz, R., DiMaio, F., Baker, D., Chambers, M.G., Su, H., Li, D., et al. (2017). Cryo-EM structure of the protein-conducting ERAD channel Hrd1 in complex with Hrd3. *Nature* *548*, 352–355.
- Schoenmakers, T.J., Visser, G.J., Flik, G., and Theuvsen, A.P. (1992). CHELATOR: an improved method for computing metal ion concentrations in physiological solutions. *Biotechniques* *12*, 870–874, 876–879.
- Schuldiner, M., Metz, J., Schmid, V., Denic, V., Rakwalska, M., Schmitt, H.D., Schwappach, B., and Weissman, J.S. (2008). The GET complex mediates insertion of tail-anchored proteins into the ER membrane. *Cell* *134*, 634–645.
- Serdiuk, T., Balasubramaniam, D., Sugihara, J., Mari, S.A., Kaback, H.R., and Müller, D.J. (2016). YidC assists the stepwise and stochastic folding of membrane proteins. *Nat. Chem. Biol.* *12*, 911–917.
- Serek, J., Bauer-Manz, G., Struhalla, G., van den Berg, L., Kiefer, D., Dalbey, R., and Kuhn, A. (2004). *Escherichia coli* YidC is a membrane insertase for Sec-independent proteins. *EMBO J.* *23*, 294–301.
- Setoguchi, K., Otera, H., and Mihara, K. (2006). Cytosolic factor- and TOM-independent import of C-tail-anchored mitochondrial outer membrane proteins. *EMBO J.* *25*, 5635–5647.

## References

- Shan, S., and Walter, P. (2005). Co-translational protein targeting by the signal recognition particle. *FEBS Lett.* *579*, 921–926.
- Shao, S., and Hegde, R.S. (2011a). Membrane protein insertion at the endoplasmic reticulum. *Annu. Rev. Cell Dev. Biol.* *27*, 25–56.
- Shao, S., and Hegde, R.S. (2011b). A calmodulin-dependent translocation pathway for small secretory proteins. *Cell* *147*, 1576–1588.
- Shao, S., Rodrigo-Brenni, M.C., Kivlen, M.H., and Hegde, R.S. (2017). Mechanistic basis for a molecular triage reaction. *Science* *355*, 298–302.
- Sharma, A., Mariappan, M., Appathurai, S., and Hegde, R.S. (2010). In vitro dissection of protein translocation into the mammalian endoplasmic reticulum. *Methods Mol. Biol.* *619*, 339–363.
- Shi, L.B., Skach, W.R., Ma, T., and Verkman, A.S. (1995). Distinct biogenesis mechanisms for the water channels MIWC and CHIP28 at the endoplasmic reticulum. *Biochemistry* *34*, 8250–8256.
- Shibatani, T., David, L.L., McCormack, A.L., Frueh, K., and Skach, W.R. (2005). Proteomic analysis of mammalian oligosaccharyltransferase reveals multiple subcomplexes that contain Sec61, TRAP, and two potential new subunits. *Biochemistry* *44*, 5982–5992.
- Shigemitsu, S., Cao, W., Terada, T., and Shimizu, K. (2016). Development of a prediction system for tail-anchored proteins. *BMC Bioinformatics* *17*, 378.
- Shimizu, Y., and Ueda, T. (2010). PURE technology. *Methods Mol. Biol.* *607*, 11–21.
- Skach, W.R. (1998). Topology of P-glycoproteins. *Methods Enzymol.* *292*, 265–278.
- Skach, W.R. (2009). Cellular mechanisms of membrane protein folding. *Nat. Struct. Mol. Biol.* *16*, 606–612.
- Skach, W.R., and Lingappa, V.R. (1993). Amino-terminal assembly of human P-glycoprotein at the endoplasmic reticulum is directed by cooperative actions of two internal sequences. *J. Biol. Chem.* *268*, 23552–23561.
- Skach, W.R., Shi, L.B., Calayag, M.C., Frigeri, A., Lingappa, V.R., and Verkman, A.S. (1994). Biogenesis and transmembrane topology of the CHIP28 water channel at the endoplasmic reticulum. *J. Cell Biol.* *125*, 803–815.
- Snapp, E.L., Reinhart, G.A., Bogert, B.A., Lippincott-Schwartz, J., and Hegde, R.S. (2004). The organization of engaged and quiescent translocons in the endoplasmic reticulum of mammalian cells. *J. Cell Biol.* *164*, 997–1007.
- Söding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* *33*, W244–8.
- Sommer, N., Junne, T., Kalies, K.-U., Spiess, M., and Hartmann, E. (2013). TRAP assists

## References

- membrane protein topogenesis at the mammalian ER membrane. *Biochim. Biophys. Acta* 1833, 3104–3111.
- Sonnhammer, E.L., von Heijne, G., and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings. Int. Conf. Intell. Syst. Mol. Biol.* 6, 175–182.
- Steel, G.J., Brownsword, J., and Stirling, C.J. (2002). Tail-anchored protein insertion into yeast ER requires a novel posttranslational mechanism which is independent of the SEC machinery. *Biochemistry* 41, 11914–11920.
- Stefanovic, S., and Hegde, R.S. (2007). Identification of a Targeting Factor for Posttranslational Membrane Protein Insertion into the ER. *Cell* 128, 1147–1159.
- Stefer, S., Reitz, S., Wang, F., Wild, K., Pang, Y., Schwarz, D., Bomke, J., Hein, C., Löhr, F., Bernhard, F., et al. (2011). Structural basis for tail-anchored membrane protein biogenesis by the Get3-receptor complex. *Science* 333, 758–762.
- Stein, A., Ruggiano, A., Carvalho, P., and Rapoport, T.A. (2014). Key steps in ERAD of luminal ER proteins reconstituted with purified components. *Cell* 158, 1375–1388.
- Suloway, C.J.M., Chartron, J.W., Zaslaver, M., and Clemons, W.M. (2009). Model for eukaryotic tail-anchored protein binding based on the structure of Get3. *Proc. Natl. Acad. Sci. U. S. A.* 106, 14849–14854.
- Suzuki, R., and Kawahara, H. (2016). UBQLN4 recognizes mislocalized transmembrane domain proteins and targets these to proteasomal degradation. *EMBO Rep.* 17, 842–857.
- Trueman, S.F., Mandon, E.C., and Gilmore, R. (2011). Translocation channel gating kinetics balances protein translocation efficiency with signal sequence recognition fidelity. *Mol. Biol. Cell* 22, 2983–2993.
- Tu, L., Wang, J., Helm, A., Skach, W.R., and Deutsch, C. (2000). Transmembrane biogenesis of Kv1.3. *Biochemistry* 39, 824–836.
- Ulbrandt, N.D., Newitt, J.A., and Bernstein, H.D. (1997). The E. coli signal recognition particle is required for the insertion of a subset of inner membrane proteins. *Cell* 88, 187–196.
- Vembar, S.S., and Brodsky, J.L. (2008). One step at a time: endoplasmic reticulum-associated degradation. *Nat. Rev. Mol. Cell Biol.* 9, 944–957.
- Venkatakrishnan, A.J., Deupi, X., Lebon, G., Tate, C.G., Schertler, G.F., and Babu, M.M. (2013). Molecular signatures of G-protein-coupled receptors. *Nature* 494, 185–194.
- Voigt, S., Jungnickel, B., Hartmann, E., and Rapoport, T.A. (1996). Signal sequence-dependent function of the TRAM protein during early phases of protein transport across the endoplasmic reticulum membrane. *J. Cell Biol.* 134, 25–35.
- Voorhees, R.M., and Hegde, R.S. (2015). Structures of the scanning and engaged states of the mammalian srp-ribosome complex. *Elife* 4, 1–21.

## References

- Voorhees, R.M., and Hegde, R.S. (2016a). Structure of the Sec61 channel opened by a signal sequence. *Science* *351*, 88–91.
- Voorhees, R.M., and Hegde, R.S. (2016b). Toward a structural understanding of co-translational protein translocation. *Curr. Opin. Cell Biol.* *41*, 91–99.
- Voorhees, R.M., Fernández, I.S., Scheres, S.H.W., and Hegde, R.S. (2014). Structure of the mammalian ribosome-Sec61 complex to 3.4Å resolution. *Cell* *157*, 1632–1643.
- Walter, P., and Blobel, G. (1981a). Translocation of proteins across the endoplasmic reticulum. II. Signal recognition protein (SRP) mediates the selective binding to microsomal membranes of in-vitro-assembled polysomes synthesizing secretory protein. *J. Cell Biol.* *91*, 551–556.
- Walter, P., and Blobel, G. (1981b). Translocation of proteins across the endoplasmic reticulum III. Signal recognition protein (SRP) causes signal sequence-dependent and site-specific arrest of chain elongation that is released by microsomal membranes. *J. Cell Biol.* *91*, 557–561.
- Walter, P., and Blobel, G. (1983). Preparation of microsomal membranes for cotranslational protein translocation. *Methods Enzymol.* *96*, 84–93.
- Walter, P., and Johnson, A.E. (1994). Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu. Rev. Cell Biol.* *10*, 87–119.
- Walter, P., Ibrahimi, I., and Blobel, G. (1981). Translocation of proteins across the endoplasmic reticulum. I. Signal recognition protein (SRP) binds to in-vitro-assembled polysomes synthesizing secretory protein. *J. Cell Biol.* *91*, 545–550.
- Walter, P., Gilmore, R., and Blobel, G. (1984). Protein translocation across the endoplasmic reticulum. *Cell* *38*, 5–8.
- Wang, P., and Dalbey, R.E. (2011). Inserting membrane proteins: the YidC/Oxa1/Alb3 machinery in bacteria, mitochondria, and chloroplasts. *Biochim. Biophys. Acta* *1808*, 866–875.
- Wang, F., Brown, E.C., Mak, G., Zhuang, J., and Denic, V. (2010). A chaperone cascade sorts proteins for posttranslational membrane insertion into the endoplasmic reticulum. *Mol. Cell* *40*, 159–171.
- Wang, F., Chan, C., Weir, N.R., and Denic, V. (2014). The Get1/2 transmembrane complex is an endoplasmic-reticulum membrane protein insertase. *Nature* *512*, 441–444.
- Wang, Q., Liu, Y., Soetandyo, N., Baek, K., Hegde, R., and Ye, Y. (2011). A ubiquitin ligase-associated chaperone holdase maintains polypeptides in soluble states for proteasome degradation. *Mol. Cell* *42*, 758–770.
- Warne, T., Serrano-Vega, M.J., Baker, J.G., Moukhametzianov, R., Edwards, P.C., Henderson, R., Leslie, A.G.W., Tate, C.G., and Schertler, G.F.X. (2008). Structure of a

## References

- beta1-adrenergic G-protein-coupled receptor. *Nature* 454, 486–491.
- Wessels, H.P., and Spiess, M. (1988). Insertion of a multispanning membrane protein occurs sequentially and requires only one signal sequence. *Cell* 55, 61–70.
- White, S.H., and Von Heijne, G. (2005). Transmembrane helices before, during, and after insertion. *Curr. Opin. Struct. Biol.* 15, 378–386.
- Whiteley, A.M., Prado, M.A., Peng, I., Abbas, A.R., Haley, B., Paulo, J.A., Reichelt, M., Katakam, A., Sagolla, M., Modrusan, Z., et al. (2017). Ubiquitin1 promotes antigen-receptor mediated proliferation by eliminating mislocalized mitochondrial proteins. *Elife* 6.
- Wideman, J.G. (2015). The ubiquitous and ancient ER membrane protein complex (EMC): tether or not? *F1000Research* 4, 624.
- Wiedemann, N., and Pfanner, N. (2017). Mitochondrial Machineries for Protein Import and Assembly. *Annu. Rev. Biochem.* 86, 685–714.
- Wiedmann, M., Kurzchalia, T. V, Hartmann, E., and Rapoport, T.A. (1987). A signal sequence receptor in the endoplasmic reticulum membrane. *Nature* 328, 830–833.
- Wiedmann, M., Goerlich, D., Hartmann, E., Kurzchalia, T. V, and Rapoport, T.A. (1989). Photocrosslinking demonstrates proximity of a 34 kDa membrane protein to different portions of preprolactin during translocation through the endoplasmic reticulum. *FEBS Lett.* 257, 263–268.
- Wild, R., Kowal, J., Eyring, J., Ngwa, E.M., Aebi, M., and Locher, K.P. (2018). Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science* (80-. ). 359, 545–550.
- Wolin, S.L., and Walter, P. (1989). Signal recognition particle mediates a transient elongation arrest of preprolactin in reticulocyte lysate. *J. Cell Biol.* 109, 2617–2622.
- Wright, G., Terada, K., Yano, M., Sergeev, I., and Mori, M. (2001). Oxidative stress inhibits the mitochondrial import of preproteins and leads to their degradation. *Exp. Cell Res.* 263, 107–117.
- Xie, Y., Langhans-Rajasekaran, S.A., Bellovino, D., and Morimoto, T. (1996). Only the first and the last hydrophobic segments in the COOH-terminal third of Na,K-ATPase alpha subunit initiate and halt, respectively, membrane translocation of the newly synthesized polypeptide. Implications for the membrane topology. *J. Biol. Chem.* 271, 2563–2573.
- Xu, Y., Cai, M., Yang, Y., Huang, L., and Ye, Y. (2012). SGTA recognizes a noncanonical ubiquitin-like domain in the Bag6-Ubl4A-Trc35 complex to promote endoplasmic reticulum-associated degradation. *Cell Rep.* 2, 1633–1644.
- Xu, Y., Liu, Y., Lee, J., and Ye, Y. (2013). A ubiquitin-like domain recruits an oligomeric chaperone to a retrotranslocation complex in endoplasmic reticulum-associated degradation. *J. Biol. Chem.* 288, 18068–18076.
- Yabal, M., Brambillasca, S., Soffientini, P., Pedrazzini, E., Borgese, N., and Makarow, M.

## References

- (2003). Translocation of the C terminus of a tail-anchored protein across the endoplasmic reticulum membrane in yeast mutants defective in signal peptide-driven translocation. *J. Biol. Chem.* *278*, 3489–3496.
- Yamagata, A., Mimura, H., Sato, Y., Yamashita, M., Yoshikawa, A., and Fukai, S. (2010). Structural insight into the membrane insertion of tail-anchored proteins by Get3. *Genes to Cells* *15*, 29–41.
- Yamagishi, M., Fujita, H., Morimoto, F., Kida, Y., and Sakaguchi, M. (2011). A sugar chain at a specific position in the nascent polypeptide chain induces forward movement during translocation through the translocon. *J. Biochem.* *149*, 591–600.
- Yamaguchi, A., Hori, O., Stern, D.M., Hartmann, E., Ogawa, S., and Tohyama, M. (1999). Stress-associated endoplasmic reticulum protein 1 (SERP1)/Ribosome-associated membrane protein 4 (RAMP4) stabilizes membrane proteins during stress and facilitates subsequent glycosylation. *J. Cell Biol.* *147*, 1195–1204.
- Yamamoto, K., Hayashishita, M., Minami, S., Suzuki, K., Hagiwara, T., Noguchi, A., and Kawahara, H. (2017). Elimination of a signal sequence-uncleaved form of defective HLA protein through BAG6. *Sci. Rep.* *7*, 14545.
- Ye, Y., Shibata, Y., Yun, C., Ron, D., and Rapoport, T.A. (2004). A membrane protein complex mediates retro-translocation from the ER lumen into the cytosol. *Nature* *429*, 841–847.
- Zalisko, B.E., Chan, C., Denic, V., Rock, R.S., and Keenan, R.J. (2017). Tail-Anchored Protein Insertion by a Single Get1/2 Heterodimer. *Cell Rep.* *20*, 2287–2293.
- Zhang, Z.-R., Bonifacino, J.S., and Hegde, R.S. (2013). Deubiquitinases sharpen substrate discrimination during membrane protein degradation from the ER. *Cell* *154*, 609–622.
- Zhao, G., and London, E. (2006). An amino acid “transmembrane tendency” scale that approaches the theoretical limit to accuracy for prediction of transmembrane helices: Relationship to biological hydrophobicity. *Protein Sci.* *15*, 1987–2001.
- Zhao, G., and London, E. (2009). Strong correlation between statistical transmembrane tendency and experimental hydrophobicity scales for identification of transmembrane helices. *J. Membr. Biol.* *229*, 165–168.
- Zhu, L., Kaback, H.R., and Dalbey, R.E. (2013). YidC protein, a molecular chaperone for LacY protein folding via the SecYEG protein machinery. *J. Biol. Chem.* *288*, 28180–28194.
- Zhu, X., Zhao, X., Burkholder, W.F., Gragerov, A., Ogata, C.M., Gottesman, M.E., and Hendrickson, W.A. (1996). Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* *272*, 1606–1614.

## Publications

**Guna, A.**, Volkmar N., Christianson, J.C., and Hegde, R.S. (2017). The ER membrane protein complex is a transmembrane domain insertase. *Science*, 359(6374): 470-473.

**Guna, A.**, Hegde, R.S. Recognition of transmembrane domains during protein biogenesis and quality control. *Current Biology*, 28(8): R498-511 (Review).

Chitwood, P.J., Juskiewicz, S., **Guna, A.**, Shao, S., Hegde, R.S. EMC is required to initiate accurate membrane protein topogenesis. (under review).

## Appendix 1: siRNA screen results

#	Gene	Full Gene Name	RefSeq AN	siRNA ID	siRNA	GFP:RFP
1	SEC61A1	Sec61 translocon alpha 1 subunit	NM_013336	s26721	A	1.02
				s26722	B	1.59
				s26723	C	1.15
2	SEC61B	Sec61 translocon beta subunit	NM_006808	s21552	A	1.37
				s21553	B	1.30
				s21554	C	1.65
3	SEC61G	Sec61 translocon gamma subunit	NM_001012456	s23910	A	1.03
				s23911	B	1.51
				s23912	C	1.19
4	SEC61A2	Sec61 translocon alpha 2 subunit variant 1	NM_018144	s30355	A	1.88
				s30356	B	1.34
				s30357	C	1.27
5	SEC62	SEC62 homolog (S. cerevisiae) preprotein translocation factor	NM_003262	s14188	A	1.69
				s14189	B	1.15
				s14190	C	1.86
6	SEC63	SEC63 homolog (S. cerevisiae) protein translocation regulator	NM_007214	s22166	A	1.24
				s22167	B	1.15
				s22168	C	1.27
7	TRAM1	translocation associated membrane protein 1	NM_014294	s23889	A	1.80
				s23890	B	1.14
				s23891	C	1.14
8	TRAM2	translocation associated membrane protein 2	NM_012288	s18677	A	1.73
				s18678	B	1.56
				s18679	C	1.98
9	SSR1	signal sequence receptor subunit alpha (translocon-associated protein alpha)	NM_003144	s13477	A	0.71
				s13478	B	0.68
				s13479	C	0.72
10	SSR2	signal sequence receptor subunit beta (translocon-associated protein beta)	NM_003145	s13480	A	1.54
				s13481	B	1.45
				s13482	C	1.08
11	SSR3	signal sequence receptor subunit gamma (translocon-associated protein gamma)	NM_007107	s13483	A	1.03
				s13484	B	1.65
				s13485	C	0.93
12	SSR4	signal sequence receptor subunit delta (translocon-associated protein delta)	NM_006280	s13486	A	1.08
				s13487	B	1.28
				s13488	C	1.10
13	RRBP1	ribosome binding protein 1 homolog 180kDa	NM_001042576	s12351	A	1.36
				s12352	B	0.92
				s12353	C	0.96
14	SEC11A	SEC11 homolog A (S. cerevisiae), signal peptidase complex subunit	NM_014300	s23904	A	0.76
				s23905	B	1.10
				s23905	C	0.68
15	SEC11C		NM_033280	s40479	A	1.59
				s40480	B	1.22

Appendix 1: siRNA screen results

		SEC11 homolog C ( <i>S. cerevisiae</i> ), signal peptidase subunit		s40481	C	1.37
16	SPCS1	signal peptidase complex subunit 1 homolog ( <i>S. cerevisiae</i> )	NM_014041	s26327 s26328 s26329	A B C	1.07 0.96 1.07
17	SPCS2	signal peptidase complex subunit 2 homolog ( <i>S. cerevisiae</i> )	NM_014752	s18918 s18919 s18920	A B C	0.91 0.84 0.84
18	SPCS3	signal peptidase complex subunit 3 homolog ( <i>S. cerevisiae</i> )	NM_021928	s34132 s34133 s34134	A B C	0.89 0.89 0.89
19	KIAA090	ER membrane protein complex subunit 1	NM_015047	s22953 s22954 s22955	A B C	0.85 0.77 0.94
20	TTC35	ER membrane protein complex subunit 2	NM_014673	s18668 s18669 s18670	A B C	1.03 1.10 1.19
21	TMEM111	ER membrane protein complex subunit 3	NM_018447	s31613 s31614 s31615	A B C	1.04 1.04 1.17
22	TMEM85	ER membrane protein complex subunit 4	NM_016454	s27733 s27734 s27735	A B C	0.97 0.96 0.91
23	MMGT1	ER membrane protein complex subunit 5	NM_173470	s41129 s41130 s41131	A B C	0.96 0.87 1.06
24	TMEM93	ER membrane protein complex subunit 6	NM_001014764	s37940 s37941 s37942	A B C	0.97 0.96 1.36
25	C15orf24	ER membrane protein complex subunit 7	NM_020154	s32292 s32293 s32294	A B C	0.88 0.98 1.10
26	COX4NB	ER membrane protein complex subunit 8	NM_006067	s20200 s20201 s20202	A B C	1.00 0.95 1.17
27	FAM158A	ER membrane protein complex subunit 9	NM_016049	s27244 s27245 s27246	A B C	1.06 0.95 1.14
28	C19orf63	ER membrane protein complex subunit 10	NM_175063	s49610 s49611 s49612	A B C	1.39 1.27 1.10
29	TMED1	transmembrane emp24 protein transport domain containing 1	NM_006858	s21699 s21700 s195083	A B C	1.59 1.49 1.35
30	TMED2	transmembrane emp24 domain trafficking protein 2	NM_006815	s21570 s21571 s21572	A B C	1.24 1.27 1.22
31	TMED3		NM_007364	s23799 s23800	A B	0.91 0.94

Appendix 1: siRNA screen results

		transmembrane emp24 protein transport domain containing 3		s23801	C	1.11
32	TMED4	transmembrane emp24 protein transport domain containing 4	NM_182547	s48156 s48157 s48158	A B C	0.92 1.14 1.03
33	TMED5	transmembrane emp24 protein transport domain containing 5	NM_016040	s27202 s27203 s27204	A B C	1.00 1.14 0.97
34	TMED6	transmembrane emp24 protein transport domain containing 6	NM_144676	s44861 s44862 s44863	A B C	1.18 1.24 0.96
35	TMED7	transmembrane emp24 protein transport domain containing 7	NM_181836	s27238 s27239 s27240	A B C	1.45 1.25 1.59
36	TMED8	transmembrane emp24 protein transport domain containing 8	NM_213601	s49298 s49299 s49300	A B C	1.78 1.41 2.22
37	TMED9	transmembrane emp24 protein transport domain containing 9	NM_017510	s29353 s29354 s29355	A B C	1.55 1.32 1.36
38	TMED10	transmembrane emp24-like trafficking protein 10 (yeast)	NM_006827	s21600 s21601 s21602	A B C	1.53 1.51 1.38
39	DERL1	Der1-like domain family, member 1	NM_024295	s35605 s35606 s35607	A B C	1.21 1.64 2.19
40	DERL2	Der1-like domain family, member 2	NM_016041	s27226 s27227 s27228	A B C	1.23 1.16 1.17
41	DERL3	Der1-like domain family, member 3	NM_198440	s40621 s40622 s230190	A B C	1.06 1.19 1.20
42	SEL1L	SEL1L ERAD ER ligase adaptor subunit	NM_005065	s12674 s12675 s12676	A B C	1.06 1.41 1.23
43	SEL1L2	sel-1 suppressor of lin-12- like 2 (C. elegans)	XM_046437	s37261 s37262 s37263	A B C	1.12 1.48 2.58
44	KIAA0746	SEL1L family member 3	NM_015187	s23343 s23344 s23345	A B C	1.32 1.16 1.26
45	C1orf163	chromosome 1 open reading frame 163	NM_023077	s35263 s35264 s35265	A B C	1.32 1.24 1.64
46	ERLIN2	ER lipid raft associated 2	NM_001003790	s22014 s22015 s22016	A B C	1.12 1.59 1.43
47	ERLIN1	ER lipid raft associated 1	NM_006459	s20839 s20840	A B	1.16 1.29

Appendix 1: siRNA screen results

				s20841	C	1.41
48	SYVN1	synovial apoptosis inhibitor 1, yeast Hrd1 homolog	NM_032431	s39019	A	1.35
				s39020	B	2.10
				s39021	C	1.24
49	RNF170	ring finger protein 170	NM_030954	s37785	A	1.10
				s37786	B	1.02
				s37787	C	1.19
50	BCAP31	B-cell receptor-associated protein 31	NM_005745	s19721	A	1.31
				s19722	B	1.13
				s19723	C	1.12
51	SEC22B	SEC22 vesicle trafficking protein homolog B (S. cerevisiae)	NM_004892	s18347	A	2.12
				s18348	B	1.35
				s18349	C	1.54
52	USE1	unconventional SNARE in the ER 1 homolog (S. cerevisiae)	NM_018467	s31658	A	1.31
				s31659	B	1.27
				s31660	C	1.71
53	STX18	syntaxin 18	NM_016930	s28733	A	1.38
				s28734	B	1.42
				s28735	C	1.69
54	BNIP1	BCL2/adenovirus E1B 19kDa interacting protein 1	NM_013980	s2053	A	1.38
				s2054	B	1.40
				s2055	C	1.25
55	MIA3	melanoma inhibitory activity family, member 3, ER export factor	NM_198551	s51657	A	1.69
				s51658	B	1.39
				s51659	C	1.20
56	ERGIC1	endoplasmic reticulum-golgi intermediate compartment (ERGIC) 1	NM_001031711	s32914	A	1.24
				s195318	B	1.54
				s195319	C	1.15
57	ERGIC2	ERGIC and golgi 2	NM_016570	s27857	A	1.02
				s27858	B	1.10
				s27859	C	1.46
58	ERGIC3	ERGIC and golgi 3	NM_015966	s28410	A	1.36
				s28411	B	1.13
				s28412	C	1.09
59	LMAN1	lectin, mannose-binding, 1	NM_005570	s8218	A	0.93
				s8219	B	1.15
				s8220	C	1.13
60	GOLT1B	golgi transport 1 homolog B (S. cerevisiae)	NM_016072	s27270	A	1.29
				s27271	B	1.02
				s27272	C	1.31
61	ATP2A2	ATPase, Ca <sup>++</sup> transporting, cardiac muscle, slow twitch 2	NM_170665	s1746	A	1.15
				s1747	B	1.13
				s1748	C	1.11
62	HM13	histocompatibility (minor) 13	NM_030789	s37580	A	1.20
				s195419	B	1.13
				s195420	C	1.29
63	FKBP11	FK506 binding protein 11, 19 kDa	NM_016594	s27896	A	1.31
				s27897	B	1.32
				s27898	C	1.28

## Appendix 1: siRNA screen results

64	SURF4	surfeit 4	NM_033161	s13649	A	1.33
				s13650	B	1.23
				s13651	C	1.56
65	CCDC47	coiled-coil domain containing 47	NM_020198	s32576	A	1.21
				s32577	B	1.18
				s32578	C	1.42
66	TMUB1	transmembrane and ubiquitin-like domain containing 1	NM_031434	s38013	A	1.25
				s38014	B	0.96
				s195441	C	1.12
67	ATP13A1	ATPase type 13A1	NM_020410	s32748	A	1.48
				s32749	B	1.08
				s32750	C	1.17
68	PGRMC1	progesterone receptor membrane component 1	NM_006667	s21310	A	1.30
				s21311	B	1.19
				s21312	C	1.20
69	PGRMC2	progesterone receptor membrane component 2	NM_006320	s704	A	1.02
				s705	B	1.19
				s706	C	1.19
70	MAGT1	magnesium transporter 1	NM_032121	s38423	A	1.12
				s38424	B	1.04
				s38425	C	1.14
71	SLC33A1	solute carrier family 33 (acetyl-CoA transporter), member 1	NM_004733	s17578	A	1.10
				s17579	B	1.14
				s17580	C	1.12
72	TP53I11	tumour protein p53 inducible protein 11	NM_006034	s18308	A	1.57
				s18309	B	1.49
				s18310	C	1.24
73	SLC37A4	solute carrier family 37 (glucose-6-phosphate transporter), member 4	NM_001467	s5449	A	1.65
				s5450	B	1.33
				s5451	C	1.18
74	C1orf9	chromosome 1 open reading frame 9	NM_016227	s28126	A	1.13
				s28127	B	1.05
				s28128	C	1.18
75	TMEM214	transmembrane protein 214	NM_001083590	s29626	A	1.16
				s29627	B	1.42
				s29628	C	1.45
76	JAGN1	jagunal homolog 1 (Drosophila)	NM_032492	s39100	A	1.22
				s39101	B	1.65
				s39102	C	1.02
77	SLMAP	sarcolemma associated protein	NM_007159	s15433	A	1.32
				s15434	B	1.17
				s15435	C	1.23
78	TXNDC14	thioredoxin domain containing 14	NM_015959	s27362	A	1.50
				s27363	B	1.20
				s27364	C	1.28
79	NOMO1	NODAL modulator 1	NM_014287	s23793	A	1.33
				s23794	B	1.00
				s23795	C	1.37
80	NOMO2	NODAL modulator 2	NM_001004060	s49378	A	1.15

Appendix 1: siRNA screen results

				s49379	B	1.52
				s49380	C	1.35
81	PGAP1	post-GPI attachment to proteins 1	NM_024989	s36870	A	1.23
				s36871	B	1.37
				s36872	C	1.31
82	LRRC59	leucine rich repeat containing 59	NM_018509	s30849	A	1.46
				s30850	B	1.60
				s30851	C	1.64
83	TMTC3	transmembrane and tetratricopeptide repeat containing 3	NM_181783	s46197	A	1.31
				s46198	B	1.31
				s46199	C	1.29
84	TMTC4	transmembrane and tetratricopeptide repeat containing 4	NM_001079669	s39610	A	1.16
				s39611	B	1.35
				s39612	C	1.67
85	ATL3	atlastin GTPase 3	NM_015459	s24754	A	1.24
				s24755	B	1.13
				s24756	C	1.43
86	KIAA1715	ER junction fotation factor lunapark	NM_030650	s37454	A	1.33
				s37455	B	1.29
				s37456	C	1.13
87	TMEM66	store-operated calcium entry associated regulatory factor SARAF	NM_016127	s28509	A	1.23
				s28510	B	1.25
				s28511	C	1.61
88	WFS1	wolframin ER transmembrane glycoprotein	NM_006005	s14856	A	1.21
				s14857	B	1.20
				s14858	C	1.03
89	ZMPSTE24	zinc metalloproteinase (STE24 homolog, S. cerevisiae)	NM_005857	s20065	A	1.07
				s20066	B	1.23
				s20067	C	1.30
90	CLPTM1	cleft lip and palate associated transmembrane protein 1	NM_001294	s3185	A	1.20
				s3186	B	1.24
				s3187	C	1.13
91	ATP1A1	ATPase, Na <sup>+</sup> /K <sup>+</sup> transporting, alpha 1 polypeptide	NM_001001586	s1718	A	1.43
				s1719	B	1.09
				s1720	C	1.83
92	RDH11	retinol dehydrogenase 11 (all-trans/9-cis/11-cis)	NM_016026	s27455	A	1.61
				s27456	B	1.33
				s27457	C	1.28
93	RDH12	retinol dehydrogenase 12 (all-trans/9-cis/11-cis)	NM_152443	s44708	A	1.23
				s44709	B	1.25
				s44710	C	1.40
94	ITPR1	inositol 1,4,5-triphosphate receptor, type 1	NM_002222	s7631	A	1.15
				s7632	B	1.33
				s7633	C	1.23
95	ITPR2	inositol 1,4,5-trisphosphate receptor, type 2	NM_002223	s7634	A	1.18
				s7635	B	1.05
				s7636	C	1.45
96	TMEM97	transmembrane protein 97	NM_014573	s26204	A	1.12
				s26205	B	1.12

Appendix 1: siRNA screen results

				s26206	C	1.12
97	REEP5	receptor accessory protein 5	NM_005669	s15454	A	1.15
				s15455	B	1.20
				s15456	C	1.12
98	REEP4	receptor accessory protein 4	NM_025232	s37270	A	1.32
				s37271	B	1.27
				s37272	C	1.02
99	TMCO1	transmembrane and coiled-coil domains 1	NM_019026	s29085	A	1.28
				s29086	B	1.10
				s29087	C	1.28
100	FITM2	fat storage-inducing transmembrane protein 2	NM_001080472	s43312	A	1.32
				s43313	B	1.16
				s43314	C	1.12
101	ABCD3	ATP-binding cassette, sub-family D (ALD), member 3	NM_001122674	s229943	A	1.19
				s229944	B	1.43
				s229945	C	1.02
102	ABCD4	ATP-binding cassette, sub-family D (ALD), member 4	NM_005050	s11618	A	1.15
				s11619	B	1.10
				s11620	C	0.97
103	EI24	etoposide induced 2.4 mRNA	NM_001007277	s18311	A	1.09
				s18312	B	1.18
				s18313	C	1.23
104	SLC39A7	solute carrier family 39 (zinc transporter), member 7	NM_001077516	s15478	A	1.13
				s15479	B	1.99
				s15480	C	1.22
105	ANO10	anoctamin 10	NM_018075	s30236	A	1.02
				s30237	B	1.05
				s30238	C	1.02
106	EPHX1	epoxide hydrolase 1, microsomal (xenobiotic)	NM_000120	s4749	A	1.07
				s4750	B	1.40
				s4751	C	1.10
107	KTN1	kinectin 1 (kinesin receptor)	NM_001079521	s8033	A	1.16
				s8034	B	1.09
				s8035	C	1.06
108	LMF2	lipase maturation factor 2	NM_033200	s40609	A	1.19
				s40610	B	1.12
				s40611	C	1.26
109	TECR	trans-2,3-enoyl-CoA reductase	NM_138501	s18269	A	1.15
				s18270	B	1.17
				s18271	C	1.49
110	HSD17B12	hydroxysteroid (17-beta) dehydrogenase 12	NM_016142	s27545	A	1.65
				s27546	B	1.19
				s27547	C	1.15
111	HSD17B2	hydroxysteroid (17-beta) dehydrogenase 2	NM_002153	s6944	A	1.18
				s6945	B	1.16
				s6946	C	1.16
112	SGPL1	sphingosine-1-phosphate lyase 1	NM_003901	s16963	A	1.11
				s16964	B	1.21
				s16965	C	1.26

Appendix 1: siRNA screen results

113	LPCAT3	lysophosphatidylcholine acyltransferase 3	NM_005768	s19799 s19800 s19801	A B C	1.21 1.18 1.39
114	AGPAT2	1-acylglycerol-3-phosphate O-acyltransferase 2 (lysophosphatidic acid acyltransferase, beta)	NM_001012727	s20702 s20703 s223130	A B C	1.32 1.36 1.52
115	CDIPT	CDP-diacylglycerol--inositol 3-phosphatidyltransferase	NM_006319	s20387 s20388 s20389	A B C	1.22 1.23 1.80
116	SPTLC1	serine palmitoyltransferase, long chain base subunit 1	NM_006415	s20710 s20711 s20712	A B C	1.85 1.53 1.48
117	SIGMAR1	sigma non-opioid intracellular receptor 1	NM_005866	s20086 s20087 s20088	A B C	2.69 1.43 1.26
118	DHCR24	24-dehydrocholesterol reductase	NM_014762	s4079 s4077 s4078	A B C	1.29 1.43 1.14
119	RPN1	ribophorin I	NM_002950	s12246 s12247 s12248	A B C	1.21 1.68 1.16
120	RPN2	ribophorin II	NM_001135771	s230093 s230094 s230095	A B C	1.05 0.81 1.07
121	DDOST	dolichyl-diphosphooligosaccharide-protein glycosyltransferase	NM_005216	s3998 s3999 s4000	A B C	1.10 1.71 1.10
122	DAD1	defender against cell death 1	NM_001344	s3902 s3903 s229514	A B C	1.26 1.12 0.99
123	OSTC	oligosaccharyltransferase complex subunit	NM_021227	s33882 s33883 s33884	A B C	0.97 0.86 0.89
124	KRTCAP2	keratinocyte associated protein 2	NM_173852	s47206 s47207 s195838	A B C	0.86 1.08 1.00
125	STT3A	STT3, subunit of the oligosaccharyltransferase complex, homolog A (S. cerevisiae)	NM_152713	s7616 s7617 s7618	A B C	1.06 0.99 1.11
126	STT3B	TT3, subunit of the oligosaccharyltransferase complex, homolog B (S. cerevisiae)	NM_178862	s47380 s47381 s47382	A B C	1.32 1.47 1.78
127	MLE	malectin	NM_014730	s18843 s18844 s18845	A B C	1.23 1.11 1.23

## Appendix 1: siRNA screen results