이학박사학위논문

진핵세포에서 translocon을 통한 막단백질의 삽입과정 연구

Translocon-mediated insertion of membrane proteins in eukaryotic cells

2016 년 2 월

서울대학교 대학원

생명과학부

이 헌 상

Translocon-mediated insertion of membrane proteins in eukaryotic cells

by

Hunsang Lee

Under the supervision of

Professor Hyun Ah Kim, Ph.D.

A thesis submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

February 2016

School of Biological Sciences

Seoul National University

ABSTRACT

Translocon-mediated insertion of membrane proteins in eukaryotic cells

Hunsang Lee

School of Biological Sciences

Seoul National University

Membrane proteins contribute up to 30% of the proteome and are major targets for therapeutics. They transduce signals and transport macromolecules across the membrane. The structural information is lacking for most membrane proteins, despite that it is required to understand their function and to design appropriate drugs that target them, due to the difficulties in crystallization and a limited number of experimental tools.

The topology (2-dimensional structure) of membrane proteins are predicted by numerous bioinformatics programs, but it must be validated *in vivo*. However, the eukaryotic system lacks the topology reporter that allows live-cell assessment. In this thesis, glycosylatable GFP (gGFP) was developed as a novel topology reporter to deduce the topology of membrane proteins in yeast (*Saccharomyces cerevisiae*) and mammalian cells. gGFP was made by introducing a sequon (N-linked glycosylation site, N-X-T/S) near the fluorophore. gGFP was non-glycosylated and fluorescent in the cytosol, but became glycosylated and non-fluorescent in the ER lumen. Hence, the fluorescence and the glycosylation status provide the direct evidence of the localization of gGFP, allowing a rapid screening of membrane protein topology.

Membrane proteins adopt the correct topology during the biogenesis. They are inserted into the lipid bilayer through translocon complexes. The key subunits of translocon complexes are reported, however, how a translocon recognizes a transmembrane segment (TMS), mediates membrane insertion and determine final topology is not fully understood. The thesis aimed to understand the mechanism of translocon mediated membrane protein insertion at the endoplasmic reticulum (ER) and mitochondrial inner membrane (IM).

The SEC61 complex mediates membrane protein insertion at the ER. When a TMS enters the translocon, Sec61 opens laterally towards the lipid bilayer at the interface between TMS2 and 7 (lateral gating helices) and allows the partitioning of a TMS into the membrane. Previous mutational analysis on Sec61 suggested that the insertion process is not a pure thermodynamic event, rather Sec61 is actively involved in the insertion process. To provide further insight into the opening and closing of Sec61 via the lateral gating helices, a systematic mutagenesis on TMS2 and 7 was performed in yeast. In the study, two groups of residues that either favor the open or closed conformation were identified. Compared to yeast SEC61, mammalian SEC61 complex contains additional subunits. To extend the investigation to mammalian system by characterizing the roles of different subunits of mammalian SEC61, the same set of model signal anchor proteins and multi-spanning membrane proteins used in the yeast study were expressed in HEK-293T and HeLa cells.

Unlike the SEC61 complex where the lateral gating helices of Sec61 regulate membrane protein insertion to the ER, how the TIM23 complex mediates membrane protein insertion into the mitochondrial IM remain elusive. Mgr2, a subunit of TIM23 complex, was termed as a “lateral gate keeper” as its expression level directly affected the insertion of Cyb2-DHFR and Mgm1. To test whether Mgr2 acts as a general gate keeper of the TIM23 complex and sets the hydrophobicity requirement for protein insertion, model mitochondrial IM proteins were expressed. Neither the insertion of other mitochondrial IM proteins nor the hydrophobicity requirement was altered at different Mgr2 expression levels. Thus, Mgr2’s role in the gate keeping of TIM23 may be specific for Cyb2-DHFR and Mgm1.

Some mitochondrial IM proteins are inserted into the membrane directly by the TIM23 (stop-transfer pathway), whereas the others are sorted to the matrix first and inserted from the matrix side (conservative sorting pathway). Detailed bioinformatics analysis revealed that conservative sorting proteins tend to carry a proline residue in the TMS. To investigate the molecular mechanism of TMS recognition by Tim23, with a particular interest in how it discriminates a TMS with a proline, a screening scheme was designed and validated for the selection of Tim23 with an enhanced tolerance for a proline residue.

Keywords: Membrane proteins, topology, GFP, glycosylation, ER, mitochondria, translocon, translocase.

Student number: 2010-20330

CONTENTS

LIST OF FIGURES AND TABLES

LIST OF ABBREVIATIONS

IM	inner membrane
OM	outer membrane
TMS	transmembrane segment
ER	endoplasmic reticulum
GFP	green fluorescent protein
PhoA	phosphatase A
OST	oligosaccaryl transferase
FPP	fluorescence protease protection
RNC	ribosome nascent chain complex
SRP	signal recognition particle
SR	signal recognition particle receptor
TRAP	translocon associated protein complex
TRAM	translocating chain associating membrane protein
TOM	translocase of the OM
IMS	intermembrane space
SAM	sorting and assembly machinery
MIA	mitochondrial IMS assembly
TIM	translocase of the IM
MPP	matrix processing peptidase

IMP	IM peptidase
Cyb2	cytochrome b2
gGFP	glycosylatable GFP
Lep	leader peptidase
H-segment	hydrophobic segment
SAS	signal anchor sequence
MFP	Mgm1 fusion protein
CCHL	cytochrome c heme lyase
FoA	5-fluoroorotic acid
DMSO	dimethyl sulfoxide
Endo H	endoglycosidase H
HRP	horseradish peroxidase
FITC	fluoresceinisothiocyanate
yEGFP	yeast enhanced GFP
HA	hemagglutinin
WT	wild-type
SP	signal peptide
UTR	untranslated region

CHAPTER I – INTRODUCTION

I.1. Membrane proteins and topology

I.1.1. Types of membrane proteins

Membrane proteins constitute about 30% of the proteome and deserves much attention as major targets for therapeutics[1-3]. Membrane proteins are classified into two main categories, peripheral and integral membrane proteins. Peripheral membrane proteins are associated with the lipid bilayer via an anchoring motif that does not span the membrane[4,5]. In contrast, integral membrane proteins adopt either α-helical or β-barrel conformation to be embedded in the hydrophobic lipid bilayer[6]. Integral membrane proteins function as channels or receptors to carry out vital cellular processes, namely signal transduction and transport of macromolecules[7]. At the plasma membrane or mitochondrial inner membrane (IM), where cells form a boundary from the extracellular space or carries out respiration, respectively, many membrane proteins adopt α-helical conformation[8]. In contrast, β-barrel proteins are abundant in bacterial or mitochondrial outer membrane (OM)[9].

I.1.2. Topology of membrane proteins

α-helical membrane proteins are sorted to different types based on the topology, where topology refers to a 2-dimensional structural information of the number of transmembrane segments (TMS) and the orientation of soluble loops relative to the plane of the membrane[10]. Single-spanning membrane proteins are classified into type I or II based on their location of the N-terminus. Multi-spanning membrane proteins are classified as polytopic[11] (Figure 1).

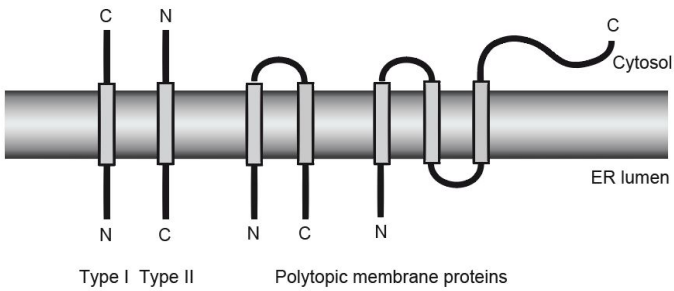


Figure 1. Topology of membrane proteins. Single-spanning membrane proteins are classified based on the orientation relative to the membrane. Type I proteins have the N-terminus in the endoplasmic reticulum (ER) lumen whereas type II proteins have the N-terminus in the cytosol. Multi-spanning membrane proteins are classified as polytopic.

Membrane proteins must adopt correct topology for their biological function and many factors dictate the topology of membrane proteins. Membrane proteins obtain their topology during the biogenesis. Firstly, charged residues on a nascent chain are key determinants of membrane protein topology. Positively charged residues are preferentially positioned at the inside (cytoplasmic side) of the cell for various reasons (positive inside rule)[12,13]. During the protein biogenesis, the exposed positively charged residues of a nascent chain may interact with negatively charged phospholipid head groups or the translocation machinery may disfavor translocation of positively charged residues. The negative membrane potential in the cytosolic side of the membrane may also influence retaining positively charged residues of a nascent chain[14] (Figure 2). Secondly, the components of the translocon complexes, which provide a passage for a nascent polypeptide to cross or insert themselves into the membrane, influences the orientation of membrane proteins[15-17]. Mutations in the translocon complexes can favor the insertion of membrane protein in certain orientation over another orientation[18,19]. Other factors that influence topogenesis of membrane proteins include hydrophobicity, length of TMS, and the lipid composition. Hydrophobic, long TMS are preferrably inserted in N_{out} forms and the lipid composition was shown to be the key determinant for topogenesis of LacY[20-23].

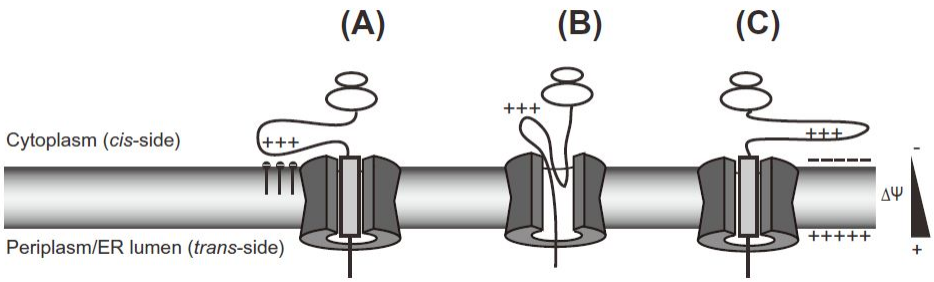


Figure 2. Some explanations for the positive inside rule. During co-translational translocation and membrane insertion from the *cis* side of the membrane, the exposed positively charged residues of a nascent chain may interact with negatively charged phospholipid head groups (A) or the translocation machinery may disfavor translocation of positively charged residues (B). The negative membrane potential in the cytosolic side of the membrane may also influence retaining positively charged residues of a nascent chain (C)[14].

I.1.3. Topology reporters

Topology reporters are biochemical tags that are fused to or engineered into the protein of interest in order to deduce the topology. There are a number of topology reporters available for both prokaryotic and eukaryotic systems[24]. For bacteria, green fluorescent protein (GFP) - phosphatase A (PhoA) fusion assay is widely used where GFP or PhoA is fused to a membrane protein. Due to the oxidizing environment of the periplasmic side, GFP is only fluorescent in the cytosol whereas PhoA is only active in the periplamic side[25,26]. Thus, two systems complement one another for a validation of membrane protein topology in bacteria.

For eukaryotic cells, glycosylation mapping is often used to determine the topology of membrane proteins entering the secretory pathway[27,28]. Glycosylation mapping takes an advantage of the spacial specificity of N-linked glycosylation at the endoplasmic reticulum (ER) lumen. A protein carrying an amino sequence of N-X-T/S (sequon, where X can be any amino acid except P) is glycosylated by oligosaccaryl transferase (OST) only when the sequon is exposed to the luminal side of the ER[29,30]. Glycosylated status of a sequon is an indicator of ER luminal localization whereas non-glycosylated status refers to the cytosolic localization. Glycosylation status of a protein is monitored by SDS-PAGE as every glycan added to a polypeptide results in a 2kDa increase in the protein mass, slowing down the mobility of a protein on a gel.

Lastly, fluorescence protease protection (FPP) assay is used to solve the protein topology of not only the secretory pathway proteins, but also proteins of the other subcellular organelles[31-33]. The assay requires the fusion of a membrane protein with a fluorescent protein, followed by a protease digestion and fluorescence measurement. The rationale is that if the part of a membrane protein where a fluorescent protein is tagged to is exposed to the protease accessible side (outside of an organelle or a cell), it would be degraded by the externally added protease, resulting in the loss of fluorescence. On the other hand, if the fluorescent protein is in the inside of an organelle or a cell, it would be protected from the added protease and remain fluorescent.

I.2. Protein targeting and insertion into the ER membrane

I.2.1. Co- and post-translational translocation pathways

The proteins sorted in the secretory pathway, both soluble and integral membrane proteins, are initially targeted to the ER membrane either by co-translational or post-translational pathway[34,35]. In co-translational pathway, once the signal sequence emerges from the ribosome, the ribosome nascent chain complex (RNC) is recognized by the signal recognition particle (SRP)[36]. SRP interacts with the SRP receptor (SR) at the ER membrane and hand the RNC over to the translocon complex[37]. In post-translational pathway, the proteins are fully synthesized in cytosol first prior to the targeting to the ER membrane[38]. Here, the cytosolic chaperones, like Hsp70, maintain the proteins in an unfolded state for the translocation across the membrane via the translocon complex[39].

I.2.2. SEC61 translocon complex

The translocon complex provides a channel for a nascent chain to cross or insert into the lipid bilayer (Figure 3). The main component of the ER translocon complex is the pore forming Sec61[40,41]. In yeast, depending on the targeting pathways, Sec61 interacts with a specific set of proteins to form distinct translocon complexes. Sec61 interacts with Sec62, Sec63, Sec71 and Sec72 to form post-translational translocon whereas Sec61's interactions with Sec62, Sec71 and Sec72 are dispensable for the formation of co-translational translocon[42-48]. Deletion of Sec61, Sec62 or Sec63 is lethal in yeast implying that the proper assembly of the translocon complex for the protein translocation and membrane insertion of various substrates is essential for cell viability[49,50].

In humans, most secretory proteins take co-translational pathway. Only the small secretory proteins, composed of less than 120 residues, are reported to take post-translational pathway as they are prone to complete the translation prior to the recognition by SRP thereby evading co-translational targeting[51]. The main channel Sec61α as well as Sec62 and Sec63 are also present and reported to interact with one another in humans[52,53].

In contrast to the yeast deletion studies, depletion of Sec62 or Sec63 does not affect the viability of mammalian tissue-cultured cells[54]. Here, Sec62 is involved in the post-translational transport of small secretory proteins whereas Sec63 plays a role in the initial phase of co-translational transport for certain substrates and modulates the steady state levels of multi-spanning membrane proteins[54,55]. There are additional components of the translocon complex in humans other than the Sec61, Sec62 and Sec63. These include ERj1, TRAP (Translocon associated protein complex) and TRAM1 (Translocating chain associating membrane protein 1). ERj1 is a J-domain containing protein and it can functionally complement Sec63 deletion in yeast[56]. TRAP is required for efficient translocation of secretory proteins with a weak signal sequence and it is involved in regulating the positive inside rule[15,57]. TRAM1 participates in an early step of protein transport and regulates the exposure of a nascent chain to the cytosol during the biogenesis[58-62].

Both yeast and human share the core channel forming subunit of the translocon, Sec61[48]. It has 10 TMS and opens laterally towards the lipid bilayer to allow the partitioning of TMS via its TMS2 and 7[63,64] (Figure 4). The key residues in modulating the gate opening and closing were reported sporadically, however, the systematic analysis of all residues lining the gating helices and how they handle substrates of different hydrophobicity is not well reported[65].

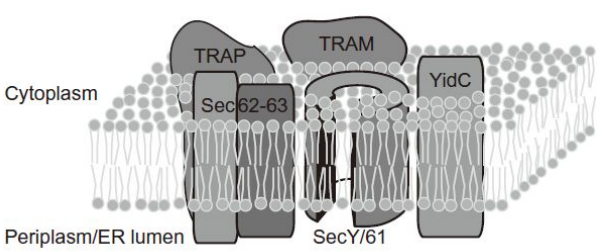


Figure 3. Translocation components involved in the membrane protein biogenesis. The Sec61/SecY translocon associates with the Sec62-Sec63 complex in yeast, the TRAM and TRAP complexes in mammalian cells, and YidC in *Escherichia coli* mediate membrane protein biogenesis[14].

I.2.3. Determinants of ER membrane protein insertion

In general, for a TMS to be in the membrane, it needs to take α-helical shape that shields the hydrophilic side chains of the amino acids from the hydrophobic lipid tails[66]. Such TMS is recognized by Sec61 and laterally released into the membrane[67]. A typical TMS is composed of about 20 hydrophobic amino acids with charged residues being disfavored[68]. The biological hydrophobicity scale, a contribution of individual amino acids in a TMS insertion into the ER membrane, was reported with isoleucine being one of the most hydrophobic residue and aspartic acid being the most hydrophilic residue[69-72].

Membrane protein insertion process is traditionally viewed as a thermodynamic event, but mutational analysis of Sec61 pore region suggest that the translocon not only provides a partitioning site, but also plays an active role in TMS insertion[63]. The mutations of the pore ring residues of Sec61 either enhanced or decreased membrane protein insertion[18]. In addition, several mutations in Sec61 were reported to affect the topogenesis of membrane proteins[19].

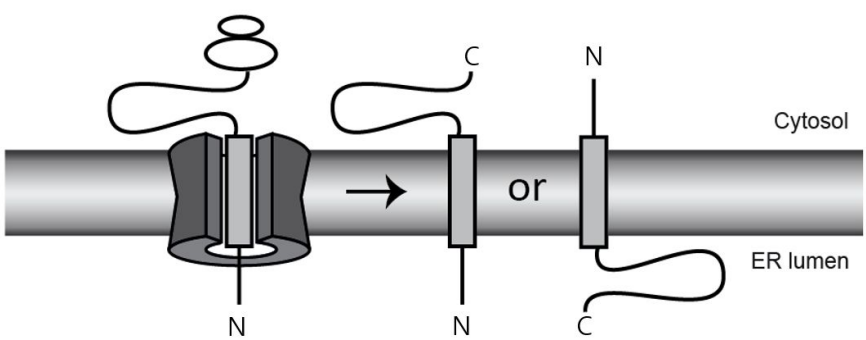


Figure 4. Sec61 mediates membrane protein insertion. Sec61 opens laterally via TMS 2 and 7 to allow partitioning of a TMS into the membrane. Single-spanning membrane proteins can be inserted into the membrane in two orientations.

I.3. Protein targeting and insertion into the mitochondrial IM

I.3.1. Protein import pathways for mitochondrial proteins

Majority of mitochondrial proteins are synthesized in the cytosol and are targeted to the mitochondria[73]. The protein targeting to mitochondria was traditionally viewed as post-translational event, but there is an accumulating evidence that co-translational targeting occurs for mitochondrial proteins[74]. These evidences include the identification of OM14 as a ribosome receptor of the OM and ribosomes actively translating mitochondrial proteins at the mitochondrial OM by site-specific ribosome profiling[75,76].

In either cases, most mitochondrial proteins initially go through the translocase of the OM (TOM) complex. Here, the proteins are sorted to four different compartments, OM, intermembrane space (IMS), IM and matrix[77]. OM proteins are inserted by TOM/SAM (sorting and assembly machinery) complex[78]. IMS proteins takes either MIA or TIM23 (mitochondrial IMS assembly or translocase of the IM) pathway. Proteins taking MIA pathway carries a cysteine motif[79]. With help from small Tim proteins, the disulfide bond formation by Mia40 drives the translocation of IMS proteins and prevents the leakage of proteins back to the cytosol[80]. In another case, an IMS protein is targeted to TIM23 complex, processed twice by MPP (matrix processing peptidase) and IMP (IM peptidase), and the mature protein is released into the IMS[81]. Carrier proteins with internal sorting signals take TIM22 pathway for their insertion into the membrane. TIM23 complex mediates the translocation or insertion of presequence carrying proteins at the mitochondrial IM[82].

I.3.2. TIM23 complex

TIM23 complex is composed of Tim23, Tim17, Tim50, Tim21, Tim44, Mgr2 and associated molecular motors such as mtHsp70[83]. Tim23 is the pore forming subunit of the complex[84-86]. Whether it forms a pore by itself or in association with Tim17 is unknown[87,88]. Both Tim23 and 17 have 4 TMS with their N- and C-terminus in the IMS[89,90]. Tim50 plays a role in presequence recognition[91-94]. Tim21 along with Mgr2 tethers TIM23 to the respiratory chain complex[95]. The molecular motors, mtHsp70, PAM16 and 18 are recruited to the translocase via Tim44 and translocate proteins to the matrix[96-98]. Membrane proteins carrying a presequence are released into the membrane by two sorting mechanisms. Some IM proteins are inserted directly by TIM23 (Stop-transfer pathway) whereas the others are fully translocated to the matrix first and re-inserted into IM from the matrix by other proteins such as Oxa1[83] (Figure 5). Recent studies showed the certain proteins, such as Mdl1, utilizes both sorting pathways[99,100]. Nevertheless, the mechanism of channel opening towards the lipid bilayer and the domain of Tim23 or other subunits of TIM23 responsible for the lateral insertion of proteins remains largely unknown. In addition, how TIM23 distinguishes a stop-transfer TMS and a conservative sorting TMS is also elusive.

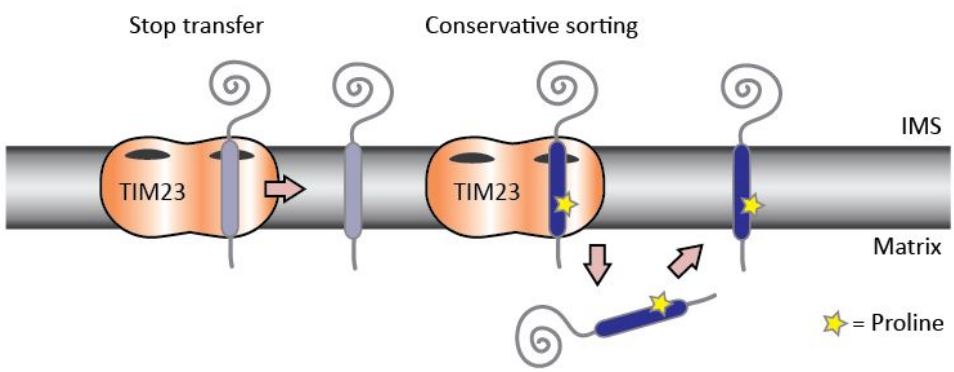


Figure 5. TIM23 dependent protein sorting pathways. Mitochondrial IM proteins of stop-transfer pathway are released into the membrane at the level of TIM23 whereas those taking the conservative sorting pathway are translocated to the matrix and inserted into the membrane afterwards. The TMS of conservative sorting pathway proteins tend to have a proline residue.

I.3.3. Determinants of mitochondrial IM protein insertion

The biological scale is reported for the TIM23 complex mediated membrane protein insertion into the mitochondrial IM[101]. There is a strong correlation with the ER biological scale (R^2=0.81), with leucine being the most hydrophobic and arginine being the most hydrophilic[101]. However, the hydrophobicity required for a TMS retention in the mitochondrial IM is influenced by the presence of the mAAA protease as it actively dislocates a TMS from the mitochondrial IM[102]. In the absence of the mAAA protease, the TIM23 complex is capable of inserting a TMS composed of 19 alanines with no leucine that is normally dislocated from the membrane by the mAAA protease[102]. As the TIM23 complex is capable of inserting a TMS composed of 19 alanine but bad at inserting TMS with a proline, one hypothesis is that TIM23 recognizes α-helical structure rather than the absolute hydrophobicity[103]. However, the protein components responsible for the difference in TMS recognition or the component that defines the hydrophobicity required for membrane protein insertion into the mitochondrial IM at the level of TIM23 remained unknown. Recently, Mgr2 was reported to delay the insertion of a model mitochondrial IM protein, Cytochrome b2-DHFR (Cyb2-DHFR) and Mgm1[104] (Figure 6). Moreover, in the absence of Mgr2, a protein that lacks the sorting signal composed of two positively charged residues, was inserted into the membrane[104]. Thus, Mgr2 was proposed as a lateral gate keeper of the TIM23 complex[104]. But whether Mgr2 acts generally on membrane protein insertion into the mitochondrial IM and sets the hydrophobicity requirement for insertion is questionable.

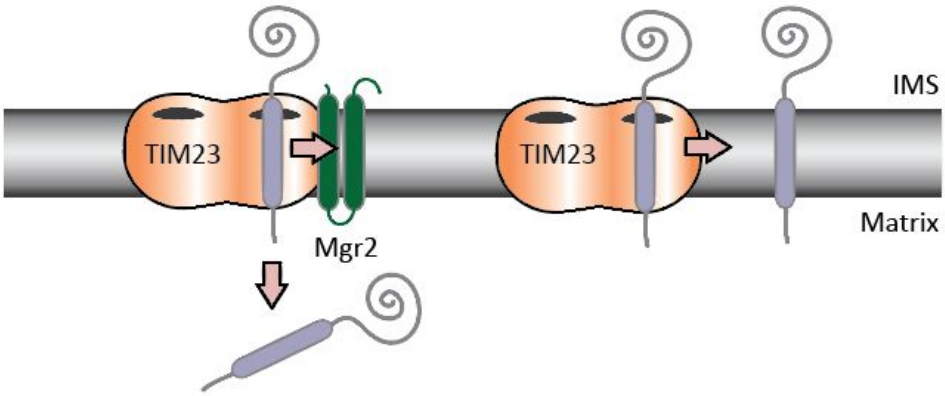


Figure 6. The schematic diagram of mitochondrial IM protein insertion in the presence or absence of Mgr2. If Mgr2 acts as a lateral gate block, the presence of Mgr2 would favor the translocation of proteins whereas the absence of Mgr2 would favor the insertion of proteins into the membrane.

I.4. Aims and experimental approaches

I.4.1. Development of a glycosylatable GFP as a topology reporter

The available topology reporters for eukaryotic cells required biochemical treatments to cells such as fixation, protease treatment or lysis[14]. Thus, there was a need for a live-cell topology reporter. The goal of the thesis was to develop a topology reporter that allows rapid live-cell assessment of membrane protein topology in eukaryotes. To develop such topology reporter, GFP was chosen as a candidate. GFP has been widely used as a fluorescent probe that allows live-cell imaging of protein localization[105]. Moreover, GFP has been proved as a topology reporter in bacterial system where GFP's fluorescence is lost in the periplasm due to the oxidizing environment that disrupts the proper folding of GFP[25]. Thus, the challenge in this study was to design GFP that is fluorescent in the cytosol but becomes non-fluorescent in the ER lumen (Figure 7). For this, the ER lumen specific protein modification, N-linked glycosylation, was utilized. N-linked glycosylation adds 2kDa glycan to a protein which can disrupt the proper folding of GFP. N-linked glycosylation sites were engineered at selected sites near the fluorophore of GFP. The engineered GFPs were expressed in the cytosol or in the ER lumen and the glycosylation and fluorescence profile were assessed[106].

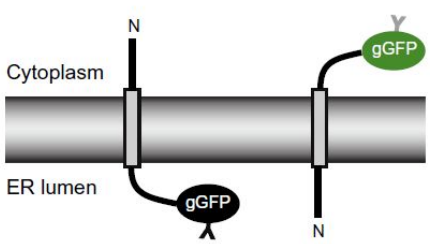


Figure 7. The schematic diagram of glycosylatable GFP (gGFP). gGFP is unglycosylated and fluorescent in the cytosol but is glycosylated and non-fluorescent in the ER lumen[14].

I.4.2. Membrane protein insertion by Sec61 translocon of yeast and human

To monitor the membrane protein insertion into the membrane via Sec61 translocon, model proteins of different characteristics were expressed in both yeast and HEK293T.

The model proteins were based on leader peptidase (Lep) from *Escherichia coli*[69]. As a model signal anchor protein, the two hydrophobic segments of Lep were replaced by 19 amino acids composed of alanines and leucines (Hydrophobic segment: H-segment). The H-segment functions as a signal anchor sequence (SAS)[17]. The hydrophobicity of the SAS is controlled by the number of leucines. SAS is flanked by N-glycosylation sites, one at the N-terminus and two at the C-terminus. These sites are only utilized upon exposure to the ER lumen, therefore serving as an indicator of membrane protein topology. The addition of glycan at each N-glycosylation sites adds about 2kDa to the protein mass, thereby slowing protein migration when analyzed by SDS-PAGE. Non-glycosylated product refers to non-targeted proteins, singly glycosylated product refers to proteins embedded in the membrane with its N-terminus facing the ER lumen, doubly glycosylated product refers to proteins embedded in the membrane with its N-terminus in cytosol and triply glycosylated product refers to fully translocated proteins[107] (Figure 8).

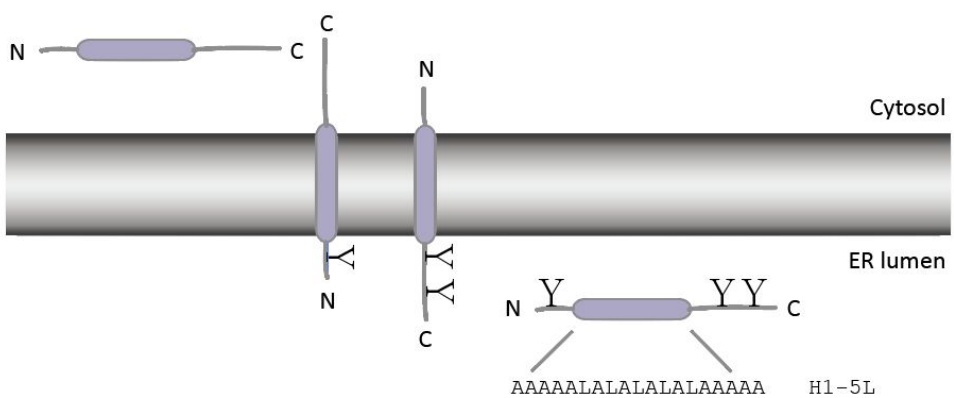


Figure 8. The schematic representation of a single-spanning model protein, LepH1. LepH1 carries a single TMS composed of *n* leucine and 19-*n* alanine that is flanked by glycosylation sites one at the N-terminus and two at the C-terminus. Y denotes

glycosylation.

To study multi-spanning membrane protein biogenesis, the H-segment was placed as the last TMS in 2nd, 3rd and 4th TM positions[69,70,108]. Here, only the insertion of the last TMS was tested as the protein targeting and initial anchoring of the protein to the ER membrane occurs by the first TMS of the protein[109]. The model proteins carry N-linked glycosylation sites to distinguish the membrane inserted form and non-inserted form. Where indicated, an additional TMS was placed after the H-segment to study the insertion of a TMS in the middle of multi-spanning membrane proteins.

Few factors governing the membrane protein insertion was tested using Lep model proteins. Hydrophobicity is the key factor in protein targeting to the ER membrane and subsequent topogenesis[110]. To investigate the targeting and topogenesis of single-spanning membrane protein of various hydrophobicity, LepH1 with H-segment composed of 1 to 10 leucine was expressed in yeast and mammalian cells. In yeast, a marginally hydrophobic segment is targeted by post-translational pathway whereas a hydrophobic segment is targeted via co-translational pathway[17]. Thus, the expression of LepH1 in yeast revealed the link between the targeting pathway and final topology of the proteins.

LepH1 was expressed in mammalian cells as well, but it should be noted that LepH1 presumably takes co-translational pathway in mammalian cells. Only the small secretory proteins of less than 120 amino acids are reported to take post-translational translocation pathway in mammalian cells[51]. Thus, the C-terminus of LepH1 was truncated to make the overall protein length 120 amino acids to study the topogenesis of post-translationally targeted single-spanning membrane proteins carrying a TMS of different hydrophobicity in mammalian cells.

Another factor that influences the membrane protein topogenesis is the C-terminal length. A long C-terminal tail that follows a SAS favors the $N_{in}C_{out}$ ($N_{Cytosol}C_{ER\text{-}lumen}$) orientation of a single-spanning membrane protein[111]. To further test the effect of varying C-terminal length of a single-spanning protein in membrane protein topogenesis, LepH1 was elongated by the sequence of yeast glycosylatable GFP (gGFP). gGFP from yeast is non-fluorescent in mammalian cells, therefore it is

regarded as an extension of 200 amino acids[112].

The role of the translocon complex in regulating membrane protein entry to the ER was also investigated in yeast[107]. SEC61 complex was proposed to insert membrane proteins via the lateral gating helices of Sec61, but how the gating is regulated remained largely questionable. Here, the residues lining yeast Sec61 were systematically replaced with an alanine residue one at a time and the effect of the mutation in membrane protein insertion and topogenesis were monitored with LepH1 and LepH2 carrying a marginally hydrophobic TMS.

I.4.3. Assessment of Mgr2 in membrane protein insertion via TIM23 complex

To study whether Mgr2 sets the threshold hydrophobicity required for membrane protein insertion into the mitochondrial IM, Mgr2 deletion (*mgr2*Δ), Mgr2 overexpression (Mgr2↑) and the corresponding wild-type (WT) (YPH499) strains were acquired from Dr. Pfanner[95,104]. To monitor membrane protein insertion and translocation in three strains above, a set of Mgm1 based model substrates were used. Mgm1 exists in two forms, *l*-Mgm1 and *s*-Mgm1[113]. *l*-Mgm1 is generated when Mgm1's 1st TMS is inserted into the membrane. *s*-Mgm1 is generated when the 1st TMS is translocated across the membrane, and the 2nd TMS is inserted into the membrane. The 2nd TMS carries a rhomboid cleavage site that is processed by Pcp1. The cleavage results in the release of *s*-Mgm1 into the IMS. In WT cells, roughly about the same amount of *l*-Mgm1 and *s*-Mgm1 exists (Figure 9).

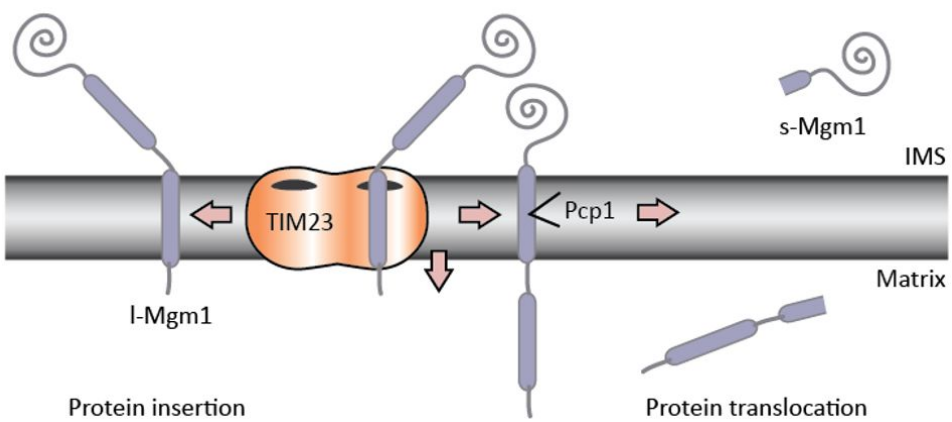


Figure 9. The schematics of Mgm1 processing. It contains an N-terminal targeting sequence, 2 TMS, and a soluble C-terminal domain. The 1st TMS of Mgm1 is sorted into the IM at the level of TIM23 complex to generate *l*-Mgm1 about 50% of the time, and translocated across the membrane to generate *s*-Mgm1.

Firstly, the effect of Mgr2 deletion or overexpression on the hydrophobicity required for membrane protein insertion was monitored using a set of Mgm1 whose 1st TMS is replaced with 19 amino acid stretch composed of *n* leucines and 19-*n* alanines (Mgm1-A/L)[101]. The number of leucine controls the hydrophobicity of the 1st TMS of Mgm1 where increasing the number of leucine makes the 1st TMS more hydrophobic. *l*-Mgm1 and *s*-Mgm1 represents membrane inserted and translocated Mgm1 respectively, thus the ratio of two isoforms (*l*-Mgm1 and *s*-Mgm1) was an indicator of membrane protein insertion efficiency. The ratio of *l*-Mgm1 and *s*-Mgm1 was compared between WT, *mgr2*Δ and Mgr2↑.

Next, the same screening was done using Mgm1 fusion proteins (MFPs) as Mgm1-A/L carries a non-native TMS that is dislocated from the mitochondrial IM by the mAAA protease[114]. The screening with mAAA independent MFPs was required to rule out the effect of the mAAA protease. MFPs contain the N-terminal domain of mitochondrial IM proteins of the stop-transfer pathway or the conservative sorting pathway fused to the C-terminal domain of Mgm1.

The 1st TMS of MFPs cover a broad range of hydrophobicity (ΔG_{app}). MFPs based on the proteins taking the stop-transfer pathway tend to be more hydrophobic compared to the ones taking the conservative pathway. Like Mgm1-A/L, the ratio of *l*-MFP and *s*-Mgm1 was compared to monitor membrane insertion and translocation. Lastly, the role of Mgr2 in recognizing the charged residues flanking a TMS was tested with Mgm1 mutants. Mgr2 permits the entry of Cyb2 with defective sorting signal where two positively charged residues preceding the TMS were mutated[104]. Mgm1 also has many charged residues that precedes or follows the 1st TMS[115]. These residues were mutated to alanine or to opposite charged residues and the insertion of these charged residue mutants in Mgr2 deletion or overexpression strains were investigated.

I.4.4. Screening for Tim23 with enhanced proline tolerance

To identify the domain of Tim23 that mediates the recognition and insertion of a TMS, both random and systematic mutagenesis screening was designed. The selection scheme utilizes a chimeric model protein made of Cyb2 and cytochrome c heme lyase (CCHL) and *tim23*Δ*cyc3*Δ strain.

Cyb2 is an IMS protein. It is first targeted to matrix via its N-terminal presequence which is removed by the MPP upon its exposure, generating i-Cyb2. i-Cyb2 is inserted into the IM and cleaved once again by IMP, releasing the mature Cyb2 (m-Cyb2) into IMS[116]. When alanine 63 (A63) is mutated to proline (P), Cyb2 is no longer sorted into the IM, instead it is translocated to matrix (Figure 10). By taking advantage of differential targeting of Cyb2, mature CCHL was fused to WT preCyb2(167) and preCyb2(167) with A63P mutation. preCyb2(167)-CCHL is inserted into membrane, whereas preCyb2(A63P-167)-CCHL is translocated to the matrix. CCHL is encoded by Cyc3 gene in *Saccharomyces cerevisiae* and adds heme to apo-cytochrome c[117]. CCHL is required at the IMS for cell survival on non-fermentable carbon source[118]. Thus, the expression of preCyb2(167)-CCHL in *cyc3*Δ rescues the growth defect on non-fermentable carbon source whereas preCyb2 (A63P-167)-CCHL does not.

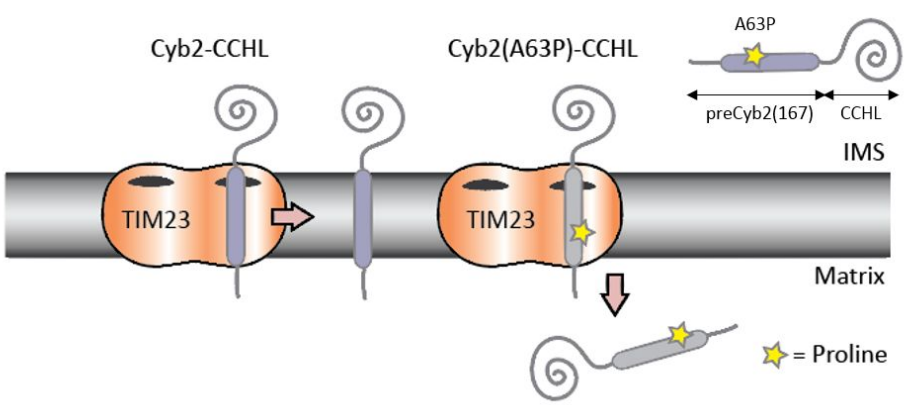


Figure 10. The sorting of preCyb2-CCHL and preCyb2(A63P)-CCHL. Cyb2 is sorted into the mitochondrial IM by TIM23. However, alanine to proline mutation at position 63 causes the translocation of the protein into the matrix.

*tim23*Δ*cyc3*Δ strain with WT TIM23 in Ura plasmid and preCyb2(A63P-167)-CCHL is viable only on fermentable carbon source. But if preCyb2 (A63P-167)-CCHL could be inserted into the IM and processed by IMP, it would support the cell growth on non-fermentable carbon source. As membrane insertion of Cyb2 is mediated by TIM23 complex, TIM23 was randomly mutagenized and introduced into *tim23*Δ*cyc3*Δ expressing preCyb2(A63P-167)-CCHL. A plasmid encoding WT TIM23 was removed by plasmid shuffling. Any colony formed after the selection on non-fermentable medium is a probable candidate exhibiting enhanced membrane insertion by Tim23 (Figure 11). Tim23 mutant capable of supporting the cell growth of *tim23*Δ*cyc3*Δ strain expressing preCyb2(A63P-167)-CCHL will be isolated and sequenced. Once the mutations are identified, it will be retransformed into *tim23*Δ*cyc3*Δ strain with a set of model proteins to validate enhanced membrane protein insertion activity by Tim23.

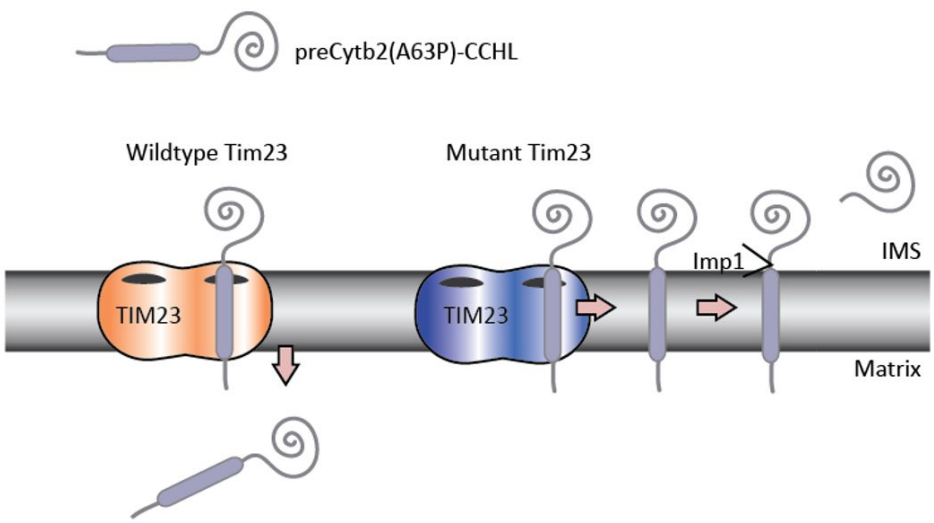


Figure 11. The selection scheme for Tim23 with enhanced membrane insertion activity. CCHL, encoded by CYC3 gene and required for cell growth in non-fermentable carbon source, was fused to the C-terminus of preCyb2(A63P). WT Tim23 translocates the fusion protein preCyb2(A63P)-CCHL to the matrix, therefore is expected to have no influence on the growth of *cyc3*Δ on non-fermentable carbon source. However, if a mutant Tim23 could insert the fusion protein into the IM, then CCHL would be exposed to the IMS and thus supports the cell growth on non-fermentable carbon source.

I.1.5. Quantitative analysis of protein import kinetics and forces

Quantitative information such as protein import kinetics or forces acting on preproteins during the import process is lacking for mitochondrial proteins. To quantitatively analyze the mitochondrial protein import process, an optical tweezer analysis was designed where an isolated mitochondria and a purified preprotein would be attached to different beads for kinetic and force measurements during the import process. Optical tweezer enables the measurement of protein import kinetics by monitoring the displacement of the bead that the substrate is attached to during the protein import at a very fine scale[119]. Protein import force is also measured by applying a pulling force in the opposite direction of the import, and the minimum force at which the import is blocked is representative of the protein import force[120].

For this, mitochondria expressing Mcr1-Avi tag will be isolated from yeast grown in non-fermentable carbon source. Mcr1 is an OM protein with its C-terminus exposed to the cytosol[121]. Biotin acceptor site, Avi tag, will be fused to the C-terminus of the Mcr1 to allow efficient biotinylation of mitochondria in the presence of BirA (Biotin ligase)[122]. The mitochondria will be biotinylated prior to isolation (*in vivo* biotinylation) or after isolation (*in vitro* biotinylation) and attached to the streptavidin coated beads[123].

Mitochondrial expressing Mcr1-Halotag will also be prepared as an alternative approach to prepare for cases of inefficient *in vivo*/*in vitro* biotinylation. Halotag interacts specifically with a set of Halo-ligands[124]. For the attachment of mitochondria expressing Mcr1-Halotag to the bead, streptavidin beads coated with biotinylated Halo-ligand will be used. The interaction between the Haloligand and Halotag will firmly hold mitochondria onto the bead.

The substrate preprotein, Cox5a will be expressed in *Escherichia coli* and purified. Purified Cox5a carries ybbR tag at the C-terminus as a CoA acceptor site. CoA-DNA will be attached to ybbR tag by Sfp[125]. The 3kb long linker DNA with 5' digoxigenin will be ligated to CoA-DNA, and the ligated product DNA linker-Cox5a will be fused to anti-digoxigenin coated polysterene beads for analysis[120].

The two beads, mitochondria attached bead and Cox5a attached bead, will be trapped in a laser beam and will be placed in close proximity to allow protein import. The import kinetics will be monitored by the displacement of the bead that Cox5a is attached to. The force loaded (the force that acts in the opposite direction of the import) will be changed to measure the protein import force.

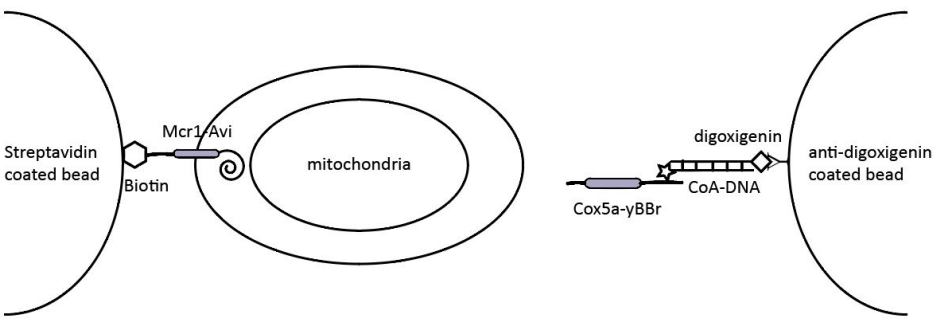


Figure 12. The experimental setup for optical tweezer analysis. Isolated mitochondria is biotinylated by BirA either *in vivo* or *in vitro* using Mcr1 with Avi tag as a biotin acceptor protein present in the OM. The mitochondria is attached to the streptavidin coated bead via biotin-streptavidin interaction. Cox5a-ybbR tag is expressed in *Escherichia coli* in the presence of CoA-DNA conjugate. Sfp adds CoA-DNA to ybbR tag and Cox5a with CoA-DNA will be purified. 3kb long DNA linker is amplified by PCR with 5' digoxigenin modified primer and ligated to Cox5a with CoA-DNA. The ligated product is attached to polysterene beads coated with anti-digoxigenin. Optical tweezer experiment is performed with these two beads.

CHAPTER II – MATERIALS AND METHODS

II.1. Yeast and mammalian cell culture

II.1.1. Glycosylatable GFP study

gGFP constructs in p424GPD vectors were transformed into W303-1α (*MATα*, *ade2*, *can1*, *his3*, *leu2*, *trp1*, *ura3*) and grown in -Trp medium at 30° C overnight[106]. Mammalian cells (HeLa or HEK-293T) were grown in medium (10% FBS in DMEM with antibiotics) at 37° C with 5% CO_2. Cells were transiently transfected with plasmids using Attractene (Qiagen) following the manufacturer's protocol[112].

II.1.2. Sec translocon study

*sec61*Δ strain (*MAT*α, *ura3-1*, *leu2-3*,*112*, *his3-11*,*15*, *trp1-1*, *ade2-1*, *can1-100*, *sec61*::*HIS3* [pH6-Sec61-YCplac33]) was transformed with pDQ1 encoding the WT SEC61 gene or mutant *sec61*[126,127]. pH6-Sec61-YCplac33 that contains WT SEC61 gene was removed by FoA (5-fluoroorotic acid) selection. The plasmids encoding model proteins were transformed into the FoA selected strains and the cells were cultured in -Leu-Trp medium at 30° C overnight unless stated[107]. Mammalian cells were handled as describes in M.1.1. HiperFect (Qiagen) was used for the transfection of siRNAs.

II.1.3. Mgr2 study

Mgm1 and MFP constructs encoded in pHP84HA vector were transformed into YPH499 (*MATa*, *ade2, lys2, his3, leu2, trp1, ura3*), *mgr2*Δ (YPH499 *mgr2::KANMX6*) and Mgr2↑ (*mgr2*Δ [pPGKMGR2]) for Mgr2 characterization[104]. The cells were cultured in -Leu or -Leu-Ura medium at 30° C or 39° C overnight. For respiratory growth, yeast cells were cultured in appropriate medium with 3% glycerol as a carbon source.

II.1.4. Tim23 random mutagenesis study

pHP84 vectors encoding CYC3, CYC3-HA, Cyb2-CCHL-HA or Cyb2(A63P)-CCHL-HA were transformed into BWY4741*(MATa, his3, leu2, met15, ura3), cyc3*Δ (BWY4741, *cyc3::KANMX6*), W303-1α and *cyc3*Δ*tim23*Δ (W303-1α, *cyc3::KANMX6, tim23::HIS3* [pRS316-TIM23]). The transformants were selected on -Leu at 30° C and were streaked on YPEG plate for the growth assay 30° C.

II.2. Proteasome inhibition assay

Six hours after transfection, the proteasome inhibitor, MG132 was added to the cells together with fresh media at a concentration of 10μM in dimethyl sulfoxide (DMSO). Equal amount of DMSO was added to control populations[112].

II.3. Protein preparation, SDS-PAGE and Western blotting

Yeast transformants were grown overnight at 30° C. Whole-cell lysates were prepared as previously described[128]. Lysates were incubated with or without endoglycosidase H (Endo H) (Roche) for 2 h at 37° C to remove N-linked oligosaccharides. SDS–PAGE and Western blot analysis with an anti-HA antibody were followed. Western blots were developed with Amersham Bioscience Advanced ECL kit on a Biorad Chemi-doc-XRS+ system (Biorad)[106].

For mammalian cells, lysates of were prepared using lysis buffer (1% NP-40 in 1X PBS with Protease Inhibitors). Endo H treatment (Roche), SDS-PAGE and Western blot analysis were performed as described previously. Horseradish peroxidase (HRP) conjugated GFP antibody (Rockland) was used to detect GFP and GFP fusion proteins (1:5000)[112].

II.4. Pulse (Chase) labeling, immuno-precipitation and autoradiography

Pulse(Chase) labeling of yeast transformants, immuno-precipitation and autoradiography were done as described in Jung *et al*[16]. Mammalian cells transfected with plasmids encoding model proteins were cultured to 90% confluency in 60mm dish. The cells were washed twice with 1X PBS and incubated in -Met-Cys medium for 45 min. 10µl of S^{35} Met-Cys was added and the cells were further incubated for 20 min. The cells were collected in 1.5ml tube with lysis buffer (1% NP-40 in 1X PBS with Protease Inhibitors) and subjected to immuno-precipitation.

II.5. Optical tweezer based *in vitro* protein import assay

II.5.1. Mitochondria isolation and *in vitro* protein import assay

The isolation of mitochondria will be performed as decribed in Weckbecker *et al*[129]. The isolated mitochondria will be biotinylated *in vivo* or *in vitro* following van Werven *et al* or manufacturer's protocol (Avidity) respectively[123]. The preparation of the substrate protein, Cox5a, and the optical tweezer analysis will be done as described in Maillard *et al*[120]. In brief, yeast cells will be cultured in semi-lactate medium to enrich mitochondria prior to isolation. For *in vivo* biotinylation, yeast transformants co-expressing Mcr1-Avi and BirA will be used and biotin will be added to the culture medium. For *in vitro* biotinylation, isolated mitochondria expressing Mcr1-Avi will be mixed with purified BirA in the presence of biotin. The biotinylated status of mitochondria will be checked by streptavidin shift assay as well as western blotting against biotin[123].

II.5.2. List of constructs made for *in vitro* protein import assay

Up to date, the materials required for an optical tweezer based *in vitro* protein import assay were prepared. To produce biotinylated mitochondria, mitochondria was isolated from W303-1α expressing Mcr1-HA-Avi. Isolated mitochondria was incubated with purified BirA in the presence of biotin for biotinylation of Avi tag. The substrate mitochondrial proteins, Cox5a and preCyb2-GFP, were expressed in *Escherichia coli* and isolated by affinity purification (Data not shown).

II.6. Fluorescence microscopy

Yeast transformants expressing gGFP fusion constructs were grown overnight in 5ml of -Trp medium at 30° C. 100µl of cells were taken from 0.7 OD_{600} culture. Cells were transferred to a 96well plate for fluorescence assessment using a Zeiss Axiovert 200M inverted microscope with a Plan-NeoFluar 100/1.30 NA oil-immersion objective lens. Fluorescence images were taken as described in Sung et al., using a standard fluoresceinisothiocyanate (FITC) filter set (excitation band pass filter, 450-490nm; beam splitter, 510nm; emission band pass filter, 515-565nm). The pictures were taken with an exposure time of 0.2 ms[106]. For HEK-293T or HeLa cells, transiently transfected cells were assessed under JuLi fluorescence cell imager (Digital Bio) for fluorescence measurement. Where indicated, cells were cultured on a cover glass for direct imaging on Axioimager A1[112].

II.7. Fluorescence measurements

Yeast transformants expressing gGFP fusion constructs were grown in 10 ml of -Trp medium at 30℃ overnight. Cells were harvested at 0.3-0.6 OD_{600} by centrifugation at 3,000g and resuspended in 200µl of YSB buffer (50mM Tris–HCl, pH 7.6, 5mM EDTA, pH 8, 10% (w/v) Glycerol and 1X Protease Inhibitor Cocktail. Cells were transferred to a 96-well microplate (Nunc) and the fluorescence measurements were taken on Perkin Elmer Envision 2102 Multilabel reader with excitation band pass filter at 460nm and cut-off FITC filter at 535nm. All fluorescence measurements were subtracted by fluorescence from whole-cell lysates of yeast transformants carrying an empty vector. Averages of at least three independent measurements were plotted with standard errors[106]. For HEK-293T or HeLa cells, transiently transfected cells were washed with 1X PBS and collected by trypsin-EDTA treatment. The collected cells were resuspended in 3.7% formaldyhyde in 1X PBS and analyzed with FACS Canto (BD)[112].

CHAPTER III – RESULTS

III.1. gGFP as a novel membrane protein topology reporter.

III.1.1. Engineering of gGFP for *Saccharomyces cerevisiae*

The rationale behind the design of gGFP was the assumption that the presence of an N-linked glycan near the GFP fluorophore might interfere with protein folding and maturation of the fluorophore, rendering the protein non-fluorescent. Based on the X-ray structure of GFP, two asparagine residues at positions 121 and 170 (N121 and N170) near the fluorophore were initially chosen[130] (Figure 13A). To introduce an N-linked glycosylation acceptor site, asparagine-X-threonine-Y (where X, Y can be any amino acid except proline), isoleucine 123 (I123) and glutamic acid 172 (E172), two residues downstream of the asparagine residues at positions 121 and 170, were individually changed to threonine in yeast enhanced GFP (yEGFP)[131]. To facilitate detection by Western blotting and immuno-fluorescence, a hemagglutinin (HA) tag was introduced directly upstream of GFP (Figure 13B). GFP fluorescence was abolished when isoleucine123 was changed to threonine.

To obtain a version of gGFP that is localized to the lumen of the ER, the signal peptide (SP) from invertase was fused to the N terminus of GFP (Figure 13B). Three constructs each carrying GFP, GFP(E172T) or SP-GFP(E172T) were transformed into the *Saccharomyces cerevisiae* strain W303-1a. Whole-cell lysates were prepared and fluorescence was measured. While fluorescence of GFP(E172T) was comparable to that of WT GFP, fluorescence of SP-GFP(E172T) was not significantly different from the whole-cell lysate prepared from a yeast transformant carrying an empty vector (Figure 13C). The version of GFP used in the study, yEGFP, in itself has much lower fluorescence when localized to the lumen of the ER[131]. When comparing the fluorescence of SP-GFP and SP-GFP(E172T), approximately a twofold reduction in fluorescence for SP-GFP(E172T) was observed (Figure 13C inset). Hence, the addition of the glycan moiety reduces the residual fluorescence of lumenally located SP-GFP(E172T) to background levels.

To further determine whether SP-GFP(E172T) was translocated to the lumen of the ER, whole-cell lysates were treated with Endo H for removal of N-glycans, followed by SDS-PAGE and Western blotting. The sample treated with Endo H showed faster migration on the Western blot, indicating that GFP(E172T) was glycosylated, and thus correctly translocated to the lumen of the ER (Figure 13D). Next, fluorescence from the cells expressing either GFP(E172T) or SP-GFP(E172T) was assessed by fluorescence microscopy. Only the cells expressing GFP(E172T), the cytosolically localized construct, exhibited a detectable fluorescence signal (Figure 13E). These results demonstrate that GFP(E172T) is unglycosylated and fluorescent in the cytosol but is glycosylated and non-fluorescent when localized in the ER lumen. This version of GFP was named as gGFP[106].

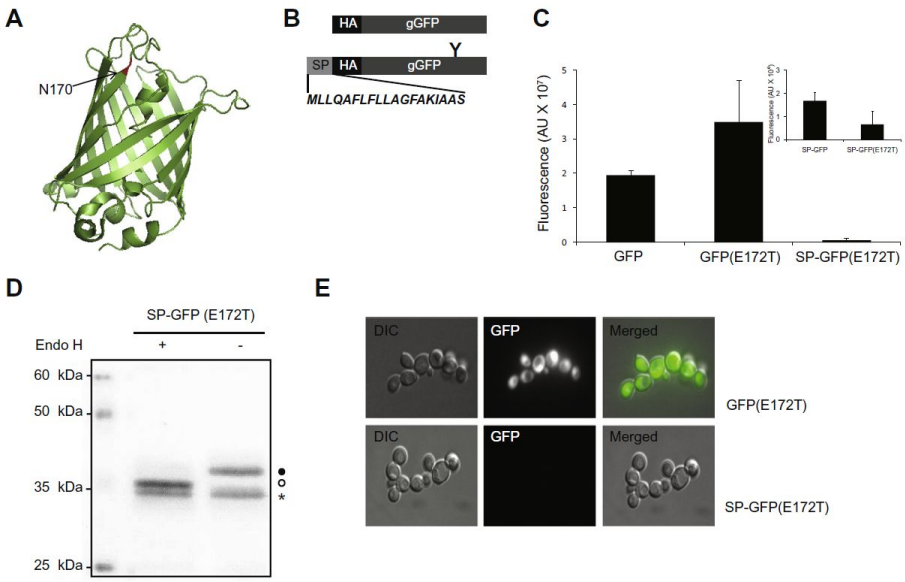


Figure 13. Engineering a gGFP. (A) Structure of yEGFP (PDB 3AI4)[132]. The red label indicates the N170 residue that becomes glycosylated in gGFP. (B) Schematic representation of GFP constructs. The glycosylation site is labeled as Y. HA denotes a triple hemagglutinin tag. The sequence of the invertase SP is shown. (C) Fluorescence measurements. Whole-cell lysates were prepared from yeast transformants expressing GFP, GFP(E172T), SP-GFP or SP-GFP(E172T). Fluorescence was measured with a 460 nm excitation and 535 nm cut-off filter, and relative fluorescence units were normalized by subtracting the values from the whole-cell lysate of a yeast transformant carrying an empty vector. Averages of at least three independent measurements are shown with standard errors. In the inset, GFP fluorescence of SP-GFP and SP-GFP(E172T) is compared. (D) SP-GFP(E172T) is glycosylated. Whole-cell lysates prepared from yeast transformants expressing GFP(E172T) or SP-GFP(E172T) were subjected to Endo H digestion, SDS–PAGE, and Western blot analysis. ● denotes a glycosylated form, ○ denotes a non-glycosylated form. *indicates an uncharacterized background band. (E) Fluorescence microscopy of yeast cells expressing GFP(E172T) or SP-GFP(E172T). Cells were viewed with DIC and GFP filters, and the images were merged[106].

For large membrane proteins, the difference in molecular weight between the presence and absence of a single N-linked glycan (2 kDa) in gGFP might be too small to be detected on SDS-gels. To overcome this difficulty, an additional glycosylation site was engineered at positions 7 or 228 in gGFP, or in HA tag. The positions of these sites were chosen such that they would least disrupt folding and maturation of the fluorophore.

To test the efficiency of N-linked glycosylation at the three sites, whole-cell lysates of yeast transformants expressing GFP(L7N/E172T), GFP(E172T/G228N), GFP(HAg/E172T), SP-GFP(L7N/E172T), SP-GFP(E172T/G228N), or SP-GFP(HAg/E172T) were prepared and subjected to Endo H digestion. SP-GFP(E172T/G228N) and SP-GFP(HAg/E172T) were efficiently glycosylated on both sites (Figure 14A). In comparison, N7 in SP-GFP(L7N/E172T) did not get efficiently glycosylated as about equal amounts of singly and doubly glycosylated proteins were apparent. Whole-cell lysates from yeast transformants expressing these constructs were then tested for fluorescence. GFP(E172T/G228N) did not fluoresce, whereas GFP(L7N/E172T) and GFP(HAg/E172T) showed comparable levels of fluorescence as GFP(E172T) (Figure 14B). Since the additional glycosylation site was efficiently glycosylated and the fluorescence pattern of gGFP was maintained in GFP(HAg/E172T), this version of gGFP can be used to assay the topology of membrane proteins of relatively large size[106].

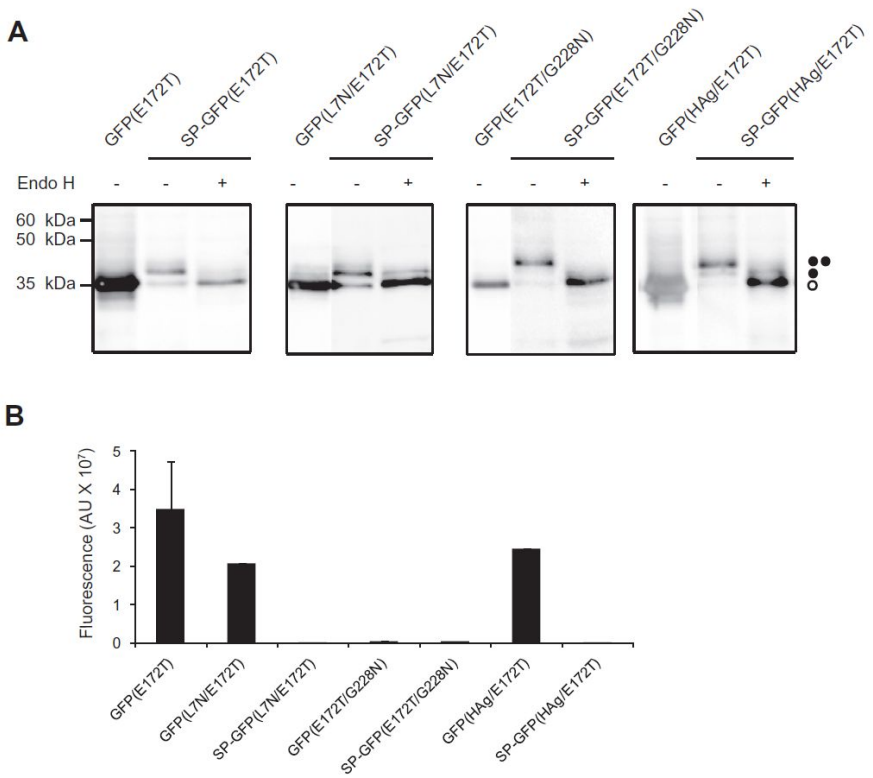


Figure 14. Engineering an additional glycosylation site into gGFP. (A) gGFP with an additional glycosylation site can be doubly glycosylated. Endo H digestion analysis of yeast transformants expressing SP-GFP(E172T), SP-GFP(L7N/E172T), SP-GFP(E172T/G228N) or SP-GFP(HAg/E172T). GFP(E172T), GFP(L7N/E172T), GFP(E172T/G228N) and GFP(HAg/E172T) were loaded as controls for an unglycosylated gGFP. ●● indicates a doubly glycosylated form, ● indicates a singly glycosylated form, ○ indicates a non-glycosylated form. (B) Fluorescence measurements. Fluorescence was measured as described in Figure 12C. Averages of two measurements are shown except for GFP(E172T). Average of three measurements is shown for GFP(E172T) as in Figure 12C[106].

III.1.2. Validation of gGFP as a topology reporter

To test whether gGFP can be used as a topology reporter, gGFP was fused to the C terminus of two membrane protein constructs based on Lep protein from *Escherichia coli*, an IM protein with two TMS (TMS1, TMS2) near the N terminus[133]. In these constructs, TMS2 was replaced by two different 19-residue long segments; one composed of polar residues that does not insert into the IM (construct Lep-C_{in}-gGFP) and the other composed of a very hydrophobic 6 L/13A segment that is known to insert efficiently into the ER membrane (Lep-C_{out}-gGFP)[69] (Figure 15A). Whole-cell lysates from yeast transformants expressing Lep-C_{in}-gGFP exhibited high fluorescence, whereas Lep-C_{out}-gGFP transformants showed only weak fluorescence (Figure 15B).

Next, to assess the glycosylation status of the two constructs, whole-cell lysates from yeast transformants expressing either Lep-C_{in}-gGFP or Lep-C_{out}-gGFP were prepared and treated with Endo H. Two glycan acceptor sites are present in the Lep part, one in the lumenally located N-terminal tail and the other just downstream of the P or H segment, hence molecules with a single glycan (1G) have an N_{out}-C_{in} membrane topology, whereas triply glycosylated molecules (3G) have an N_{out}-C_{out} orientation (Figure 15A). As expected, Lep-C_{in}-gGFP was singly glycosylated and Lep-C_{out}-gGFP was triply glycosylated (Figure 15C). A cleaved and glycosylated form of Lep-C_{out}-gGFP was also detected. Previously, it was shown that this form is generated by signal peptidase cleavage in the 6L/13A segment, releasing the C-terminal domain of the protein to the lumen. These results demonstrate that gGFP fusion neither interferes with correct protein targeting nor with the membrane topology, thus can be used for topology mapping[106].

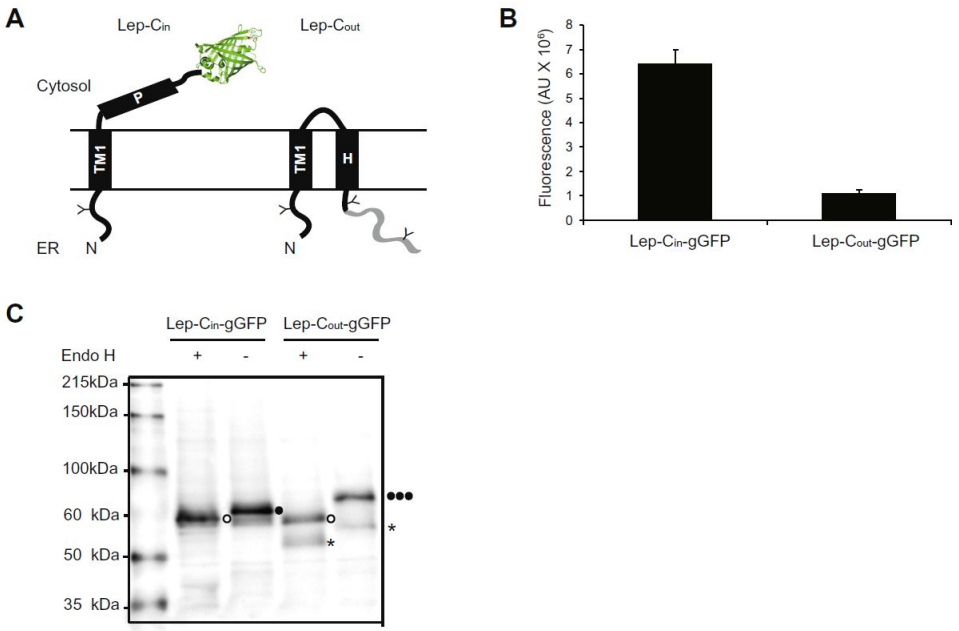


Figure 15. gGFP can be used as a membrane topology reporter. (A) Schematic representation of model proteins, Lep-C_{in} and Lep-C_{out}[69]. The glycosylated sites are labeled as Y. TM1 indicates TM segment 1 of *Escherichia coli* Lep, P stands for a polar segment, and H stands for a hydrophobic segment[133]. (B) Fluorescence measurements. Whole-cell lysates were prepared from yeast transformants expressing Lep-C_{in}-gGFP or Lep-C_{out}-gGFP. Fluorescence was measured as in Figure 12C. (C) Endo H digestion analysis of yeast transformants expressing Lep-C_{in}-gGFP or Lep-C_{out}-gGFP. ●●● indicates a triply glycosylated form, ● indicates a singly glycosylated form, ○ indicates a non-glycosylated form, and *indicates a cleaved product[106].

III.1.3. Development of gGFP for mammalian cell culture.

To test the applicability of yeast gGFP in mammalian cells, yeast gGFP was cloned into pcDNA3.1 and transfected into HeLa cells. However, yeast gGFP was not expressed in HeLa cells, which may have been due to the differences in codon usage between yeast and mammalian systems.

To overcome problems in gGFP expression in HeLa cells, pEGFP-N1 plasmid encoding mammalian EGFP was used to introduce a glycosylation site by engineering an E173T substitution (equivalent to E172T in yEGFP, which was reported to abolish fluorescence upon N-linked glycosylation)[106] (Figure 16A). Mammalian EGFP contains an additional valine residue at position 2 compared to yEGFP, thus the residue number is one higher for mammalian EGFP.

To check whether the engineered N-linked glycosylation site (N171- I172-T173) is utilized in the ER lumen, the cleavable SP of secreted yeast invertase was fused to EGFP variants at their N-terminus for translocation into the ER lumen. Previously, SP of yeast invertase was shown to be functional in mammalian cells[134]. In addition, the ER retention signal, KDEL sequence, was added to the C-terminus to prevent a gGFP fusion protein from being secreted[135] (Figure 16B). The plasmid encoding EGFP(WT) or EGFP(E173T) with or without the SP was transfected into HEK-293T cells and their fluorescence pattern was examined by fluorescence microscopy and FACS analysis. Both EGFP(WT) and EGFP(E173T) exhibited fluorescence in the cytosol (Figure 16C, D). The fluorescence from EGFP(E173T) was more intense compared to that of the WT EGFP. SP-EGFP(E173T) fluoresced in the ER lumen, indicating that the E173T mutation alone did not abolish fluorescence even though being glycosylated[112] (Figure 16E).

A study by Losfeld *et al* reported another version of gGFP that was developed for clinical use[136]. This version of GFP carries an engineered N-linked glycosylation site at position 145. Once this site is glycosylated, GFP loses fluorescence in mammalian cells. Therefore, we prepared EGFP(WT) with two N-linked glycosylation sites, N145 and N171, by introducing an N147T substitution in EGFP(E173T) (Figure 16A). With two N-linked glycosylation sites in the EGFP sequence, the size difference between glycosylated and unglycosylated EGFP would be more prominent, thus enhance the applicability as a potential membrane topology reporter when fused to larger membrane proteins.

The protein EGFP(N147T/E173T) was tested for fluorescence and glycosylation patterns. EGFP(N147T/E173T) exhibited fluorescence in the cytosol whereas its ER version, SP-EGFP(N147T/E173T), showed no fluorescence under a fluorescence microscopy (Figure 16C). FACS analysis offered more detailed fluorescence measurements of EGFP variants (Figure 16D). Mean fluorescence was reduced in SP-EGFP(N147T/E173T) compared to the cytosolic version but we noticed a significant difference in maximum fluorescence between the cytosolic and the ER version of EGFP(N147T/E173T), as well as between detectable and non-detectable fluorescent EGFP variants under fluorescence microscopy (Figure 16C, D). While mean fluorescence may differ depending on protein concentration, maximum fluorescence may more accurately indicate changes in fluorescence intensity due to N-linked glycosylation, thus more reliable to assess glycosylation-dependent fluorescence changes of gGFP. Endo H digestion of whole-cell lysates showed that SP-EGFP(N147T/E173T) was efficiently glycosylated in the ER lumen (Figure 16E). As EGFP(N147T/E173T) exhibits the characteristics of yeast gGFP whose fluorescence is selectively lost only upon glycosylation in the ER lumen, we refer to EGFP(N147T/E173T) as a mammalian gGFP[112].

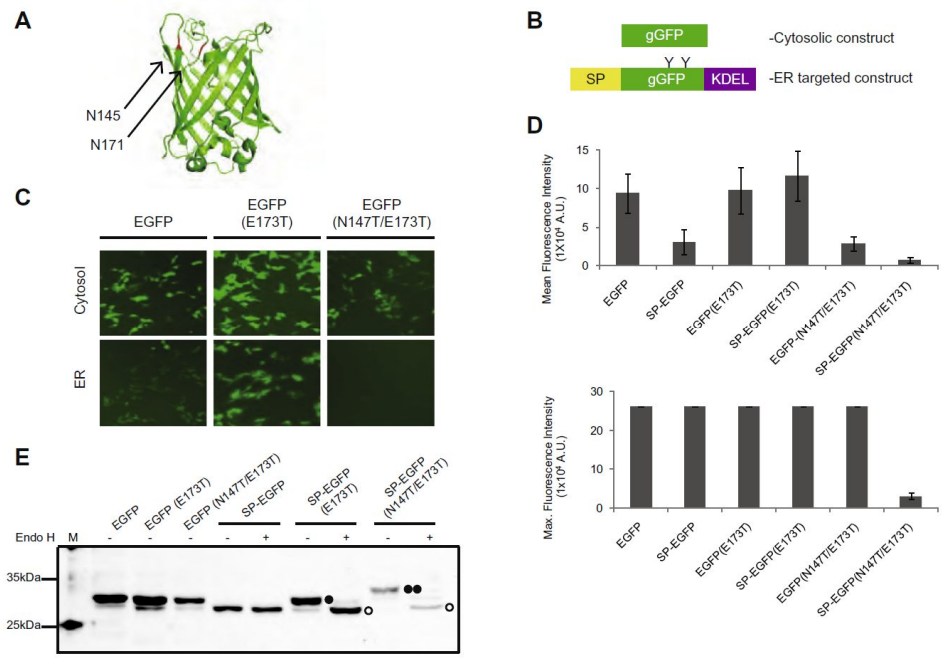


Figure 16. gGFP for mammalian cells. (A) The structure of EGFP (PDB 2Y0G) with engineered glycosylation sites, N145 and N171, marked in red[137]. (B) gGFP constructs. SP denotes a cleavable signal peptide (SP) of an invertase and KDEL is the ER retention sequence. The engineered glycosylation sites are labelled with Ys. (C) Fluorescence microscopy of HEK-293T cells expressing EGFP(WT), EGFP(E173T), EGFP(N147T/E173T) in the cytosol and SP-EGFP, SP-EGFP(E173T) or SP-EGFP(N147T/E173T) in the ER. Cells were viewed with a GFP filter for GFP fluorescence after 24 h of transfection under Juli cell imager. (D) The cells were analyzed by FACS Canto. Mean (top) and maximum (bottom) fluorescence of EGFP(WT), EGFP(E173T), EGFP(N147T/E173T), SP-EGFP, SP-EGFP(E173T) or SP-EGFP(N147T/E173T) are plotted with standard deviations. (E) SP-EGFP(N147T/E173T) is efficiently glycosylated. Whole-cell lysates prepared from HEK-293 cells expressing EGFP(WT), EGFP(E173T), EGFP(N147T/E173T), SP-EGFP, SP-EGFP(E173T) or SP-EGFP(N147T/E173T) were subjected to Endo H digestion, SDS–PAGE and Western blot analysis. ●● Denotes doubly glycosylated, ● singly glycosylated and ○ non-glycosylated form. M marks the standard molecular weight marker loaded for analysis[112].

To validate the applicability of mammalian gGFP as a membrane protein topology reporter, gGFP was fused to a set of model membrane proteins derived from *Escherichia coli* Lep of known membrane topology. The derived Lep variant (LepH3) contains 3 TMS, where the last TMS is the test segment of varying hydrophobicity made of leucines and alanines (H-segment)[70]. It has N-linked acceptor at two sites; in loop 2 and in the C-terminus (Figure 17A). LepH3 with 3 leucines and 16 alanines in the test segment (LepH3-C_{out}) is expected to have its C-terminus translocated to the ER lumen whereas LepH3 with 19 leucines (LepH3-C_{in}) is expected to have its test segment inserted into the membrane, thereby leaving its C-terminus in the cytosol.

The gGFP was fused to the C-terminus of LepH3-C_{out} and LepH3-C_{in}, then expressed in HEK-293T cells. When the fusion proteins were expressed in HEK-293T cells, LepH3-C_{in}-gGFP was fluorescent whereas LepH3-C_{out}-gGFP was not (Figure 17B, C). Endo H digestion of lysates confirmed that the C-terminus of LepH3-C_{in}-gGFP remained in the cytosol whereas that of LepH3-C_{out}-gGFP was translocated to the ER lumen (Figure 17D). In sum, these results demonstrate that mammalian gGFP is a convenient *in vivo* membrane topology reporter[112].

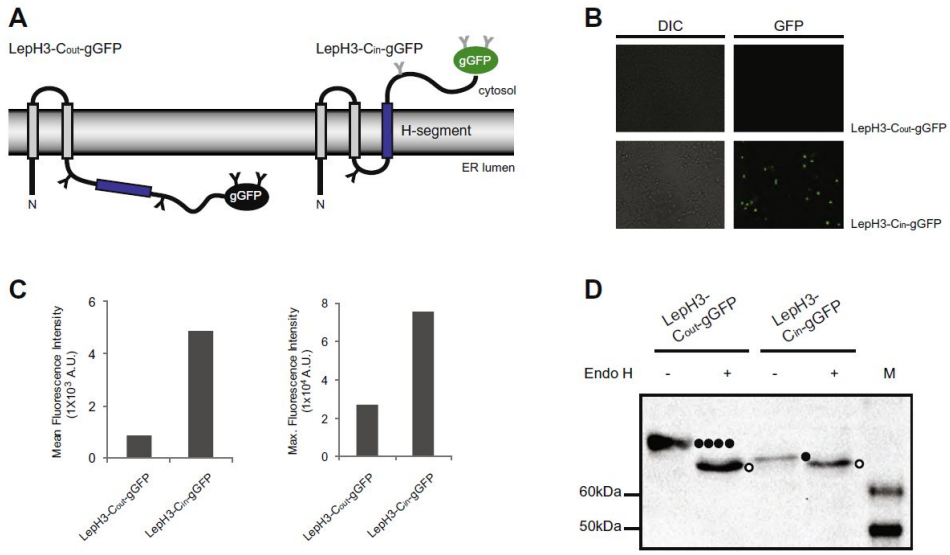


Figure 17. gGFP fused to model membrane proteins. (A) Schematics of LepH3-C_{out}-gGFP and LepH3-C_{in}-gGFP in the ER membrane. Glycosylation sites are marked with Ys. Utilized glycosylation sites are colored in black and unutilized sites are colored in grey. (B) Fluorescence microscopy of HEK-293T cells expressing Lep-H3-C_{out}-gGFP and Lep-H3-C_{in}-gGFP. Cells were viewed with a GFP filter for GFP fluorescence after 24 h of transfection under Juli cell imager. Brightness and contrast were adjusted for clearer pictures. (C) Mean (left) and maximum (right) fluorescence of LepH3-C_{out}-gGFP and LepH3-C_{in}-gGFP. The cells were analyzed by FACS Canto and the average of two independent measurements is plotted. (D) Whole-cell lysates prepared from HEK-293T cells expressing LepH3-C_{out}-gGFP or LepH3-C_{in}-gGFP were subjected to Endo H digestion, SDS-PAGE and Western blot analysis. ●●●● Denotes quadruply glycosylated (C_{out}), ● singly glycosylated (C_{in}) and ○ non-glycosylated form. M marks the standard molecular weight marker loaded for analysis[112].

III.1.4. Topology assessment of disease related proteins by gGFP

Mammalian gGFP was tagged to the C-terminus of Up-Regulated Gene 7 (URG7), Multi-drug Resistance Protein 6 ($MRP6_{102}$), SP-C(Val) and SP-C(Leu) to assay their orientation in HEK-293T cells (Figure 18A). Three out of four test proteins, $MRP6_{102}$, SP-C(Val) and SP-C(Leu), exhibited no fluorescence (Figure 18B, C). Endo H digestion analysis revealed that the proteins were efficiently glycosylated, thereby assaying the fusion joint (the localization of gGFP) to the ER lumen (Figure 18D). In contrast, URG7 exhibited fluorescence, suggesting that its C-terminus is located in the cytosol (Figure 18B, C). The C-terminus of URG7 can reside in the cytosol in two different forms, untargeted and membrane inserted form in an N_{out}/C_{in} orientation. Endo H digestion of URG7 lysate showed that both glycosylated and unglycosylated forms are present *in vivo* (Figure 18D). Hence, it is concluded that URG7 targeting *in vivo* is not efficient, leaving about 50% in the cytosol, but once targeted, it is oriented with an N_{out}/C_{in} orientation.

In contrast to our gGFP study of SP-C where the protein is inserted in an N_{in}/C_{out} form, the *in vitro* topology study with microsomes revealed that SP-C is embedded in the membrane in two different orientations (unpublished data). To test whether the discrepancy between the results from two systems arises from the presence or absence of the quality control system, HEK-293T cells were treated with proteasome inhibitor MG132. While the majority of SP-C was inserted into the membrane in an N_{in}/C_{out} form, the unglycosylated SP-C was also detected in the presence of MG132 (Figure 19A). These results suggest that SP-C is either inserted in two different membrane orientations or some were not targeted to the membrane, and the cells remove the incorrectly inserted, N_{out}/C_{in} form, or an untargeted form by the proteasome *in vivo*.

To distinguish these two possibilities, an additional N-linked glycosylation site was engineered at the N-terminus of SP-C and the protein was expressed in HEK-293T cells. If the unglycosylated SP-C was incorrectly inserted as N_{out}/C_{in} form, with an N-linked glycosylation site at the N-terminus, it would be glycosylated. However, unglycosylated product was still detected, thus suggesting that a small fraction of SP-C is untargeted *in vivo* and removed by the proteasome[112] (Figure 19A, B).

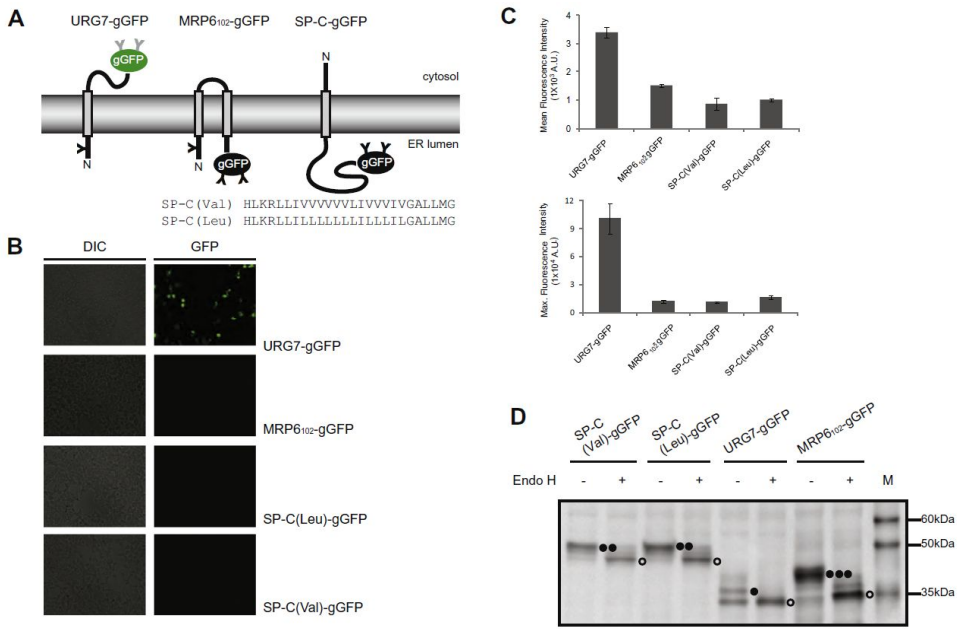


Figure 18. *In vivo* topology assessment of URG7, $MRP6_{102}$, SP-C(Val) and SP-C(Leu). (A) Schematics of URG7, $MRP6_{102}$, SP-C(Val) and SP-C(Leu) with gGFP tag in the ER membrane. The TMS sequences of SP-C(Val) and SP-C(Leu) are shown. (B) Fluorescence microscopy of HEK-293T cells expressing URG7-gGFP, $MRP6_{102}$-gGFP, SP-C(Leu)-gGFP and SP-C(Val)-gGFP. Cells were viewed as described in Figure 16B. (C) Mean (top) and maximum (bottom) fluorescence of URG7, $MRP6_{102}$, SP-C(Val) and SP-C(Leu) with gGFP. The cells were analyzed by FACS Canto and the average of three independent measurements plus the standard deviations is shown. (D) Whole-cell lysates prepared from HEK-293 cells expressing URG7, $MRP6_{102}$, SP-C(Val) and SP-C(Leu) with gGFP were subjected to Endo H digestion, SDS–PAGE and Western blot analysis. ●●● Denotes triply glycosylated form of $MRP6_{102}$-gGFP, ●● doubly glycosylated form of SP-C(Leu/Val)-gGFP, ● singly glycosylated form of URG7-gGFP and ○ non-glycosylated form (N_{out}/C_{in}). M marks the standard molecular weight marker loaded for analysis[112].

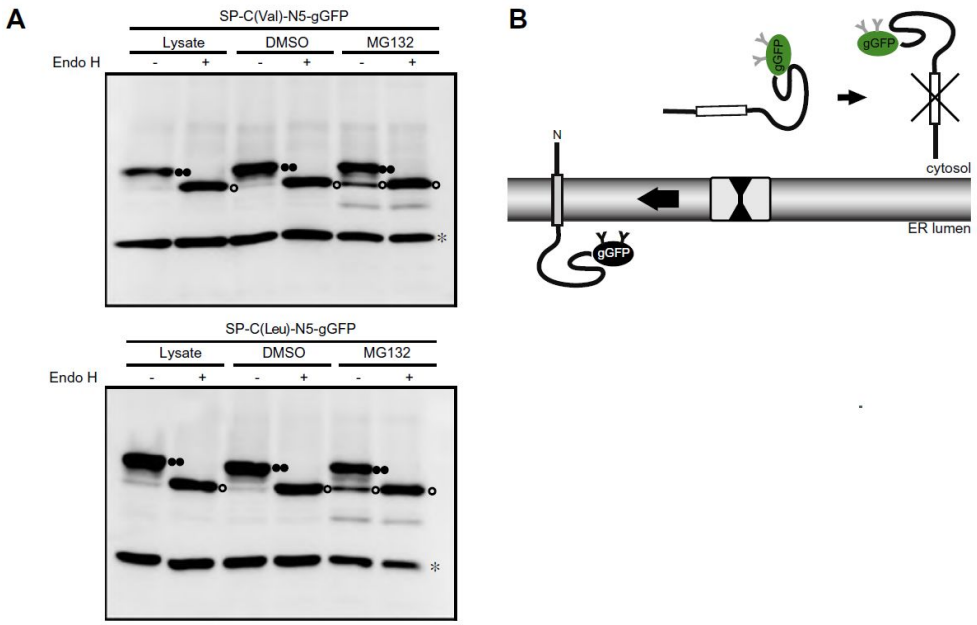


Figure 19. SP-C is inserted into the membrane in N_{in}/C_{out} orientation *in vivo* upon targeting to the membrane. (A) Whole-cell lysates prepared from HEK-293T cells expressing SP-C(Leu/Val)-N5-gGFP in the presence or absence of MG132 were subjected to Endo H digestion, SDS–PAGE and Western blot analysis. SP-C(Leu/Val)-N5-gGFP contains an additional glycosylation site at the N-terminus compared to SP-C(Leu/Val)-gGFP used in Figure 17. for detection of proteins in N_{out}/C_{in} form. ●● Denotes a doubly glycosylated form (N_{in}/C_{out}) and ○ a non-glycosylated form (Untargeted). * Denotes an unspecific band. (B) Schematic representation of SP-C biogenesis. Untargeted SP-C is degraded via the proteasome dependent pathway *in vivo* leaving only the N_{in}/C_{out} form in the membrane[112].

III.2. Structural profiling of the lateral gating helices of yeast Sec61

III.2.1. The residues lining TMS7 of Sec61 affects the targeting and insertion of a single-spanning membrane

To probe which residues in TMS2b and TMS7 of Sec61 are critical for the lateral gating function, systematic mutagenesis of these helices was performed. Residues 82-99 in TMS2b and 290-308 in TMS7 were substituted with an Ala residue one at a time[107]. I specifically profiled the TMS7 residues of Sec61.

First, the efficiency of targeting and membrane insertion of LepH1, a model single-spanning membrane protein, was determined. The TMS was replaced with H-segment of 5L/14A as it enables sensitive detection of the insertion efficiencies of Sec61 gate mutants. A plasmid encoding LepH1(5L/14A) was transformed into the Sec61 mutant strains, and each transformant was subjected to 5 min of S^{35}Met radiolabeling and subsequent immuno-precipitation and analyzed by autoradiography (Figure 20A). The LepH1(5L/14A) segment is moderately hydrophobic and has 48% targeting efficiency (glycosylated products/total products) in WT Sec61[107].

Compared with WT, the targeting efficiency of the model protein was increased by 10% for the I293A, F304A, and S307A Sec61 TMS7 mutants. Interestingly, in these mutants the model protein that was integrated into the membrane with an $N_{in}C_{out}$ orientation was enhanced (45-54% in the mutants compared with 33% in WT) (Figure 20B). Another group of Sec61 mutants decreased the insertion of the Lep-H1 model protein. In Sec61 TMS7 mutants N290A, P292A, L295A, and N302A, 20% less of the LepH1 model protein was integrated into the membrane compared with WT. These mutations also decreased the relative amount of LepH1 with an $N_{in}C_{out}$ membrane topology[107] (Figure 20B).

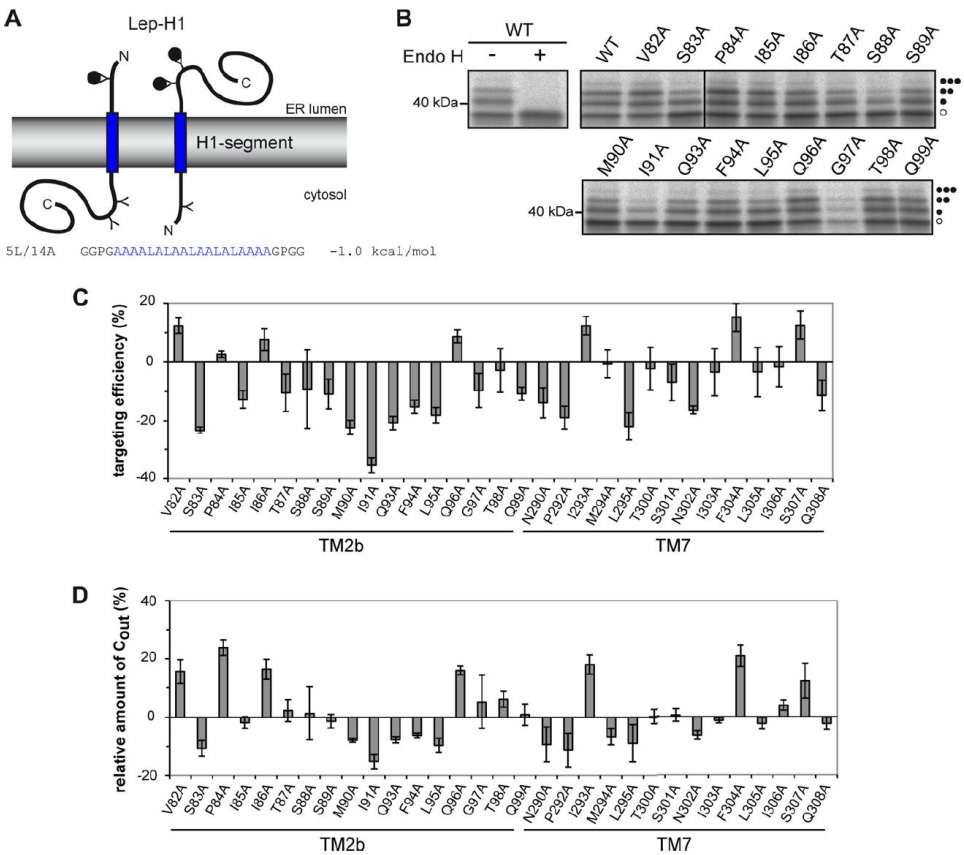


Figure 20. The targeting and membrane insertion of a single-spanning membrane protein in Sec61 gating helix mutants. (A) LepH1(5L/14A) model protein. H-segment sequence with its apparent free energy of insertion (ΔG) is indicated. (B) WT cells expressing the model protein were radiolabeled with ^{35}S Met for 5 min and subjected to immuno-precipitation and SDS-PAGE and analyzed by autoradiography. Before SDS-PAGE, one sample was split into two samples, one of which was treated with Endo H to remove all glycans. Different glycosylation states are indicated with closed circles. Open circle indicates an unglycosylated form. (C) Deviations of Lep-H1(5L/14A) membrane targeting efficiencies of the Sec61 mutant strains normalized to the WT level are plotted. The targeting efficiencies were calculated as the amount of glycosylated products over the total products. The average plus S.E. for at least three independent measurements is shown. (D) Relative amount (%) of Lep-H1(5L/14A) with an N_{in}-C_{out} membrane topology was calculated as 2G/(1G + 2G) × 100% using the same data set

in (B) and normalized against the WT level[107].

III.2.2. The residues lining TMS7 of Sec61 affects the insertion of a 2nd TMS of a double spanning membrane protein

The TMS in single-spanning membrane proteins initiates translocation, opens the Sec61 lateral gate, and positions itself in the $N_{in}C_{out}$ or $N_{out}C_{in}$ orientation before membrane insertion. However, the mechanism of the membrane insertion of the downstream TMS in multi-spanning membrane proteins may differ because the initial opening of the channel is not required[107].

To test this, LepH2 model protein was used. LepH2 contains two hydrophobic segments with two N-linked glycosylation sites, one at the N terminus and the other at the downstream of the second TMS. The first TMS is the original first TMS of *Escherichia coli* Lep and is shown to efficiently target the protein to the ER[69]. The second segment was replaced by H-segment, consisting of 2L/17A (H2-2L) or 3L/16A (H2-3L) (Figure 21A). We selected H-segments with 2L/17A and 3L/16A, the H-segment membrane insertion efficiency of which are 20 and 70%, respectively.

For WT and all the mutant strains, the model protein was efficiently targeted, as little unglycosylated products were detected. For H2-2L and H2-3L, compared with the WT level, I observed a group of mutations that decreased (N290A, P292A, L295A, S301A, N302A, and Q308A) the membrane integration of the H-segment (Figure 21 A, B). In the I293A and F304A Sec61 mutants, targeting and membrane insertion of the Lep-H1 protein were defective, but membrane insertion of the Lep-H2 protein was only mildly impaired. In the S301A and Q308A Sec61 mutants, membrane insertion of the LepH2-3L model protein was reduced by 30% compared with the WT, whereas the targeting defect with the LepH1 model protein was moderate. It was noticed that the targeting and membrane insertion of LepH1 was more severe in the TMS2b mutants compared with the TMS7 mutants. However, the membrane insertion of the H-segment in LepH2-3L construct was more severely affected in the TMS7 mutants than the TMS2b mutants. These results may indicate that the initial binding of a SAS may be mediated more by TMS2b, whereas the membrane insertion of a succeeding TMS is more influenced by TMS7[107].

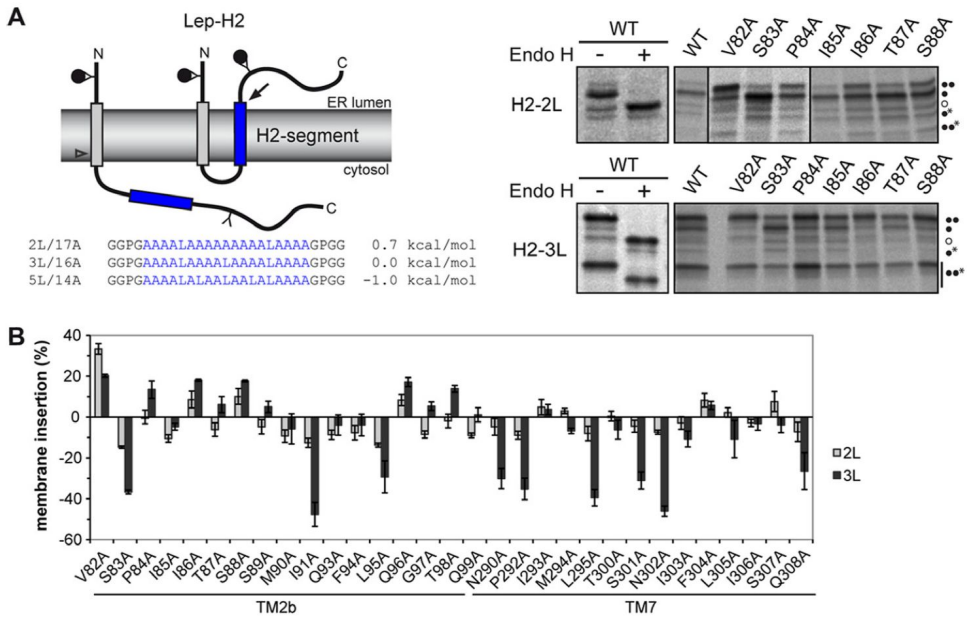


Figure 21. Membrane insertion of the second TMS of LepH2 model proteins in the Sec61 TMS2 and TMS7 mutants. (A) LepH2 has two potential TMS. H2-segment is shown as a blue rectangular box, and sequences with their apparent free energy of insertion (ΔG) are indicated. N-Linked glycosylation sites are indicated as Y, and occupied sites are shown with closed circles. Open and closed arrows indicate the cleavage sites for cleaved products, 1G* and 2G*, respectively. Cells were radiolabeled for 5 min with ^{35}S Met, subjected to immuno-precipitation and SDS-PAGE, and analyzed by autoradiography. Endo H treatment was carried out to remove all glycans before SDS-PAGE. Different glycosylation states are indicated with *closed circles*, including different cleaved products indicated with an asterisk (right). The open circle indicates an unglycosylated full-length protein that is untargeted. (B) H2-segment membrane insertion efficiency (%) was calculated as (2G + 2G*)/(1G + 2G + 1G* + 2G*) × 100% and normalized against the WT level. The average of at least three independent measurements plus S.E. is shown. (C) LepH2-2L and LepH2-3L were expressed in Sec61 WT and double mutant strains and analyzed as described in (A). H2-segment membrane insertion (%) was calculated as 2G/(1G + 2G) × 100%. (D) LepH2-5L was expressed in Sec61 WT and I83A/I91A double mutant cells. Cells were treated as in (A), and membrane insertion of H2-segment was calculated as in (B)[107].

III.3. Profile of membrane protein insertion into mammalian ER

III.3.1. Membrane insertion profile of single-spanning membrane proteins

Unlike *Saccharomyces cerevisiae* where signal anchor proteins are targeted post- or co-translationally depending on the hydrophobicity of the SAS, a signal anchor protein targeting to the ER is shown to depend on the protein length in mammalian cells[51,110].

To study the targeting and insertion of signal anchor proteins in mammalian cells in greater detail, a set of LepH1 with variable hydrophobicity and C-terminal length were expressed in HEK-293T. The targeting of LepH1 was shown with H-segment with 3L/16A, and were inserted mainly in 1G form ($N_{out}C_{in}$). The insertion profile was overall similar in HeLa and HEK-293T cells for LepH1 with a marginally hydrophobic SAS (Figure 22A). The more hydrophobic LepH1 with H-segment of 7L/12A and 10L/9A were mainly inserted in 2G form ($N_{in}C_{out}$) in HeLa cells. However, the pulse labeling of LepH1 in HEK-293T revealed that 10L SAS is inserted in 1G form ($N_{out}C_{in}$) (Figure 22B). Whether the differential insertion of LepH1-10L is a result of cell to cell variation (HeLa vs HEK-293T) or the protein detection technique (Western blotting vs pulse labeling) needs to be re-investigated.

To confirm the glycosylation status of LepH1, Endo H digestion was performed. Strikingly, Endo H digestion revealed an Endo H sensitive product whose deglycosylated form was smaller than the non-targeted, unglycosylated full-length protein. The cleaved product was prevalent in LepH1 with a marginally hydrophobic TMS, 3L/16A to 6L/13A (Figure 22A, B). We speculate that LepH1 with a marginally hydrophobic TMS is a substrate of signal peptidase complex (SPC) as LepH2 gets processed by SPC[107]. If the assumption is true and that the C-terminal end of the TMS is cleaved, this implies that the majority of LepH1 is inserted as 2G form ($N_{in}C_{out}$) and is subsequently processed by SPC. In sum, LepH1 is primarily inserted in $N_{in}C_{out}$ form, with an exception of LepH1-10L/9A in HEK-293T where the majority was inserted as $N_{out}C_{in}$ form. $N_{in}C_{out}$ form is processed by SPC in hydrophobicity dependent manner, with a less hydrophobic TMS being cleaved more efficiently than a more hydrophobic TMS.

Next, the truncated LepH1 composed of 121 amino acids were expressed in HEK-293T to study the protein targeting by post-translational pathway. It is previously reported that polypeptides shorter than 120 amino acids are targeted post-translationally in mammalian cells[51]. Just like the full-length LepH1, marginally hydrophobic signal anchors (5L/14A and 6L/13A) were cleaved whereas hydrophobic signal anchor (7L/12A) was not (Figure 22C). Thus, regardless of the targeting pathways, the topogenesis of signal anchor protein and the subsequent cleavage of SAS is unaffected.

Lastly, the effect of the C-terminal length following the TMS on membrane protein topogenesis was examined. It was previously reported that a long C-terminal tail favors the protein topology of $N_{in}C_{out}$ whereas a short C-terminal tail favors $N_{out}C_{in}$ orientation[111]. LepH1 fused to yeast gGFP was expressed in HeLa or HEK-293T. Yeast gGFP was non-functional in mammalian cells and thus regarded as an extension of the C-terminus[112]. Like full length LepH1 or truncated LepH1, the targeting to the ER was observed for LepH1 with 3L/16A H-segment (Figure 22D). However, the orientation of LepH1-gGFP relative to the membrane differed when expressed in HEK-293T, with the majority of LepH1-gGFP inserted in 1G form ($N_{out}C_{in}$) regardless of the hydrophobicity (3L/16A to 7L/12A). No cleavage product was observed (Figure 22D, Middle). In contrast, LepH1-gGFP was cleaved efficiently when expressed in HeLa, and based on Endo H digestion analysis, LepH1-gGFP was inserted in $N_{in}C_{out}$ form (Figure 22D, Bottom). The experiment needs to be repeated to confirm cell to cell variation. Based on the data obtained from HeLa cells, it is likely that the subsequent cleavage of LepH1 is independent of the C-terminal tail length and occurs when LepH1 is inserted in $N_{in}C_{out}$ form.

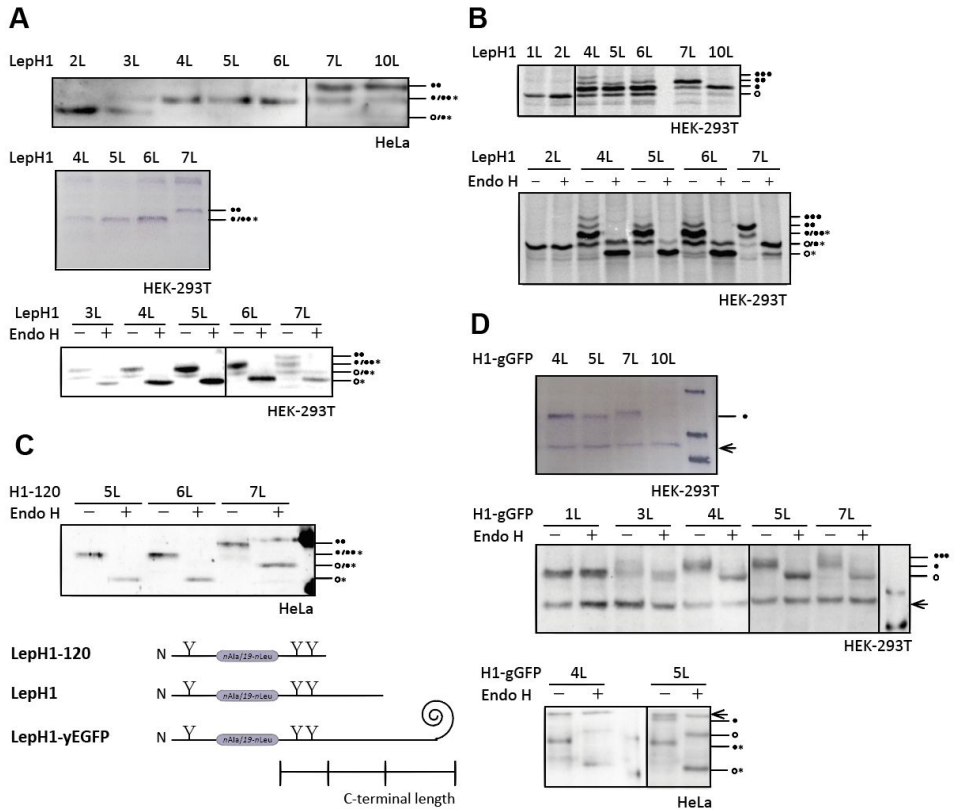


Figure 22. The targeting and insertion of a single-spanning membrane protein in mammalian cells. LepH1 full length, truncated and gGFP tagged constructs were expressed in HeLa or HEK-293T. (A) Top. HeLa cells expressing LepH1 were lysed with NP-40, separated on SDS-PAGE, Western blotted and immuno-decorated with α-HA. Middle. As in Top, but LepH1 was expressed in HEK-293T. Bottom. Endo H digestion of the samples from Middle. (B) Top. HEK-293T cells expressing LepH1 were pulsed labeled for 20 minutes, immuno-precipitated and subjected to autoradiography. Bottom. Endo H digestion of the samples from Top. (C) As in (A) Top, except that the expressed construct is the truncated LepH1 (H1-120). The schematics of LepH1 full length, truncated and gGFP tagged constructs are depicted. (D) As in (A), except that the expressed construct is LepH1-gGFP. ○ non-glycosylated, ● glycosylated, * cleavage, ←non-specific band.

III.3.2. Membrane insertion profile of multi-spanning membrane proteins

To investigate the insertion of multi-spanning membrane proteins, LepH2 and LepH3 were expressed in HEK-293T. LepH2 is a double spanning membrane protein and has a test segment composed of *n*-leucine and *19-n* alanine as the 2nd TMS. LepH3 has one more TMS when compared to LepH2 and has a test segment as the 3rd TMS. The insertion of the test segment is monitored by the glycosylation status, with LepH2 being singly glycosylated (1G) when the test TMS is non-inserted and doubly glycosylated (2G) when inserted. For LepH3, it is 2G when the test TMS is non-inserted and 1G when inserted (Figure 23A).

LepH2 with a test segment of 2L/17A and 3L/16A were efficiently targeted to the ER, inserted into the membrane and cleaved efficiently (Figure 23A). In comparison, LepH1-2L could not be targeted/inserted into the ER membrane. This agrees with the previous report that once the protein is targeted to the ER, the hydrophobicity required for the insertion of a downstream TMS is not high[107,109]. Pulse labeling revealed that the cleavage of LepH2 occurs in time dependent manner (Figure 23B). When cells expressing LepH2 were labeled for 5 minutes, no cleavage product was observed for LepH2-2L and a little for LepH2-3L. However, when the cells were labeled for 20 minutes, more than half of membrane inserted proteins (2G forms) were cleaved (Figure 23B).

LepH3 with a test segment of 19L efficiently targeted to the ER and inserted into the membrane. However, LepH3 with a test segment of 3L/16A or 5L/11A were efficiently targeted to the ER, but the insertion of the test segments into the membrane were very poor (Figure 23C, D). In comparison, a test segment with 2L/17A was efficiently inserted into the membrane when placed in the context of LepH2 (Figure 23A, B). As the key difference between LepH2 and LepH3 is the number of TMS which makes the orientation of the test segment different upon membrane insertion, the results suggest that the hydrophobicity required for a TMS insertion is much lower when the insertion occurs in $N_{in}C_{out}$ orientation (as in LepH2) than $N_{out}C_{in}$ (as in LepH3).

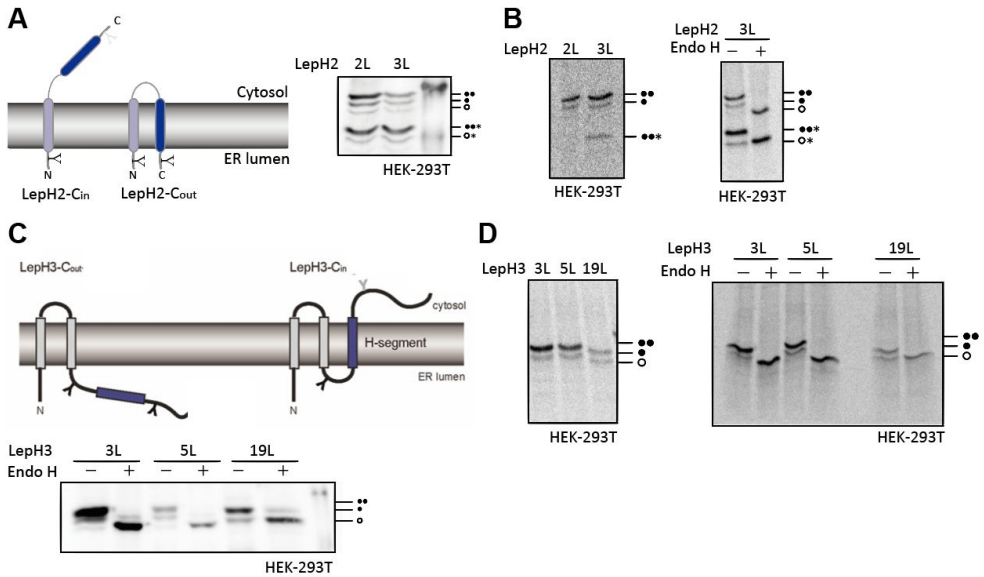


Figure 23. The insertion of 2nd or 3rd TMS into the membrane. LepH2 and LepH3 were expressed in HEK-293T. (A) Left. Schematics of LepH2. Right. HEK-293T cells expressing LepH2-2L or 3L were lysed with NP-40, separated on SDS-PAGE and immuno-decorated with α-HA. (B) Left. HEK-293T cells expressing LepH2-2L or 3L were pulsed labeled for 5 minutes, immuno-precipitated and subjected to autoradiography. Right. Endo H digestion of 20 minute pulse labeled LepH2-3L from HEK-239T cells. (C) Top. Schematics of LepH3. Bottom. HEK-293T cells expressing LepH3-3L, 5L or 19L were lysed with NP-40, separated on SDS-PAGE, Western blotted and immuno-decorated with α-HA. (D) Left. HEK-293T cells expressing LepH3-3L, 5L or 19L were pulsed labeled for 20 minutes, immuno-precipitated and subjected to autoradiography. Right. Endo H digestion of the samples from Left. ○ non-glycosylated, ● glycosylated, * cleavage.

Membrane insertion of a TMS may differ for a TMS that is found in the midst of a multi-spanning membrane protein than a TMS that serves as the last TMS. To test this, LepIV and LepV, which span the membrane four and five times with the test segment in the 3rd and 4th TMS respectively, were expressed in HEK-293T. The test segment is inserted in $N_{out}C_{in}$ in LepIV and $N_{in}C_{out}$ in LepV (Figure 24).

This case, unlike LepH2 and LepH3, the orientation of the test TMS did not have a significant influence on the threshold hydrophobicity for membrane insertion. Up to 4L/15A, the membrane insertion efficiency of the test TMS was low and 7L/12A was required for an efficient insertion both for LepIV and LepV (Figure 24).

To study the insertion of the test segment in the absence of the last TMS, LepIV and LepV were truncated (LepIV TC and LepV TC, respectively). Membrane insertion of the test segment was complete with 7L/12A in both cases, and partial with 4L/15A test segment (Figure 25). Again the insertion of the 4L/15A test segment was more efficient in LepV TC where the test segment insertion occurs in $N_{in}C_{out}$ compared to LepIV TC where the test segment insertion occurs in $N_{out}C_{in}$. This is in agreement with the previous observation with LepH2 and LepH3.

In sum, the insertion of the last TMS occurs more efficiently than the TMS that precedes the last TMS for unknown reasons. For the insertion of the last TMS of a multi-spanning membrane protein, the insertion in $N_{in}C_{out}$ requires less hydrophobicity than insertion in $N_{out}C_{in}$ form. The experiments need to be repeated to confirm the observations.

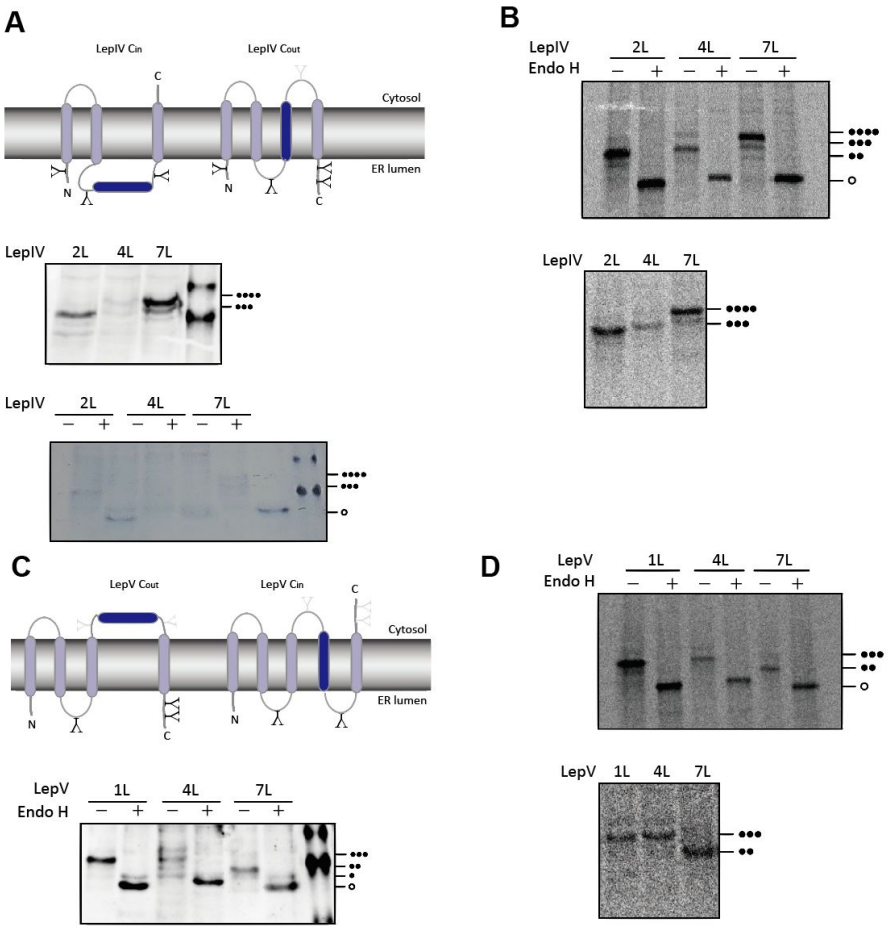


Figure 24. The insertion of a TMS in the middle of multi-spanning membrane proteins. LepIV and LepV were expressed in HEK-293T. (A) Top. Schematics of LepIV. Middle. HEK-293T cells expressing LepIV-2L, 4L or 7L were lysed with NP-40, separated on SDS-PAGE, Western blotted and immuno-decorated with α-HA. Bottom. Endo H digestion of the samples from middle. (B) Top. HEK-293T cells expressing LepIV-2L, 4L or 7L were pulsed labeled for 20 minutes, immuno-precipitated and subjected to autoradiography. Where indicated Endo H digestion was performed. Bottom. The same as in Top. Endo H untreated samples were run side by side for size comparison. (C) Top. Schematics of LepV. Middle. HEK-293T cells expressing LepV-1L, 4L or 7L were lysed with NP-40, separated on SDS-PAGE, Western blotted and immuno-decorated with α-HA. Where indicated, Endo H digestion was performed. (D) The same as in (B) except that the expressed model proteins are LepV-1L, 4L and 7L. ○ non-glycosylated, ● glycosylated.

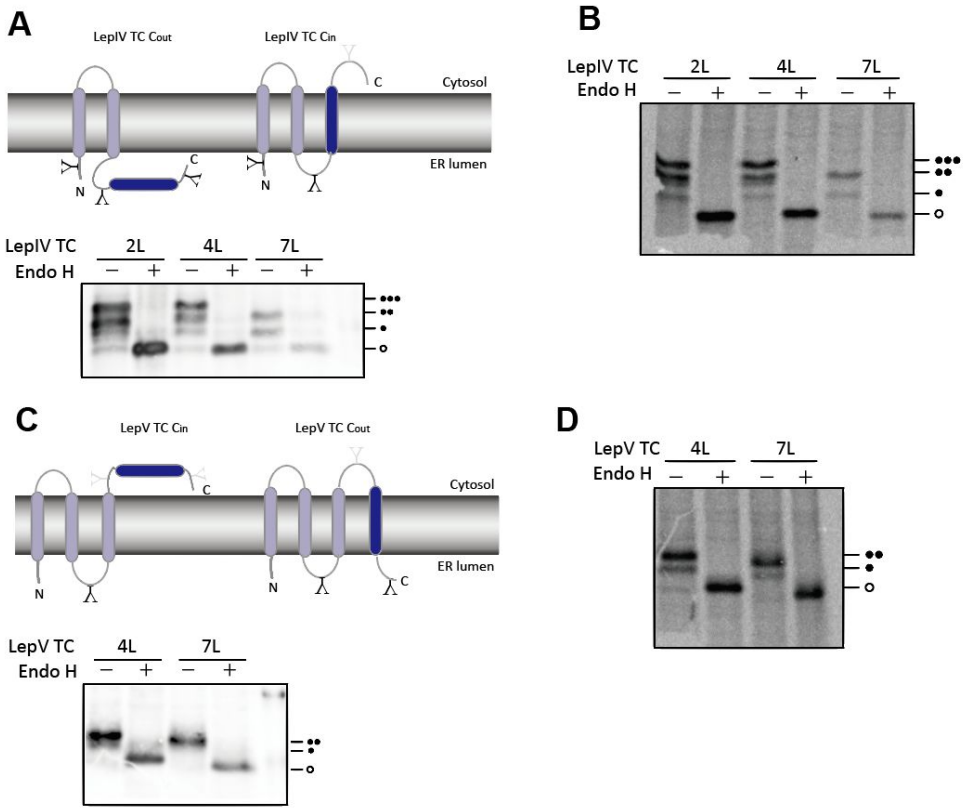


Figure 25. The insertion of 3rd or 4th TMS. LepIV-TC and LepV-TC were expressed in HEK-293T. (A) Top. Schematics of LepIV-TC. Bottom. HEK-293T cells expressing LepIV-TC-2L, 4L or 7L were lysed with NP-40, separated on SDS-PAGE, Western blotted and immuno-decorated with α-HA. Endo H was treated where indicated. (B) HEK-293T cells expressing LepIV-TC-2L, 4L or 7L were pulsed labeled for 20 minutes, immuno-precipitated and subjected to autoradiography. Where indicated Endo H digestion was performed. (C) Top. Schematics of LepV-TC. Bottom. HEK-293T cells expressing LepV-TC-4L or 7L were lysed with NP-40, separated on SDS-PAGE, Western blotted and immuno-decorated with α-HA. Where indicated, Endo H digestion was performed. (D) The same as in (B) except that the expressed model proteins are LepV-TC-4L and 7L. ○ non-glycosylated, ● glycosylated.

III.3.3. siRNA mediated silencing of Sec62

To functionally characterize the subunits of the mammalian SEC61 complex, siRNA mediated knock down of individual subunits was designed. The scheme was to deplete the target component of SEC61 by siRNA followed by transient expression of a model membrane protein. The insertion of a model membrane protein in the absence of the chosen subunit of SEC61 would uncover the role of the silenced subunit in membrane protein biogenesis.

As a starting point, Sec62 silencing was attempted. Sec62 was reported to facilitate the C-terminal translocation of a marginally hydrophobic TMS of membrane proteins in yeast[16,17]. To figure out whether the Sec62's role is conserved in mammalian cells, siRNA targeting the untranslated region (UTR) of SEC62 mRNA were introduced into cells and the silencing was monitored by Western blotting with Sec62 and GAPDH antibodies (control). 70% silencing efficiency (the amounts of Sec62 normalized by GAPDH were compared) was achieved when the cells were incubated with 50 nM siRNA for 48 hours (Figure 26). Transfection of a model protein at 48 hour-time point is required to observe the effect of Sec62 silencing.

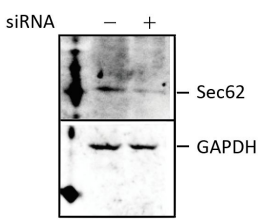


Figure 26. Sec62 knock down. HEK-293T were transfected with a scrambled siRNA (negative control) or siRNA targeting SEC62 mRNA. The cells were maintained for 48 hours after transfection and lysed with NP-40. Cell lysates were analyzed by SDS-PAGE and Western blotting. Sec62 and GAPDH were detected with specific antibodies.

III.3.4. Rescue of Sec62 knock-down

Although a negative control siRNA was used, the possibility of an off-target effect by SEC62 siRNA still existed. To confirm that the observed phenotype was a direct consequence of Sec62 silencing, a rescuing experiment was designed.

Here, after the silencing of Sec62 by transfection of siRNA targeting the UTR of SEC62, a plasmid encoding cDNA of SEC62 would be transfected. Total RNA extracted from HEK-293T was used to make cDNA library and SEC62 was amplified using gene specific primers. The cDNA of SEC62 was subsequently cloned into pcDNA 3.1 plasmid. The rescue of the observed defect by the expression of SEC62 from the plasmid would rule out the off-target effect.

The designed rescuing scheme also enables the identification of an active domain of Sec62 involved in membrane protein biogenesis by systematic mutagenesis on cDNA of SEC62. The mutant Sec62 that is incapable of rescuing the defect would highlight the key domain of Sec62.

III.3.5. Expression of human translocon subunits in yeast

Mammalian SEC61 complex has more subunits compared to yeast SEC61, yet most studies on SEC61 complex were done in yeast due to the ease of genetic manipulation.

In the thesis, rescue of the SEC61 subunit deletion strains of *Saccharomyces cerevisiae* were attempted by the expression of mammalian homologues. The complementation assay would provide a direct evidence of a conserved function of proteins. SEC61α, SEC62 and TRAM1 were amplified from cDNA library and cloned into p424GPD vector by homologous recombination. Due to time constraints, the functional complementation assay has not been performed.

III.4. Lateral gating of TIM23 by Mgr2 is limited to Mgm1 and Cyb2

III.4.1. Expression level of Mgr2 is low

Mgr2 was shown to regulate the membrane insertion of Mgm1 and Cyb2. Thus, it was proposed as a lateral gate keeper of the TIM23 complex[104]. To study the role of Mgr2 in the gate keeping of TIM23 complex, Mgr2 deletion (*mgr2*Δ), Mgr2 overexpression (Mgr2↑) and the corresponding WT strain were obtained from Dr. Pfanner[104]. To confirm the strains, cells were cultured in a fermentable carbon source (glucose) and the proteins were prepared for Western blotting. Mgr2 was properly overexpressed in Mgr2↑, however, the endogenous Mgr2 was barely detectable in WT strain (Figure 27A). WT and *mgr2*Δ strains were further confirmed by PCR.

Mitochondrial health is at optimum when cells respire. To check the expression of Mgr2 at optimal mitochondrial health, cells were grown in non-fermentable carbon source (glycerol) and the proteins were prepared for Western blotting. However, the expression level of Mgr2 remained low and was indifferent from the cells grown in fermentable carbon source (Figure 27B). Thus, Mgr2 is expressed at a low level in *Saccharomyces cerevisiae* regardless of the carbon source.

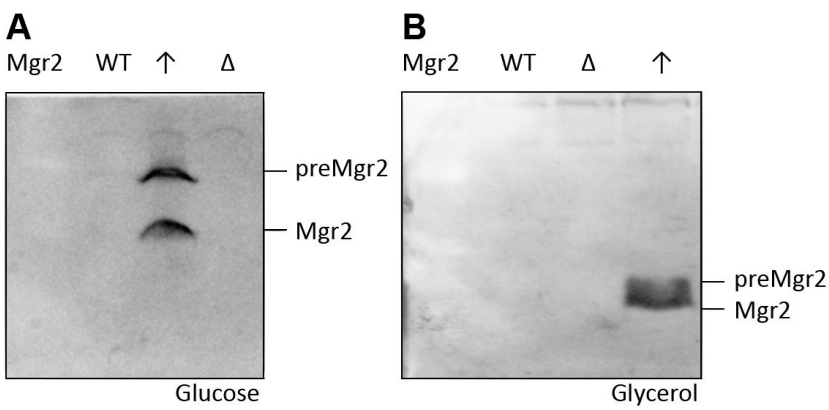


Figure 27. Mgr2 expression in WT, *mgr2*Δ and Mgr2↑. (A) Proteins were TCA precipitated from overnight cultures of WT, *mgr2*Δ and Mgr2↑ in glucose. Proteins were separated on 16% Tricine gel to distinguish 16kDa preMgr2 from 9kDa Mgr2. The protein bands were detected using α-Mgr2. (B) The same as in (A) except that

the cells were grown in glycerol and the proteins were separated on 15% Tris-Cl gel. WT, Wildtype; Δ, *mgr2*Δ; ↑, Mgr2↑.

III.4.2. Mgr2 does not set the hydrophobicity requirement for membrane protein insertion into mitochondrial IM

To test whether Mgr2 sets the hydrophobicity requirement for membrane protein insertion into IM, a set of Mgm1 whose 1st TMS is replaced by 19 alanine/leucine stretch (Mgm1-A/L) was expressed in WT, *mgr2*Δ and Mgr2↑. If Mgr2 acts as a lateral gate keeper, the overexpression of Mgr2 would increase the protein translocation (i.e. decrease the insertion) whereas the absence of Mgr2 would have an opposite effect. However, no significant difference was observed in membrane protein translocation efficiency in three strains monitored by the amount of *s*-Mgm1 regardless of the hydrophobicity of the tested TMS (Figure 28A, B). Thus, Mgr2 does not appear to set the hydrophobicity required for the membrane protein insertion into mitochondrial IM.

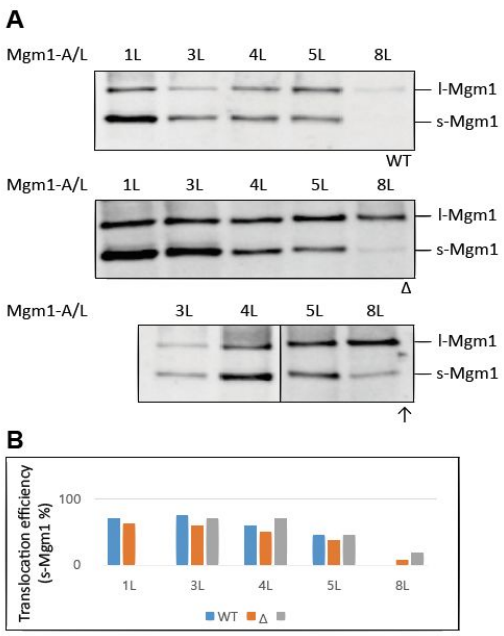


Figure 28. No difference in the membrane insertion of Mgm1 carrying varying hydrophobicity among WT, *mgr2*Δ and Mgr2↑ strains. (A) Mgm1-A/L with indicated L were expressed in WT, *mgr2*Δ and Mgr2↑. Proteins were TCA precipitated from overnight cultures in glucose. Proteins were separated on 6.5% Tris-Cl gel to distinguish *l*-Mgm1 and *s*-Mgm1. The protein bands were detected using α-HA. (B) The quantification of *s*-Mgm1 from (A). WT, Wildtype; Δ, *mgr2*Δ; ↑, Mgr2↑.

III.4.3. Mgr2 does not affect the sorting of mitochondrial IM proteins taking the stop-transfer or conservative sorting pathway

Mgm1-A/L is dislocated from the mitochondrial IM by the mAAA protease[102]. To rule out the effect of the mAAA protease and to determine if Mgr2 affects insertion of other mitochondrial IM proteins, MFPs taking the stop-transfer pathway and conservative sorting pathway were selected[114]. The selected MFPs were derived from native IM proteins and were previously shown to be uninfluenced by the mAAA. In addition, the selected MFPs cover a broad range of hydrophobicity (Table 1).

Stop transfer proteins

Name	ΔG
8L	-2.713
She9 (1TM) MFP	-1.985
7L	-1.760
6L	-1.540
Cox5b MFP	-1.163
5L	-0.999
Oms1 MFP	-0.890
4L	-0.490
Cox5a FL MFP	-0.228
3L	-0.006
Sco2 MFP	0.023
Sco1 MFP	0.479
2L	0.699
Yme1	1.063
1L	1.063
Dld1	1.479

Conservative sorting proteins

Name	ΔG
5L	-0.999
Yta10 1TM MFP	-0.679
Oxa1 1-4TM MFP	-0.549
4L	-0.490
...	
1L	1.063
Cox18 1TM MFP	1.261
Mrs2 1TM MFP	1.334
Mba1 MFP	2.321
Tcm62 MFP	2.842

Table 1. The ΔG values of the 1st TMS of MFPs taking the stop-transfer pathway (left) and conservative sorting pathway (right). ΔG values of Mgm1-A/L for comparison. The TMS of stop-transfer proteins are in general more hydrophobic (less ΔG) than that of conservative sorting proteins.

The hypothesis was that the insertion of stop-transfer proteins would be reduced by Mgr2 overexpression due to the jamming of the lateral gate of the TIM23 complex. On the other hand, the translocation of conservative sorting proteins would be reduced by Mgr2 deletion as the lateral gate remains unoccupied with Mgr2, therefore probably remains in an open conformation which may favor the insertion of conservative sorting proteins. However, the sorting of stop-transfer proteins into IM was unaltered by Mgr2 overexpression as evident by the generation of *l*-MFPs (Figure 29A). The sorting of conservative sorting proteins to the matrix was also unaffected by Mgr2 deletion (apart from Yta10-MFP which was sorted into the membrane regardless of Mgr2 level) (Figure 29B). In sum, Mgr2 has negligible effect on the insertion of mitochondrial IM proteins regardless of the sorting pathway.

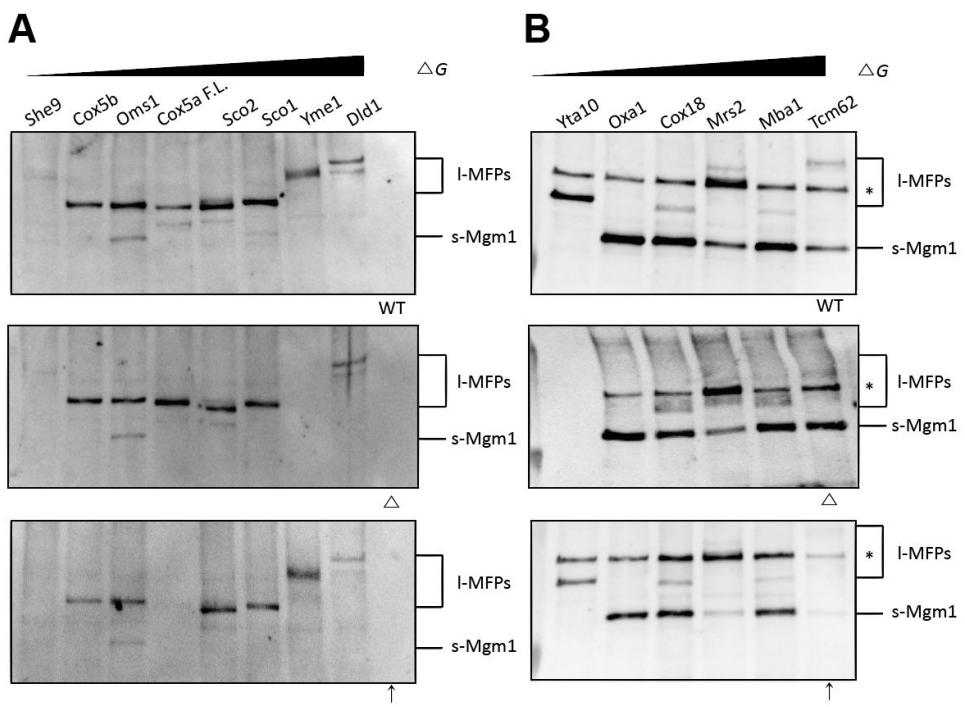


Figure 29. The membrane insertion of MFPs is unaffected in WT, *mgr2*Δ and Mgr2↑. (A) MFPs taking the stop-transfer pathway were expressed in WT, *mgr2*Δ and Mgr2↑. Proteins were TCA precipitated from overnight cultures in glucose. Proteins were separated on 6.5% Tris-Cl gel to distinguish *l*-MFPs and *s*-Mgm1. The protein bands were detected using α-HA. (B) The same as in (A) except that the expressed proteins take conservative sorting pathway. WT, Wildtype; Δ, *mgr2*Δ;

↑, Mgr2↑; *, non-specific.

III.4.4. Mgr2 senses the charges preceding the TMS of Mgm1

The sorting of Mgm1-A/L and MFPs were not influenced by the Mgr2 expression level. This led to a new hypothesis that Mgr2's function may be specific to previously tested proteins, Mgm1 and Cyb2[104]. To confirm this, Mgm1 and Cyb2 were expressed in WT, *mgr2*Δ and Mgr2↑. Membrane insertion of Mgm1 (*l*-Mgm1) was reduced by Mgr2 overexpression (Figure 30A). Surprisingly, in this study, Mgr2 deletion did not increase the amount of *l*-Mgm1. This could be due to the low expression level of Mgr2 in WT making it indistinguishable from *mgr2*Δ. The sorting of Cyb2 was analyzed by Western blotting assay. The processing of Cyb2 monitored by the band size was unaltered regardless of Mgr2 level, however the overall expression level of Cyb2 was affected (Figure 30B). Cyb2 expression was at the highest in the absence of Mgr2 and lowest upon overexpression of Mgr2. A detailed analysis with a control protein is needed to confirm the observation.

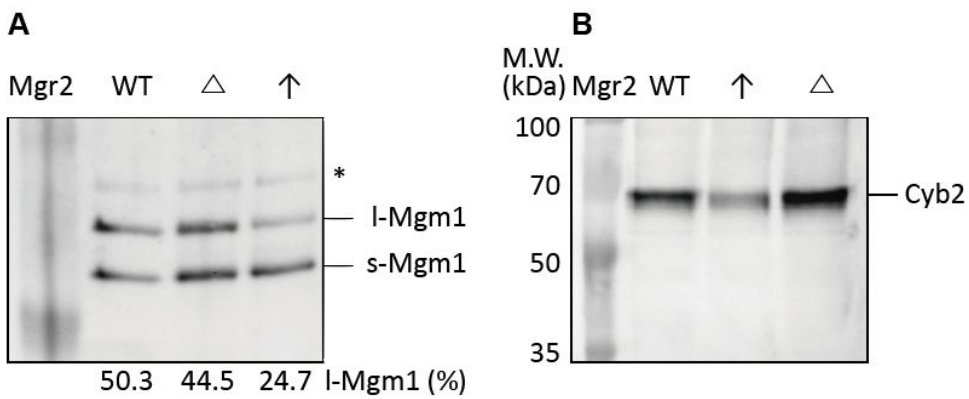


Figure 30. Mgr2 affects the sorting of Mgm1. (A) Mgm1 was expressed in WT, *mgr2*Δ and Mgr2↑. Proteins were TCA precipitated from overnight cultures in glucose. Proteins were separated on 6.5% Tris-Cl gel to distinguish *l*-Mgm1 and *s*-Mgm1. The protein bands were detected using α-HA. (B) The same as in (A) except that the expressed protein is Cyb2-HA. WT, Wildtype; Δ, *mgr2*Δ; ↑, Mgr2↑, *, non-specific.

The sorting of Mgm1's 1st TMS is heavily dependent on the flanking charged residues[115]. As Mgr2 was previously reported to block the insertion of Cyb2 with mutated N-terminal flanking charges, I hypothesized that Mgr2 recognizes the charged residues that flanks the TMS of Mgm1[104].

To test this, the sorting of mutant Mgm1 proteins were analyzed in WT, *mgr2*Δ and Mgr2↑. The mutations either replaced the charged residues to neutral residues or the residues of opposite charges. Mutations at the N-terminus of Mgm1 resulted in Mgr2 independent sorting, with the exceptions of R78A, R78AR79A GGM:VVL mutations (Figure 31A). However, the mutations of the C-terminal flanking charged residues did not change the Mgm1 sorting dependency on Mgr2 (Figure 31B). This suggests that the charged residues that flanks the matrix side of the TMS of Mgm1 plays a key role in Mgr2 regulated Mgm1 insertion into the membrane. This is in great agreement with the previous observation that the mutations of the sorting signal composed of two charged residues at the matrix side is recognized by Mgr2[104]. In addition, the overall expression level of Mgm1 variants was affected. Like Cyb2, the expression was at lowest upon overexpression of Mgr2 (Figure 31A, B). Again, detailed analysis with a control protein is needed to confirm the observation.

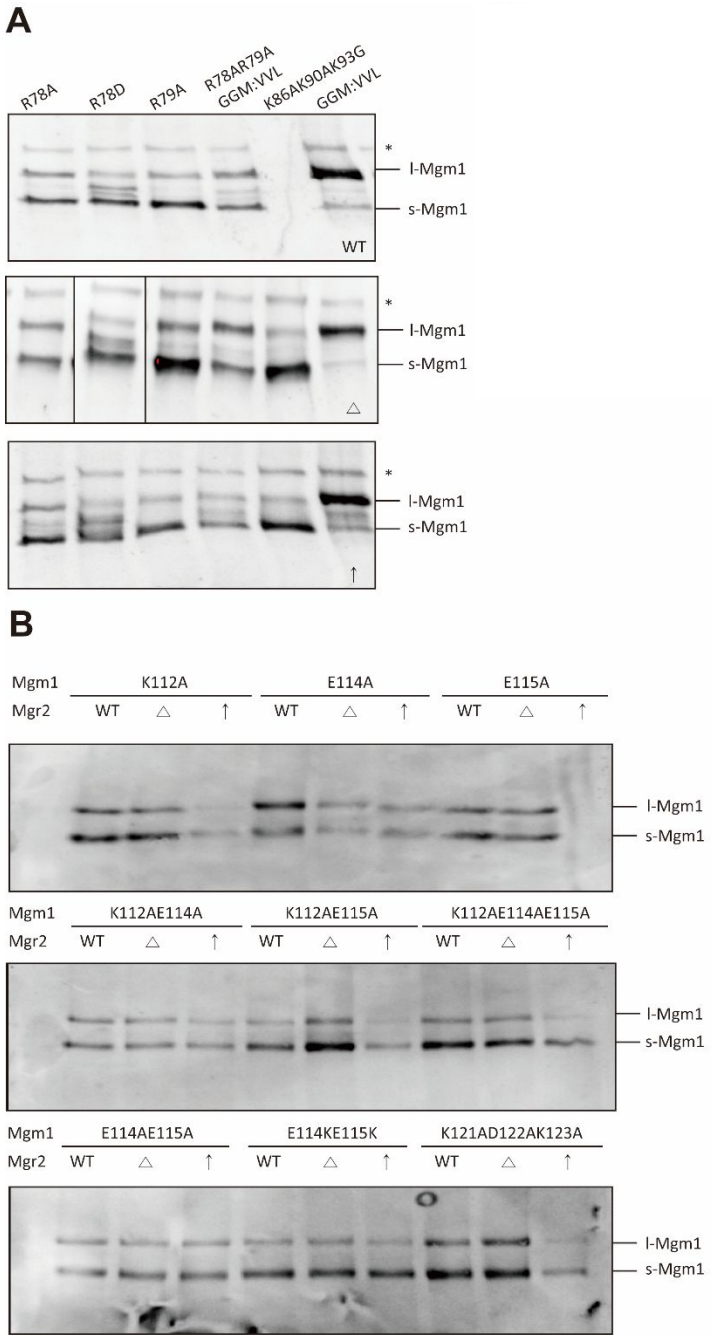


Figure 31. Mgr2 senses the flanking charged residues of Mgm1. (A) The N-terminal flanking charged mutants of Mgm1 were expressed in WT, *mgr2*Δ and Mgr2↑. Proteins were prepared and separated as described in Figure 25. (B) The same as in (A) except that the expressed protein were the C-terminal flanking charged residues. WT, Wildtype; Δ, *mgr2*Δ; ↑, Mgr2↑, *, non-specific.

III.4.5. Mitochondrial IM protein sorting in glycerol is unaffected by Mgr2

Membrane potential affects the mitochondrial protein import and strong membrane potential is built across the mitochondrial IM when the cells are forced to respire in non-fermentable carbon source medium[138]. To test whether the Mgr2's role in regulating the protein entry into the mitochondrial IM is pronounced upon respiration condition, the sorting of a selected MFPs, Mgm1-WT and Mgm1 flanking charged mutants were tested in WT, *mgr2*Δ and Mgr2↑. The sorting of all tested proteins in glycerol media were unaffected by Mgr2 expression level (Figure 32). Thus, this further confirms that Mgr2 does not regulate all of mitochondrial IM protein insertion, rather its role is limited to Mgm1 and Cyb2.

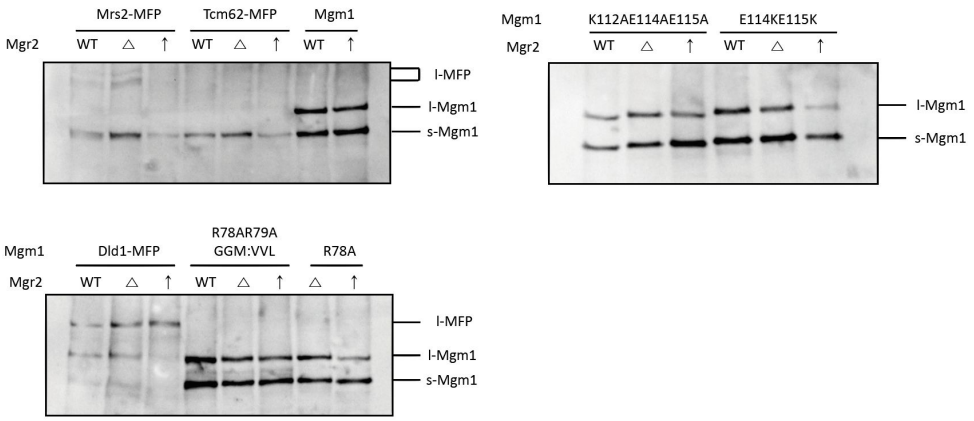


Figure 32. Mitochondrial IM protein sorting in glycerol is unaffected by Mgr2 expression level. The sorting of Mrs2-MFP, Tcm62-MFP, Dld1-MFP, Mgm1-WT, Mgm1-R78A, Mgm1-R78A R79A GGM:VVL, Mgm1-K112AE114AE115A and Mgm1-E114KE115K were tested in WT, *mgr2*Δ and Mgr2↑ cultured in glycerol. Proteins were prepared and separated as described in Figure 25.

III.4.6. Sorting kinetics of mitochondrial IM proteins.

In the previous study, the sorting of Cyb2-DHFR into the mitochondrial IM was delayed upon overexpression of Mgr2 and facilitated in the absence of Mgr2[104]. Western blot gives the homeostatic levels of proteins and is incapable of providing the sorting kinetics. Thus, pulse-chase experiment was done for Mgm1-R78A to analyze the protein sorting during the early biogenesis step. Here, the overexpression of Mgr2 delayed the generation of *s*-Mgm1. It took 20 minutes after the initial pulse labeling to reach about 50% of *s*-Mgm1 (Figure 33). Thus, Mgr2 delays the sorting of proteins to the membrane at least for Mgm1. The sorting kinetics of other MFPs will be tested in the near future.

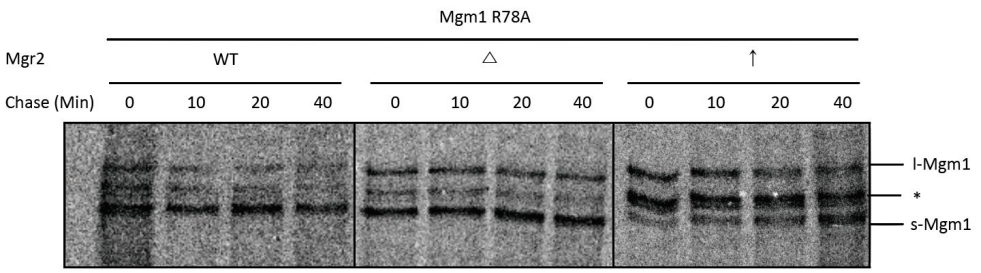


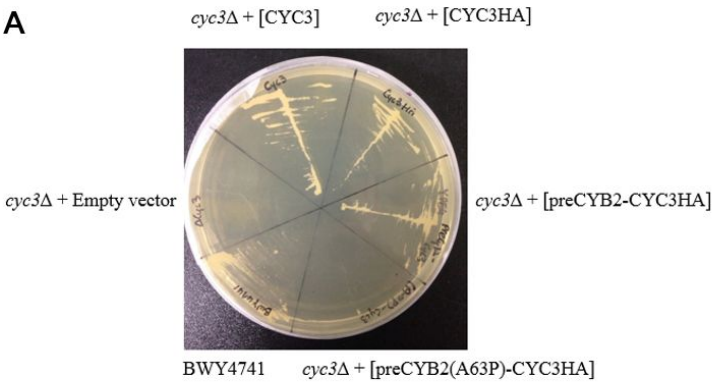
Figure 33. The differential sorting of Mgm1-R78A in WT, *mgr2*Δ and Mgr2↑. Yeast transformants expressing Mgm1-R78A was pulsed labeled for 10 minutes at 30° C and chased for 10, 20 or 40 minutes. The samples were immuno-precipitated and subjected to autoradiography. An unknown band, marked as *, was detected in between *l*-Mgm1 and *s*-Mgm1 and was evident in Mgr2↑.

III.5. TIM23 mutagenesis screening

III.5.1. Validation of TIM23 mutagenesis selection scheme

The selection scheme described in I.4.4. was validated by the growth assay. *cyc3*Δ did not grow on non-fermentable carbon source. The growth defect was recovered by expressing CYC3, CYC3HA and preCyb2-CYCHA but not with an empty vector or a vector encoding preCyb2(A63P)-CYC3HA (Figure 34). The same result was obtained for *tim23*Δ*cyc3*Δ[pRS316TIM23]. As preCyb2-CYC3HA supports the growth once it is in the membrane, the transformants expressing preCyb2(A63P)-CYC3HA, in theory, is capable of growing on non-fermentable carbon source if there exists a mutant Tim23 that is capable of inserting the fusion protein. Both random and systematic mutagenesis of Tim23 is being done.

A



B

	WT (BWY4741)	*cyc3*Δ + [Empty vector]	*cyc3*Δ + [CYC3]	*cyc3*Δ + [CYC3HA]	*cyc3*Δ+ [preCYB2 -CYC3HA]	*cyc3*Δ+ [preCYB2 (A63P)- CYC3HA]
Glucose	O	O	O	O	O	O
Glycerol	O	X	O	O	O	X

Figure 34. Validation of TIM23 random mutagenesis screening. (A) Yeast strain (*cyc3*Δ) was transformed with an empty plasmid or a plasmid encoding CYC3, CYC3HA, preCyb2-CYC3HA or preCyb2(A63P)-CYC3HA. The yeast transformants were streaked on YPEG (non-fermentable carbon source) and incubated at 30℃ for two days. *cyc3*Δ did not grow on YPEG. The growth defect was rescued by expression of CYC3, CYC3HA and preCyb2-CYC3HA. The expression of preCyb2(A63P)-CYC3HA did not rescue the growth defect, indicating that the fusion protein was translocated to the matrix. (B) The summary of the growth assay.

CHAPTER IV – DISCUSSION

IV.1. Development of an ER topology reporter

IV.1.1. gGFP as a novel ER topology reporter of eukaryotic cells

Membrane proteins function as transporters or signaling receptors and carry out vital cellular processes[7]. Obtaining the structure of membrane proteins has always been important to understand the function of proteins and to design appropriate drugs that target them. Yet, most membrane proteins lack the structural information due to the experimental challenges in overexpression, crystallization[14].

In this thesis, gGFP that fluoresces in the cytosol but becomes glycosylated and non-fluorescent in the ER lumen, was developed as a eukaryotic topology reporter and used to validate $N_{out}C_{in}$ topology of URG7, $N_{out}C_{out}$ of MRP6 and $N_{in}C_{out}$ of SP-C[106,112]. gGFP is advantageous over many of the previously used topology reporters, such as FPP assay, as it allows rapid fluorescence based live-cell assessment of membrane protein topology without the need of cell fixation or lysis. In addition, the glycosylation status of gGFP not only serves as a back to back control assessment of membrane protein topology, but also rules out the possible misinterpretation of the fluorescence based assessment that arises from partial targeting of membrane proteins, proteins of mixed topology or low expression of a fusion protein. However, it should be noted that the use of gGFP is limited to the ER targeted membrane proteins as N-linked glycosylation is specific to the ER. The developed gGFP offers a new tool to facilitate the topology determination of ER membrane proteins *in vivo* and has potential to be used in global screening of ER membrane protein topology.

IV.2. Protein insertion into the ER membrane by the SEC61 complex

IV.2.1. The role of lateral gating helices of Sec61 in membrane protein insertion

Next, Sec61, the pore forming subunit of the SEC61 complex, was characterized by systematically replacing the residues lining TMS2 and 7 to an alanine residue. TMS2 and 7 of Sec61, termed as the lateral gating helices, opens laterally towards the membrane to allow partitioning of a TMS[63-65]. The mutagenesis screening

classified residues into two groups, the ones that enhance membrane protein insertion and the others that impair it[107]. The importance of the lateral gating helices in mediating membrane protein insertion were highlighted in the previous studies, where the pore ring residues and several residues of the lateral gating helices were reported to fine tune the threshold hydrophobicity for membrane protein insertion and influence topogenesis. Compared to the previous studies, this study provided the first extensive screening of the entire lateral gating helices. The fact that the several residues identified in this study, such as P292 and N302, overlap with the previously reported phenotypes in regulating membrane protein insertion ensures the sensitivity and reliability of the screening. In addition to the systematic grouping lateral gating residues into two groups, TMS2 affected the insertion of signal anchor proteins to a greater degree whereas TMS7 affected the insertion of multi-spanning membrane proteins more. These observation suggest a model where TMS2 plays a role in the initial opening of the channel and TMS7 governs the insertion of downstream TMS.

IV.2.2. A platform to study the SEC61 complex mediated membrane protein insertion in mammalian cell-lines

To explore the mammalian SEC61 complex, the targeting and topology profile of a set of model membrane proteins used in yeast system were examined. Here, the signal anchor proteins were targeted with 3 leucine, similar to yeast, but were cleaved more efficiently in mammalian cells. In addition, the hydrophobicity required for the insertion of 2nd TMS was also lowered in mammalian cells. The differences may be attributable to the subunits of the SEC61 complex specific to the mammalian cells such as TRAM or TRAP. This study provides a platform to monitor the model membrane protein insertion profile in the mammalian cells lacking a specific SEC61 complex subunit, thereby identifying the role of such subunits in mammalian cells. Understanding the mammalian SEC61 translocon complex is important for numerous reasons. For about 30% of proteins, they are initially targeted to the ER during the biogenesis *en route* to the secretory pathway and their translocation across or insertion into the ER membrane is mediated by the SEC61 translocon complex[34,35,40,41]. Moreover, failures in the proper assembly of the translocon complex is devastating. It leads to defects in protein secretion which

in turn raises several health issues including digestion problems and diabetes. In addition, mutations in the components of the translocon have been linked to several physiological diseases[139]. For instance, overexpression of Sec62 and Sec63 is observed in various cancer cells. Sec63 is also mutated in polycyctic liver diseases. One of the nucleotide exchange factors, Sil1, which assists BiP during protein import, is linked to neurodegeneration[140]. Further studies with a collection of translocon subunit knock-out cell lines are needed to reveal the roles of different subunits of SEC61 complex in targeting and topogenesis of membrane proteins.

IV.3. Protein insertion into the mitochondrial IM by the TIM23 complex

IV.3.1. Mgr2 may not act as a general gate keeper of the TIM23 complex

Lastly, the key factors that regulate the TIM23 complex mediated entry of mitochondrial IM proteins into the IM, with specific emphasis on the role of Mgr2 was examined. Mitochondrial IM is the protein richest membrane in cells and governs respiration[141]. Several metabolic disorders and neurodegenerative diseases have been linked to mitochondrial dysfunction[142]. The proper assembly of IM proteins is crucial for optimal mitochondrial health, yet, how IM proteins are recognized by TIM23 complex and inserted into the membrane remain unanswered. Recently, Mgr2 was assigned as a lateral gate keeper of the TIM23 complex as the sorting of Mgm1 and Cyb2 into the membrane was delayed upon overexpression of Mgr2 and facilitated upon deletion[104]. However, when the insertion of other IM proteins was tested in this thesis, their insertion was uninfluenced by Mgr2 levels. In addition, the hydrophobicity required for a TMS insertion into mitochondrial IM remained unchanged. This may be due to the low endogenous expression level of Mgr2 in cells. The fact that endogenous Mgr2 is barely detectable in cells suggests that not all of the TIM23 complex contain Mgr2, therefore for most mitochondrial proteins, Mgr2 may not be required. In addition, for Mgm1, the substrate of Mgr2, its 1st TMS is unique in a sense that it has a boarder lining features for its insertion into the IM with 50% insertion efficiency. This suggests that Mgr2 may only be required for such proteins where the protein insertion efficiency need to be tightly

regulated. In sum, it seems that Mgr2's role in gate keeping of TIM23 complex is limited to Mgm1 and Cyb2.

IV.3.2. The molecular mechanism of the TIM23 complex mediated protein import and insertion into the membrane

To further identify the mechanism of TIM23 mediated membrane protein insertion, a genetic screening was devised and validated to select for Tim23 mutants that is capable of inserting a TMS with a proline. A TMS with a proline is frequently found in conservatively sorted proteins and these TMS escape the recognition by the TIM23 complex, but how the TIM23 complex discriminates a TMS with a proline is unknown[103]. Thus, uncovering the domain of Tim23 that distinguishes the TMS of stop-transfer and conservative sorting is the first step towards understanding the TMS recognition by TIM23, hence mechanism of membrane protein insertion into the mitochondrial IM. In addition, information on speed and forces that act on a polypeptide during the import into mitochondria have not been quantitatively analyzed. As the quantitative analysis of the import process is difficult *in vivo*, an optical tweezer based *in vitro* protein import assay was designed. The kinetics and the forces will be recorded by tracking a single protein import into isolated mitochondria with different force loaded. This experimental setup is expandable to monitor the protein import speed upon various stress conditions or different presequence. The proposed studies will shed light on the molecular mechanism of mitochondrial protein import by providing a numerical data of kinetics and forces.

In sum, this thesis attempted to provide a novel tool to solve membrane protein topology and resolve the molecular mechanism of translocon mediated membrane protein insertion at the ER and mitochondrial IM. Follow-up studies will solve the previously unreported topology of membrane proteins, the domains responsible for the recognition of TMS by TIM23 complex.

REFERENCES

1 Mitaku, S., Ono, M., Hirokawa, T., Boon-Chieng, S. & Sonoyama, M. Proportion of membrane proteins in proteomes of 15 single-cell organisms analyzed by the SOSUI prediction system. *Biophys Chem* **82**, 165-171 (1999).

2 Wallin, E. & von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci* **7**, 1029-1038, doi:10.1002/pro.5560070420 (1998).

3 Weinglass, A. B., Whitelegge, J. P. & Kaback, H. R. Integrating mass spectrometry into membrane protein drug discovery. *Curr Opin Drug Discov Devel* **7**, 589-599 (2004).

4 Mayor, S. & Riezman, H. Sorting GPI-anchored proteins. *Nat Rev Mol Cell Biol* **5**, 110-120, doi:10.1038/nrm1309 (2004).

5 Johnson, J. E. & Cornell, R. B. Amphitropic proteins: regulation by reversible membrane interactions (review). *Mol Membr Biol* **16**, 217-235 (1999).

6 von Heijne, G. Principles of membrane protein assembly and structure. *Prog Biophys Mol Biol* **66**, 113-139 (1996).

7 Klabunde, T. & Hessler, G. Drug design strategies for targeting G-protein-coupled receptors. *Chembiochem* **3**, 928-944, doi:10.1002/1439-7633(20021004)3:10<928::AID-CBIC928>3.0.CO;2-5 (2002).

8 Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**, 567-580, doi:10.1006/jmbi.2000.4315 (2001).

9 Imai, K., Fujita, N., Gromiha, M. M. & Horton, P. Eukaryote-wide sequence analysis of mitochondrial beta-barrel outer membrane proteins. *BMC Genomics* **12**, 79, doi:10.1186/1471-2164-12-79 (2011).

10 von Heijne, G. Membrane-protein topology. *Nat Rev Mol Cell Biol* **7**, 909-918, doi:10.1038/nrm2063 (2006).

11 Ott, C. M. & Lingappa, V. R. Integral membrane protein biosynthesis: why topology is hard to predict. *J Cell Sci* **115**, 2003-2009 (2002).

12 von Heijne, G. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* **341**, 456-458, doi:10.1038/341456a0 (1989).

13 Heijne, G. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J* **5**, 3021-3027 (1986).

14 Lee, H. & Kim, H. Membrane topology of transmembrane proteins: determinants and experimental tools. *Biochem Biophys Res Commun* **453**, 268-276, doi:10.1016/j.bbrc.2014.05.111 (2014).

15 Sommer, N., Junne, T., Kalies, K. U., Spiess, M. & Hartmann, E. TRAP assists membrane protein topogenesis at the mammalian ER membrane. *Biochim Biophys Acta* **1833**, 3104-3111, doi:10.1016/j.bbamcr.2013.08.018 (2013).

16 Jung, S. J., Kim, J. E., Reithinger, J. H. & Kim, H. The Sec62-Sec63 translocon facilitates translocation of the C-terminus of membrane proteins.

J Cell Sci **127**, 4270-4278, doi:10.1242/jcs.153650 (2014).
17 Reithinger, J. H., Kim, J. E. & Kim, H. Sec62 protein mediates membrane insertion and orientation of moderately hydrophobic signal anchor proteins in the endoplasmic reticulum (ER). *J Biol Chem* **288**, 18058-18067, doi:10.1074/jbc.M113.473009 (2013).
18 Junne, T., Kocik, L. & Spiess, M. The hydrophobic core of the Sec61 translocon defines the hydrophobicity threshold for membrane integration. *Mol Biol Cell* **21**, 1662-1670, doi:10.1091/mbc.E10-01-0060 (2010).
19 Junne, T., Schwede, T., Goder, V. & Spiess, M. Mutations in the Sec61p channel affecting signal sequence recognition and membrane protein topology. *J Biol Chem* **282**, 33201-33209, doi:10.1074/jbc.M707219200 (2007).
20 Bogdanov, M., Heacock, P. N. & Dowhan, W. A polytopic membrane protein displays a reversible topology dependent on membrane lipid composition. *EMBO J* **21**, 2107-2116, doi:10.1093/emboj/21.9.2107 (2002).
21 Eusebio, A., Friedberg, T. & Spiess, M. The role of the hydrophobic domain in orienting natural signal sequences within the ER membrane. *Exp Cell Res* **241**, 181-185, doi:10.1006/excr.1998.4042 (1998).
22 Wahlberg, J. M. & Spiess, M. Multiple determinants direct the orientation of signal-anchor proteins: the topogenic role of the hydrophobic signal domain. *J Cell Biol* **137**, 555-562 (1997).
23 Goder, V. & Spiess, M. Molecular mechanism of signal sequence orientation in the endoplasmic reticulum. *EMBO J* **22**, 3645-3653, doi:10.1093/emboj/cdg361 (2003).
24 Green, N., Fang, H., Kalies, K. U. & Canfield, V. Determining the topology of an integral membrane protein. *Curr Protoc Cell Biol* **Chapter 5**, Unit 5 2, doi:10.1002/0471143030.cb0502s00 (2001).
25 Feilmeier, B. J., Iseminger, G., Schroeder, D., Webber, H. & Phillips, G. J. Green fluorescent protein functions as a reporter for protein localization in Escherichia coli. *J Bacteriol* **182**, 4068-4076 (2000).
26 Manoil, C. & Beckwith, J. A genetic approach to analyzing membrane protein topology. *Science* **233**, 1403-1408 (1986).
27 Chang, X. B., Hou, Y. X., Jensen, T. J. & Riordan, J. R. Mapping of cystic fibrosis transmembrane conductance regulator membrane topology by glycosylation site insertion. *J Biol Chem* **269**, 18572-18575 (1994).
28 Laudon, H. *et al.* A nine-transmembrane domain topology for presenilin 1. *J Biol Chem* **280**, 35352-35360, doi:10.1074/jbc.M507217200 (2005).
29 Dempski, R. E., Jr. & Imperiali, B. Oligosaccharyl transferase: gatekeeper to the secretory pathway. *Curr Opin Chem Biol* **6**, 844-850 (2002).
30 Kornfeld, R. & Kornfeld, S. Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem* **54**, 631-664, doi:10.1146/annurev.bi.54.070185.003215 (1985).
31 Lorenz, H., Hailey, D. W. & Lippincott-Schwartz, J. Fluorescence protease protection of GFP chimeras to reveal protein topology and subcellular localization. *Nat Methods* **3**, 205-210, doi:10.1038/nmeth857 (2006).
32 Lorenz, H., Hailey, D. W. & Lippincott-Schwartz, J. Addressing membrane protein topology using the fluorescence protease protection (FPP) assay. *Methods Mol Biol* **440**, 227-233, doi:10.1007/978-1-59745-178-9_17

(2008).
33 Lorenz, H., Hailey, D. W., Wunder, C. & Lippincott-Schwartz, J. The fluorescence protease protection (FPP) assay to determine protein localization and membrane topology. *Nat Protoc* **1**, 276-279, doi:10.1038/nprot.2006.42 (2006).
34 Blobel, G. & Dobberstein, B. Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J Cell Biol* **67**, 835-851 (1975).
35 Blobel, G. & Dobberstein, B. Transfer of proteins across membranes. II. Reconstitution of functional rough microsomes from heterologous components. *J Cell Biol* **67**, 852-862 (1975).
36 Wild, K., Halic, M., Sinning, I. & Beckmann, R. SRP meets the ribosome. *Nat Struct Mol Biol* **11**, 1049-1053, doi:10.1038/nsmb853 (2004).
37 Nyathi, Y., Wilkinson, B. M. & Pool, M. R. Co-translational targeting and translocation of proteins to the endoplasmic reticulum. *Biochim Biophys Acta* **1833**, 2392-2402, doi:10.1016/j.bbamcr.2013.02.021 (2013).
38 Park, E. & Rapoport, T. A. Mechanisms of Sec61/SecY-mediated protein translocation across membranes. *Annu Rev Biophys* **41**, 21-40, doi:10.1146/annurev-biophys-050511-102312 (2012).
39 Plath, K. & Rapoport, T. A. Spontaneous release of cytosolic proteins from posttranslational substrates before their transport into the endoplasmic reticulum. *J Cell Biol* **151**, 167-178 (2000).
40 Van den Berg, B. *et al.* X-ray structure of a protein-conducting channel. *Nature* **427**, 36-44, doi:10.1038/nature02218 (2004).
41 Mothes, W., Prehn, S. & Rapoport, T. A. Systematic probing of the environment of a translocating secretory protein during translocation through the ER membrane. *EMBO J* **13**, 3973-3982 (1994).
42 Harada, Y., Li, H., Wall, J. S., Li, H. & Lennarz, W. J. Structural studies and the assembly of the heptameric post-translational translocon complex. *J Biol Chem* **286**, 2956-2965, doi:10.1074/jbc.M110.159517 (2011).
43 Deshaies, R. J., Sanders, S. L., Feldheim, D. A. & Schekman, R. Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex. *Nature* **349**, 806-808, doi:10.1038/349806a0 (1991).
44 Stirling, C. J., Rothblatt, J., Hosobuchi, M., Deshaies, R. & Schekman, R. Protein translocation mutants defective in the insertion of integral membrane proteins into the endoplasmic reticulum. *Mol Biol Cell* **3**, 129-142 (1992).
45 Panzner, S., Dreier, L., Hartmann, E., Kostka, S. & Rapoport, T. A. Posttranslational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p. *Cell* **81**, 561-570 (1995).
46 Lyman, S. K. & Schekman, R. Binding of secretory precursor polypeptides to a translocon subcomplex is regulated by BiP. *Cell* **88**, 85-96 (1997).
47 Beckmann, R. *et al.* Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex. *Science* **278**, 2123-2126 (1997).
48 Becker, T. *et al.* Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science* **326**, 1369-1373, doi:10.1126/science.1178535 (2009).

49 Deshaies, R. J. & Schekman, R. A yeast mutant defective at an early stage in import of secretory protein precursors into the endoplasmic reticulum. *J Cell Biol* **105**, 633-645 (1987).
50 Rothblatt, J. A., Deshaies, R. J., Sanders, S. L., Daum, G. & Schekman, R. Multiple genes are required for proper insertion of secretory proteins into the endoplasmic reticulum in yeast. *J Cell Biol* **109**, 2641-2652 (1989).
51 Lakkaraju, A. K. *et al.* Efficient secretion of small proteins in mammalian cells relies on Sec62-dependent posttranslational translocation. *Mol Biol Cell* **23**, 2712-2722, doi:10.1091/mbc.E12-03-0228 (2012).
52 Meyer, H. A. *et al.* Mammalian Sec61 is associated with Sec62 and Sec63. *J Biol Chem* **275**, 14550-14557 (2000).
53 Tyedmers, J. *et al.* Homologs of the yeast Sec complex subunits Sec62p and Sec63p are abundant proteins in dog pancreas microsomes. *Proc Natl Acad Sci U S A* **97**, 7214-7219 (2000).
54 Lang, S. *et al.* Different effects of Sec61alpha, Sec62 and Sec63 depletion on transport of polypeptides into the endoplasmic reticulum of mammalian cells. *J Cell Sci* **125**, 1958-1969, doi:10.1242/jcs.096727 (2012).
55 Mades, A. *et al.* Role of human sec63 in modulating the steady-state levels of multi-spanning membrane proteins. *PLoS One* **7**, e49243, doi:10.1371/journal.pone.0049243 (2012).
56 Kroczynska, B., Evangelista, C. M., Samant, S. S., Elguindi, E. C. & Blond, S. Y. The SANT2 domain of the murine tumor cell DnaJ-like protein 1 human homologue interacts with alpha1-antichymotrypsin and kinetically interferes with its serpin inhibitory activity. *J Biol Chem* **279**, 11432-11443, doi:10.1074/jbc.M310903200 (2004).
57 Fons, R. D., Bogert, B. A. & Hegde, R. S. Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. *J Cell Biol* **160**, 529-539, doi:10.1083/jcb.200210095 (2003).
58 Gorlich, D., Hartmann, E., Prehn, S. & Rapoport, T. A. A protein of the endoplasmic reticulum involved early in polypeptide translocation. *Nature* **357**, 47-52, doi:10.1038/357047a0 (1992).
59 Gorlich, D. & Rapoport, T. A. Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. *Cell* **75**, 615-630 (1993).
60 Voigt, S., Jungnickel, B., Hartmann, E. & Rapoport, T. A. 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 (1996).
61 Hegde, R. S., Voigt, S., Rapoport, T. A. & Lingappa, V. R. TRAM regulates the exposure of nascent secretory proteins to the cytosol during translocation into the endoplasmic reticulum. *Cell* **92**, 621-631 (1998).
62 Sauri, A., McCormick, P. J., Johnson, A. E. & Mingarro, I. Sec61alpha and TRAM are sequentially adjacent to a nascent viral membrane protein during its ER integration. *J Mol Biol* **366**, 366-374, doi:10.1016/j.jmb.2006.11.052 (2007).
63 Heinrich, S. U., Mothes, W., Brunner, J. & Rapoport, T. A. The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell* **102**, 233-244 (2000).

64 Bowie, J. U. Solving the membrane protein folding problem. *Nature* **438**, 581-589, doi:10.1038/nature04395 (2005).
65 Trueman, S. F., Mandon, E. C. & Gilmore, R. A gating motif in the translocation channel sets the hydrophobicity threshold for signal sequence function. *J Cell Biol* **199**, 907-918, doi:10.1083/jcb.201207163 (2012).
66 von Heijne, G. Membrane proteins: from sequence to structure. *Annu Rev Biophys Biomol Struct* **23**, 167-192, doi:10.1146/annurev.bb.23.060194.001123 (1994).
67 White, S. H. & von Heijne, G. How translocons select transmembrane helices. *Annu Rev Biophys* **37**, 23-42, doi:10.1146/annurev.biophys.37.032807.125904 (2008).
68 von Heijne, G. The membrane protein universe: what's out there and why bother? *J Intern Med* **261**, 543-557, doi:10.1111/j.1365-2796.2007.01792.x (2007).
69 Lundin, C., Kim, H., Nilsson, I., White, S. H. & von Heijne, G. Molecular code for protein insertion in the endoplasmic reticulum membrane is similar for N(in)-C(out) and N(out)-C(in) transmembrane helices. *Proc Natl Acad Sci U S A* **105**, 15702-15707, doi:10.1073/pnas.0804842105 (2008).
70 Hessa, T. *et al.* Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377-381, doi:10.1038/nature03216 (2005).
71 Hessa, T. *et al.* Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026-1030, doi:10.1038/nature06387 (2007).
72 Hessa, T., Reithinger, J. H., von Heijne, G. & Kim, H. Analysis of transmembrane helix integration in the endoplasmic reticulum in S. cerevisiae. *J Mol Biol* **386**, 1222-1228 (2009).
73 Neupert, W. Protein import into mitochondria. *Annu Rev Biochem* **66**, 863-917, doi:10.1146/annurev.biochem.66.1.863 (1997).
74 Verner, K. Co-translational protein import into mitochondria: an alternative view. *Trends Biochem Sci* **18**, 366-371 (1993).
75 Lesnik, C., Cohen, Y., Atir-Lande, A., Schuldiner, M. & Arava, Y. OM14 is a mitochondrial receptor for cytosolic ribosomes that supports co-translational import into mitochondria. *Nat Commun* **5**, 5711, doi:10.1038/ncomms6711 (2014).
76 Williams, C. C., Jan, C. H. & Weissman, J. S. Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. *Science* **346**, 748-751, doi:10.1126/science.1257522 (2014).
77 Becker, T., Bottinger, L. & Pfanner, N. Mitochondrial protein import: from transport pathways to an integrated network. *Trends Biochem Sci* **37**, 85-91, doi:10.1016/j.tibs.2011.11.004 (2012).
78 Walther, D. M. & Rapaport, D. Biogenesis of mitochondrial outer membrane proteins. *Biochim Biophys Acta* **1793**, 42-51, doi:10.1016/j.bbamcr.2008.04.013 (2009).
79 Stojanovski, D., Bragoszewski, P. & Chacinska, A. The MIA pathway: a tight bond between protein transport and oxidative folding in mitochondria. *Biochim Biophys Acta* **1823**, 1142-1150, doi:10.1016/j.bbamcr.2012.04.014 (2012).
80 Bragoszewski, P. *et al.* Retro-translocation of mitochondrial

intermembrane space proteins. *Proc Natl Acad Sci U S A* **112**, 7713-7718, doi:10.1073/pnas.1504615112 (2015).
81 Schmidt, O., Pfanner, N. & Meisinger, C. Mitochondrial protein import: from proteomics to functional mechanisms. *Nat Rev Mol Cell Biol* **11**, 655-667, doi:10.1038/nrm2959 (2010).
82 Neupert, W. A perspective on transport of proteins into mitochondria: a myriad of open questions. *J Mol Biol* **427**, 1135-1158, doi:10.1016/j.jmb.2015.02.001 (2015).
83 Mokranjac, D. & Neupert, W. The many faces of the mitochondrial TIM23 complex. *Biochim Biophys Acta* **1797**, 1045-1054, doi:10.1016/j.bbabio.2010.01.026 (2010).
84 Milisav, I., Moro, F., Neupert, W. & Brunner, M. Modular structure of the TIM23 preprotein translocase of mitochondria. *J Biol Chem* **276**, 25856-25861, doi:10.1074/jbc.M102132200 (2001).
85 Alder, N. N., Jensen, R. E. & Johnson, A. E. Fluorescence mapping of mitochondrial TIM23 complex reveals a water-facing, substrate-interacting helix surface. *Cell* **134**, 439-450, doi:10.1016/j.cell.2008.06.007 (2008).
86 Truscott, K. N. *et al.* A presequence- and voltage-sensitive channel of the mitochondrial preprotein translocase formed by Tim23. *Nat Struct Biol* **8**, 1074-1082, doi:10.1038/nsb726 (2001).
87 Kubrich, M. *et al.* The polytopic mitochondrial inner membrane proteins MIM17 and MIM23 operate at the same preprotein import site. *FEBS Lett* **349**, 222-228 (1994).
88 Berthold, J. *et al.* The MIM complex mediates preprotein translocation across the mitochondrial inner membrane and couples it to the mt-Hsp70/ATP driving system. *Cell* **81**, 1085-1093 (1995).
89 Emtage, J. L. & Jensen, R. E. MAS6 encodes an essential inner membrane component of the yeast mitochondrial protein import pathway. *J Cell Biol* **122**, 1003-1012 (1993).
90 Ryan, K. R., Menold, M. M., Garrett, S. & Jensen, R. E. SMS1, a high-copy suppressor of the yeast mas6 mutant, encodes an essential inner membrane protein required for mitochondrial protein import. *Mol Biol Cell* **5**, 529-538 (1994).
91 Yamamoto, H. *et al.* Tim50 is a subunit of the TIM23 complex that links protein translocation across the outer and inner mitochondrial membranes. *Cell* **111**, 519-528 (2002).
92 Mokranjac, D. *et al.* Tim50, a novel component of the TIM23 preprotein translocase of mitochondria. *EMBO J* **22**, 816-825, doi:10.1093/emboj/cdg090 (2003).
93 Mokranjac, D. *et al.* Role of Tim50 in the transfer of precursor proteins from the outer to the inner membrane of mitochondria. *Mol Biol Cell* **20**, 1400-1407, doi:10.1091/mbc.E08-09-0934 (2009).
94 Tamura, Y. *et al.* Tim23-Tim50 pair coordinates functions of translocators and motor proteins in mitochondrial protein import. *J Cell Biol* **184**, 129-141, doi:10.1083/jcb.200808068 (2009).
95 Gebert, M. *et al.* Mgr2 promotes coupling of the mitochondrial presequence translocase to partner complexes. *J Cell Biol* **197**, 595-604, doi:10.1083/jcb.201110047 (2012).
96 Schiller, D., Cheng, Y. C., Liu, Q., Walter, W. & Craig, E. A. Residues of

Tim44 involved in both association with the translocon of the inner mitochondrial membrane and regulation of mitochondrial Hsp70 tethering. *Mol Cell Biol* **28**, 4424-4433, doi:10.1128/MCB.00007-08 (2008).

97 Hutu, D. P. *et al.* Mitochondrial protein import motor: differential role of Tim44 in the recruitment of Pam17 and J-complex to the presequence translocase. *Mol Biol Cell* **19**, 2642-2649, doi:10.1091/mbc.E07-12-1226 (2008).

98 Schiller, D. Pam17 and Tim44 act sequentially in protein import into the mitochondrial matrix. *Int J Biochem Cell Biol* **41**, 2343-2349, doi:10.1016/j.biocel.2009.06.011 (2009).

99 Park, K., Jung, S. J., Kim, H. & Kim, H. Mode of membrane insertion of individual transmembrane segments in Mdl1 and Mdl2, multi-spanning mitochondrial ABC transporters. *FEBS Lett* **588**, 3445-3453, doi:10.1016/j.febslet.2014.08.001 (2014).

100 Bohnert, M. *et al.* Cooperation of stop-transfer and conservative sorting mechanisms in mitochondrial protein transport. *Curr Biol* **20**, 1227-1232, doi:10.1016/j.cub.2010.05.058 (2010).

101 Botelho, S. C. *et al.* TIM23-mediated insertion of transmembrane alpha-helices into the mitochondrial inner membrane. *EMBO J* **30**, 1003-1011, doi:10.1038/emboj.2011.29 (2011).

102 Botelho, S. C., Tatsuta, T., von Heijne, G. & Kim, H. Dislocation by the m-AAA protease increases the threshold hydrophobicity for retention of transmembrane helices in the inner membrane of yeast mitochondria. *J Biol Chem* **288**, 4792-4798, doi:10.1074/jbc.M112.430892 (2013).

103 Meier, S., Neupert, W. & Herrmann, J. M. Proline residues of transmembrane domains determine the sorting of inner membrane proteins in mitochondria. *J Cell Biol* **170**, 881-888, doi:10.1083/jcb.200505126 (2005).

104 Ieva, R. *et al.* Mgr2 functions as lateral gatekeeper for preprotein sorting in the mitochondrial inner membrane. *Mol Cell* **56**, 641-652, doi:10.1016/j.molcel.2014.10.010 (2014).

105 Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686-691, doi:10.1038/nature02026 (2003).

106 Lee, H., Min, J., von Heijne, G. & Kim, H. Glycosylatable GFP as a compartment-specific membrane topology reporter. *Biochem Biophys Res Commun* **427**, 780-784, doi:10.1016/j.bbrc.2012.09.138 (2012).

107 Reithinger, J. H., Yim, C., Kim, S., Lee, H. & Kim, H. Structural and functional profiling of the lateral gate of the Sec61 translocon. *J Biol Chem* **289**, 15845-15855, doi:10.1074/jbc.M113.533794 (2014).

108 Ojemalm, K., Halling, K. K., Nilsson, I. & von Heijne, G. Orientational preferences of neighboring helices can drive ER insertion of a marginally hydrophobic transmembrane helix. *Mol Cell* **45**, 529-540, doi:10.1016/j.molcel.2011.12.024 (2012).

109 Heinrich, S. U. & Rapoport, T. A. Cooperation of transmembrane segments during the integration of a double-spanning protein into the ER membrane. *EMBO J* **22**, 3654-3663, doi:10.1093/emboj/cdg346 (2003).

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

111 Kocik, L., Junne, T. & Spiess, M. Orientation of internal signal-anchor sequences at the Sec61 translocon. *J Mol Biol* **424**, 368-378, doi:10.1016/j.jmb.2012.10.010 (2012).
112 Lee, H. *et al.* Live-cell topology assessment of URG7, MRP6(1)(0)(2) and SP-C using glycosylatable green fluorescent protein in mammalian cells. *Biochem Biophys Res Commun* **450**, 1587-1592, doi:10.1016/j.bbrc.2014.07.046 (2014).
113 Herlan, M., Bornhovd, C., Hell, K., Neupert, W. & Reichert, A. S. Alternative topogenesis of Mgm1 and mitochondrial morphology depend on ATP and a functional import motor. *J Cell Biol* **165**, 167-173, doi:10.1083/jcb.200403022 (2004).
114 Park, K., Botelho, S. C., Hong, J., Osterberg, M. & Kim, H. Dissecting stop transfer versus conservative sorting pathways for mitochondrial inner membrane proteins in vivo. *J Biol Chem* **288**, 1521-1532, doi:10.1074/jbc.M112.409748 (2013).
115 Osterberg, M., Calado Botelho, S., von Heijne, G. & Kim, H. Charged flanking residues control the efficiency of membrane insertion of the first transmembrane segment in yeast mitochondrial Mgm1p. *FEBS Lett* **585**, 1238-1242, doi:10.1016/j.febslet.2011.03.056 (2011).
116 Beasley, E. M., Muller, S. & Schatz, G. The signal that sorts yeast cytochrome b2 to the mitochondrial intermembrane space contains three distinct functional regions. *EMBO J* **12**, 2303-2311 (1993).
117 Dumont, M. E., Ernst, J. F., Hampsey, D. M. & Sherman, F. Identification and sequence of the gene encoding cytochrome c heme lyase in the yeast Saccharomyces cerevisiae. *EMBO J* **6**, 235-241 (1987).
118 Bernard, D. G., Gabilly, S. T., Dujardin, G., Merchant, S. & Hamel, P. P. Overlapping specificities of the mitochondrial cytochrome c and c1 heme lyases. *J Biol Chem* **278**, 49732-49742, doi:10.1074/jbc.M308881200 (2003).
119 Moffitt, J. R., Chemla, Y. R., Smith, S. B. & Bustamante, C. Recent advances in optical tweezers. *Annu Rev Biochem* **77**, 205-228, doi:10.1146/annurev.biochem.77.043007.090225 (2008).
120 Maillard, R. A. *et al.* ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell* **145**, 459-469, doi:10.1016/j.cell.2011.04.010 (2011).
121 Meineke, B. *et al.* The outer membrane form of the mitochondrial protein Mcr1 follows a TOM-independent membrane insertion pathway. *FEBS Lett* **582**, 855-860, doi:10.1016/j.febslet.2008.02.009 (2008).
122 Beckett, D., Kovaleva, E. & Schatz, P. J. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci* **8**, 921-929, doi:10.1110/ps.8.4.921 (1999).
123 van Werven, F. J. & Timmers, H. T. The use of biotin tagging in Saccharomyces cerevisiae improves the sensitivity of chromatin immunoprecipitation. *Nucleic Acids Res* **34**, e33, doi:10.1093/nar/gkl003 (2006).
124 Stagge, F., Mitronova, G. Y., Belov, V. N., Wurm, C. A. & Jakobs, S. SNAP-, CLIP- and Halo-tag labelling of budding yeast cells. *PLoS One* **8**, e78745, doi:10.1371/journal.pone.0078745 (2013).
125 Yin, J. *et al.* Genetically encoded short peptide tag for versatile protein

labeling by Sfp phosphopantetheinyl transferase. *Proc Natl Acad Sci U S A* **102**, 15815-15820, doi:10.1073/pnas.0507705102 (2005).
126 Goder, V., Junne, T. & Spiess, M. Sec61p contributes to signal sequence orientation according to the positive-inside rule. *Mol Biol Cell* **15**, 1470-1478, doi:10.1091/mbc.E03-08-0599 (2004).
127 Pilon, M., Schekman, R. & Romisch, K. Sec61p mediates export of a misfolded secretory protein from the endoplasmic reticulum to the cytosol for degradation. *EMBO J* **16**, 4540-4548, doi:10.1093/emboj/16.15.4540 (1997).
128 Kim, H., Melen, K. & von Heijne, G. Topology models for 37 Saccharomyces cerevisiae membrane proteins based on C-terminal reporter fusions and predictions. *J Biol Chem* **278**, 10208-10213, doi:10.1074/jbc.M300163200 (2003).
129 Weckbecker, D. & Herrmann, J. M. Methods to study the biogenesis of membrane proteins in yeast mitochondria. *Methods Mol Biol* **1033**, 307-322, doi:10.1007/978-1-62703-487-6_20 (2013).
130 Yang, F., Moss, L. G. & Phillips, G. N., Jr. The molecular structure of green fluorescent protein. *Nat Biotechnol* **14**, 1246-1251, doi:10.1038/nbt1096-1246 (1996).
131 Cormack, B. P. *et al.* Yeast-enhanced green fluorescent protein (yEGFP): a reporter of gene expression in Candida albicans. *Microbiology* **143 (Pt 2)**, 303-311, doi:10.1099/00221287-143-2-303 (1997).
132 Suzuki, N. *et al.* Crystallization of small proteins assisted by green fluorescent protein. *Acta Crystallogr D Biol Crystallogr* **66**, 1059-1066, doi:10.1107/S0907444910032944 (2010).
133 Wolfe, P. B., Wickner, W. & Goodman, J. M. Sequence of the leader peptidase gene of Escherichia coli and the orientation of leader peptidase in the bacterial envelope. *J Biol Chem* **258**, 12073-12080 (1983).
134 Bergh, M. L., Cepko, C. L., Wolf, D. & Robbins, P. W. Expression of the Saccharomyces cerevisiae glycoprotein invertase in mouse fibroblasts: glycosylation, secretion, and enzymatic activity. *Proc Natl Acad Sci U S A* **84**, 3570-3574 (1987).
135 Munro, S. & Pelham, H. R. A C-terminal signal prevents secretion of luminal ER proteins. *Cell* **48**, 899-907 (1987).
136 Losfeld, M. E., Soncin, F., Ng, B. G., Singec, I. & Freeze, H. H. A sensitive green fluorescent protein biomarker of N-glycosylation site occupancy. *FASEB J* **26**, 4210-4217, doi:10.1096/fj.12-211656 (2012).
137 Royant, A. & Noirclerc-Savoye, M. Stabilizing role of glutamic acid 222 in the structure of Enhanced Green Fluorescent Protein. *J Struct Biol* **174**, 385-390, doi:10.1016/j.jsb.2011.02.004 (2011).
138 Schmidt, O. *et al.* Regulation of mitochondrial protein import by cytosolic kinases. *Cell* **144**, 227-239, doi:10.1016/j.cell.2010.12.015 (2011).
139 Hassdenteufel, S., Klein, M. C., Melnyk, A. & Zimmermann, R. Protein transport into the human ER and related diseases, Sec61-channelopathies. *Biochem Cell Biol* **92**, 499-509, doi:10.1139/bcb-2014-0043 (2014).
140 Zimmermann, R., Muller, L. & Wullich, B. Protein transport into the endoplasmic reticulum: mechanisms and pathologies. *Trends Mol Med* **12**, 567-573, doi:10.1016/j.molmed.2006.10.004 (2006).
141 Becker, T., Gebert, M., Pfanner, N. & van der Laan, M. Biogenesis of

mitochondrial membrane proteins. *Curr Opin Cell Biol* **21**, 484-493, doi:10.1016/j.ceb.2009.04.002 (2009).

142 Hroudova, J., Singh, N. & Fisar, Z. Mitochondrial dysfunctions in neurodegenerative diseases: relevance to Alzheimer's disease. *Biomed Res Int* **2014**, 175062, doi:10.1155/2014/175062 (2014).

국문초록

진핵세포에서 translocon을 통한 막단백질의 삽입과정 연구

이헌상

서울대학교 자연과학대학 생명과학부

막단백질은 전체 단백질 중 약 30% 정도를 차지하고 있는 단백질로서 제약분야의 주요 표적이 되고, 신호전달과 물질수송 등의 다양한 역할을 수행한다. 따라서 막단백질의 구조를 밝히는 것은 막단백질의 역할에 대한 연구와 신약 개발에 있어 가장 중요한 부분이지만 대다수의 막단백질 구조는 아직 규명되지 못하였다. 이는 막단백질 구조를 밝히는 데 필요한 다양한 실험 기법의 부재와 막단백질 결정화 과정에서의 어려움에서 기인한다.

여러 생물정보학 프로그램들을 통해 막단백질의 2차구조인 위상을 예측할 수 있으나, 이는 결국 세포 내에서 실험적으로 증명되어야 한다. 그러나 그 동안 살아있는 진핵 세포 내에서의 단백질 위상 구조 결정은 실험 기법의 부재로 인해 불가능하였다. 이 논문에서는 새로운 위상 구조 결정 기법으로 효모와 동물 세포에 적용 가능한 글리코실화 녹색 형광 단백질 (glycosylatable green fluorescent protein, gGFP)을 개발하였다. gGFP는 형광단 근처에 몇 개의 글리코실화 자리를 가지고 있어 세포질에서는 글리코실화되지 않고 형광을 발하지만 소포체 내에서는 글리코실화되고 형광이 사라진다. 그러므로 글리코실화 유무에 따른 형광을 추적하여 gGFP의 세포 내 위치를 직접적으로 관찰할 수 있다. 먼저 기존에 2차구조가 알려진 막단백질의 카르복시 말단에 gGFP를 부착하여 위상 구조 결정 기법으로서의 gGFP의 가용성을

확인하였다. 나아가 인간 질병과 연관된 3가지의 막단백질, Urg7, Mrp6, SP-C의 위상을 결정하였다. 추가적으로 소수의 SP-C는 잘못된 방향으로 삽입이 되지만 잘못 삽입된 SP-C는 즉시 프로테아좀에 의해 분해된다는 것을 관찰하였다. 결론적으로 gGFP는 빠르고 정확하게 막단백질 위상을 살아있는 세포 내에서 관찰할 수 있는 새로운 기법이다.

막단백질의 위상은 생성단계에서 결정된다. 막단백질은 막에 존재하는 단백질 전달 통로 복합체를 통하여 막에 삽입된다. 단백질 전달 통로 복합체의 아단위들은 잘 알려져 있지만, 이것들이 어떻게 막관통 분절을 인식하여 특정 위상으로 삽입시키는지는 아직 명확히 밝혀지지 않았다. 그리하여 이 논문에서는 소포체와 미토콘드리아 내막에서의 단백질 전달 통로 복합체에 의한 막단백질 삽입과정 기작을 연구하였다.

소포체 막으로의 막단백질 삽입은 SEC61 복합체에 의해 이루어진다. 막단백질 삽입을 위해 Sec61은 두 번째와 일곱 번째 막관통 분절 사이(측면 통로)로 열린다. Sec61 변이들을 이용한 이전 연구들은 막단백질 삽입 과정이 단순한 열역학적 현상이 아닌 Sec61에 의한 능동적인 삽입 과정임을 의미했다. Sec61 측면 통로가 어떻게 열리고 닫히는지에 대한 자세한 기작 연구를 위해 Sec61의 두 번째와 일곱 번째 막관통 분절의 아미노산을 하나씩 변이시킨 후 효모 내에서 모델 단백질의 위상 변화를 관찰하였다. 이 연구 결과를 바탕으로 Sec61 변이들을 열린 입체구조를 가지는 집단과 닫힌 입체구조를 가지는 집단으로 구분하였는데, 단일 막관통 단백질의 삽입은 두 번째 막관통 분절이 변이된 효모들에서 더 영향을 받는 반면 다중 막관통 단백질의 경우에는 일곱 번째 막관통분절이 변이된 효모들에서 더 영향을 받는다는 것을 관찰하였다. 종합적으로 이 연구는 소포체로의 막단백질 삽입과 위상 결정에 관여하는 Sec61 측면 통로의 핵심 아미노산들을 밝혀내었다.

동물 세포의 SEC61 복합체에는 효모에는 없는 추가적인 아단위들이 존재한다. 이 논문에서는 SEC61에 대한 연구를 동물 세포로 확장시키기 위하여, 또 각각의 아단위들의 역할 분석을 위하여 효모에서 쓰였던 단일 막관통 단백질과 다중 막관통 단백질들을 HEK-293T와 HeLa 세포에 발현시켰다. 복합 막 관통 단백질의 경우, 막 삽입을 위한 역치 소수성이 효모에 비해 동물세포에서 더 낮음을 관찰했다. 추가적으로 중간적인 소수성을 띤 단일 막관통 단백질의 경우 동물 세포에서 잘린 형태가 증가하였다. 이러한 동물 세포에서 발생하는 소포체로의 막단백질 삽입 과정을 완전히 이해하기 위해서는 각각의 SEC61 복합체의 아단위를 없애거나 변이시킨 후 단백질 삽입 과정이 어떻게 변하는지를 보는 후속 연구가 필요하다.

미토콘드리아 내막으로의 단백질 삽입은 주로 TIM23 복합체에 의해서 이루어 진다. SEC61 복합체의 경우와 달리 TIM23 복합체는 알려진 측면 통로가 없었으나 최근 TIM23 복합체의 아단위인 Mgr2의 양에 따라 내막 단백질인 Mgm1과 Cyb2의 막 삽입과정이 지연 또는 촉진된다는 것이 밝혀졌다. 이로 인해 Mgr2는 TIM23 복합체의 측면 통로 개폐자로 명명되었다. 이 논문에서는 Mgr2가 다른 막단백질의 삽입 과정에도 영향을 미치는지, 또 그렇다면 Mgr2가 미토콘드리아 내막으로의 막단백질 삽입에 필요한 소수성 역치를 결정하는지를 연구하였다. Mgr2의 발현량은 미토콘드리아 내막으로의 단백질 삽입과이에 필요한 소수성에 전혀 영향을 주지 못함에 따라 Mgr2의 측면 통로 개폐자로서의 역할은 Mgm1와 Cyb2와 같은 단백질에 특정되어 있음을 증명하였다.

TIM23 복합체를 통해 삽입되는 막단백질들은 크게 종결 이동 경로 또는 보존 분류 경로를 따른다. 종결 이동 경로 단백질들은 TIM23 복합체에 의해 직접적으로 막으로 삽입이 되는 반면, 보존 분류 경로 단백질들은 기질로 먼저 수송된 후 내막으로 삽입된다. 생물정보학적 분석은 보존 분류 경로의 단백질의 막관통 분절에서 프롤린 아미노산이

흔하게 발견된다는 것을 밝혀내었다. 이 결과를 바탕으로 본 논문에서는 Tim23가 어떻게 막 관통분절의 프롤린 아미노산을 인식하고 기질로 분류하는지를 연구하기 위하여 계획적 유전자 변이 검사와 무작위 유전자 변이 검사를 설계하였다. 이는 막관통 분절의 프롤린 아미노산에 대한 저항성을 가진 Tim23 변이를 찾아내기 위한 것이다. 추가적으로 미토콘드리아로의 단백질 수송 속도와 그 힘을 측정할 수 있는 광 집게 분석을 설계하였다. 이 논문에서 설계된 연구들은 미토콘드리아에서의 단백질 수송과 내막으로 삽입되는 과정에 대한 분자적 단위의 이해를 도울 것이다.

주요 단어: 막단백질, 위상, 녹색 형광 단백질, 글리코실화, 소포체, 미토콘드리아, 단백질 전달 통로 복합체.

학번: 2010-20330