Master's Thesis of Science

Analysis of the Signal Sequence Engagement at the Translocon in *Saccharomyces cerevisiae*

출아효모에서 표적화 경로에 따른 신호서열-전좌효소 맞물림 기작의 차이 분석

February 2020

Graduate School of Biological Science
Seoul National University
Membrane Biology Major

Seoyoon Dymphna Jeong

Analysis of the Signal Sequence Engagement at the Translocon in *Saccharomyces cerevisiae*

by

Seoyoon Dymphna Jeong

Under the supervision of

Professor Hyun Ah Kim, Ph.D.

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

February 2020

School of Biological Science
Seoul National University

Abstract

In yeast, 30% of proteome is targeted to the endoplasmic reticulum (ER) as the first stage secretory pathway. The translocation into ER could be separated into three steps: delivery from the cytosol to ER surface, an engagement of substrate and translocon, and after initiation of translocation.

This study focused on the early stage docking process. To explain the dynamics of translocation, the head-in and inversion model and looped conformation model have been suggested. Thus, the purpose of this study is examining the suitability of models. For this, I classified the signal sequences depended on which model signals sequences followed. The test substrates were derived from CPY from modifying the length of N-region and the hydrophobicity of signal sequence.

I sought that Sec62 dependent and SRP independent signal sequences were inhibited their translocation by long N-region, and the positive charge rescued translocation of them. The translocation of SRP dependent signal sequences was originally not affected by the length of N-region. However, when the positive charge of N-region was eliminated, translocation of SRP dependent signal sequences shown Sec62 and Sec71/72 dependence and were affected by N-region length.

Therefore, these were suggested that 1) post-translocational substrate follows the head-in and inversion model, which is easily affected by the length of N-region, and 2) co-translocational substrate follows two models alternatively pursuant to charge distribution around the signal sequence.

Furthermore, the effects of N-region length and charge distribution were generally occurred at the other natural signal sequences, not only for CPY and its derivatives.

Keyword : Saccharomyces cerevisiae, secretory pathway, signal sequence, translocation, Sec71, Sec72

Table of Contents

LIST OF FIGURES AND TABLES

INTRODUCTION

I.1 Study Background

I.1.1 Protein translocation across the secretory pathway

Eukaryotic cells have spatially and functionally divided membrane structures within, such as the endoplasmic reticulum (ER), the Golgi apparatus, the mitochondria, the peroxisome, the nucleus, and the chloroplast, especially of plants. Thus, the proper distribution of the intracellular materials is especially crucial in eukaryotic cells. This is called 'cell logistics'.

Eukaryotic cytoplasm is 'inside,' and the counter space across the membrane is 'outside' including lumens and extracellular space. Topological change of protein from inside to outside is 'secretion.' And the major mechanism of intracellular protein transport is 'secretory pathway' (Fig.1A, highlight box). This pathway has been an interesting issue of cell biology because it is deeply related to the biogenesis of membrane-embedded protein, most of the luminal and secreted proteins.

The endoplasmic reticulum is the starting site of the secretory pathway [1], [2]. Therefore, targeting to the endoplasmic reticulum is the common process for every secretory pathway. Approximately 30% of proteomes of yeast (*S. Cerevisiae*) pass through the ER along the secretory pathway [3]–[6].

The translocation substrate enters the endoplasmic reticulum through translocon aka translocase. This translocon, called the Sec61 translocation complex, forms an hourglass-shaped pore. This protein, also called the pore complex, is well conserved in all species of bacteria, archaea and eukaryotic cells [3], [7], [8]. The Sec61 complex consists of three proteins. Sec61p is homologous with bacterial SecY and mammals Secα and has ten transmembrane domains (TM). Sec61p binds with the other two single-spanning transmembrane protein, Sbh1 (Sec61β, SecG) and Sss1 (Sec61γ, SecE). This Sec61 translocation complex has a notable kinetic characteristic.

Sec61 TM2 and TM7 form a lateral gate of translocating enzyme complex. When the lateral gate opens, the space inside the complex is exposed toward the phospholipids. At this time, the hydrophobicity gradient is generated (Fig.1B). Sufficiently hydrophobic sections of the polypeptide of the translocation substrate can be embedded into the membrane horizontally to the phospholipid tail through this lateral gate. For example, the cleavable signal sequence, signal anchor, and transmembrane domains are inserted. Sec61p has the ribosome interaction site at the portions exposed to the cytoplasm.

It contains a flexible plug domain toward ER lumen direction, and its most narrow part of hourglass shape pore constructs hydrophobic ring surrounded by hydrophobic amino acid residues. The plug and the hydrophobic ring prevent the translocation passage firmly when Sec61 is in the “closed” state. Sec61 is changed into the “open” state when the lateral gate is getting broaden, and the plug is bent.

Sec61 trimeric complex of yeast remains in static by peripheral proteins such as Sec62p, Sec63p, Sec71p and Sec72p (Fig.2B). Even though mammalian homologous for those are present, it is not abundantly found across species as Sec61. The SEC translocation complex are composed of Sec61 complex and those four subunits. The structure of the SEC translocation complex has been actively studied in recent years with advances in cryo-EM technology [9]–[12].

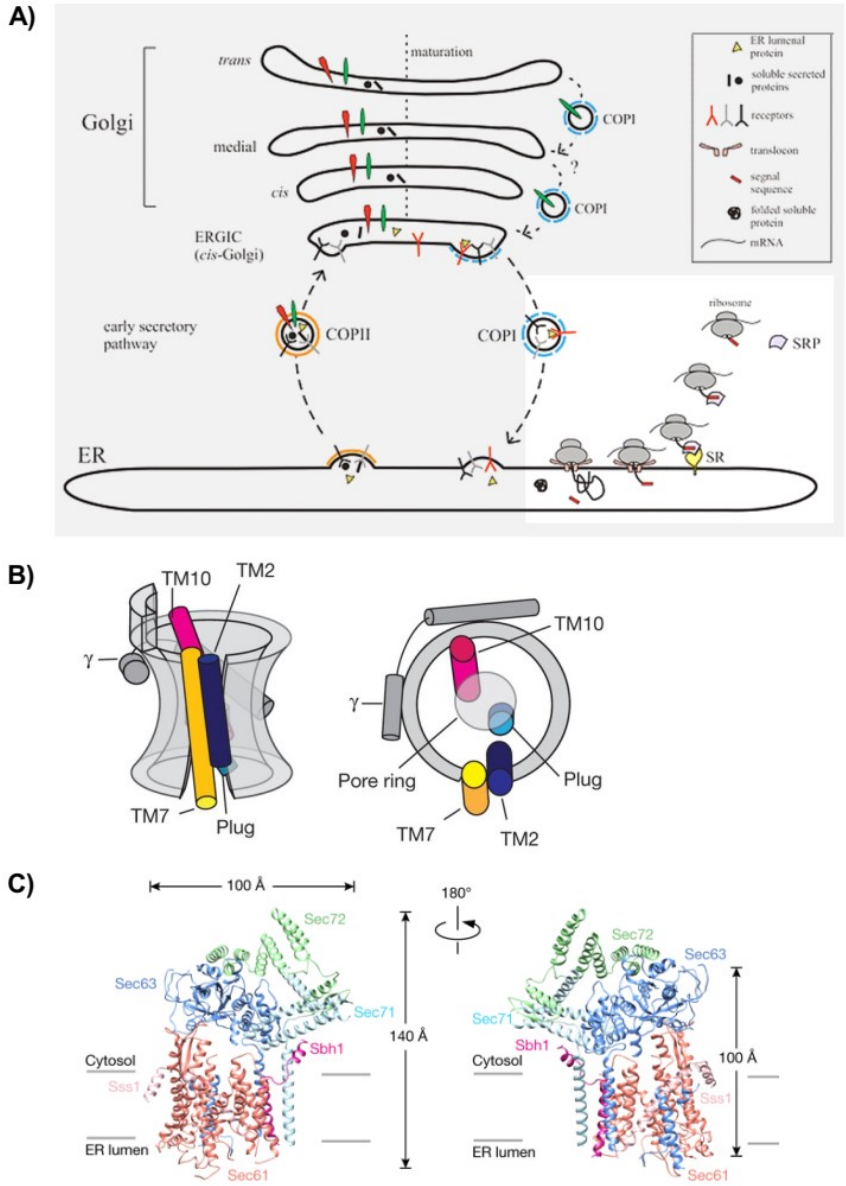


Figure 1. Schematics of the secretory pathway, and structure of the SEC translocon

(A) Cytosolic ribosomes translated the secretory proteins; then newly synthesized proteins undergo to the in-line maturation process. ER takes the role of the initial stage (retrieved from [13]). (B) The illustration of idle mammalian Sec61 complex in idle state. TM2 (deep blue) and TM7 (orange) are geometrically in skew position, and the gap between two TMs is called a lateral gate. The transparent grey circle within sectional view remarks a hydrophobic ring (retrieved from [14]). (C) Post-translocon structure. Red-to-pink colored are Sec61 complex (retrieved from [12]).

I.1.2 Subcomplexes of the SEC translocon

Subunits of SEC translocon, except the Sec61 complex, are Sec62p, Sec63p, Sec71p, and Sec72p. Although all of four proteins are interacting, mainly two groups, Sec62/63p and Sec71/72p work as subcomplexes (Fig.2). Sec62p, Sec63p, and Sec71p are membrane proteins that float around the Sec61 complex.

Two TMs of Sec62p are facing against the lateral gate. Meanwhile, the Sec63p poses in the opposite direction, where the hinge unit of Sec61p is. The two proteins form a complex by using an electromagnetic force between the positive charge at the amino-terminus of Sec62p and the negative charge of the carboxyl terminus of Sec63p.

Additionally, cytoplasmic domains of Sec63p and Sec71p interact with Sec61p of 6/7 loop. Loop 6/7 posed between TM6 and TM7. Loop6/7 provides a driving force of Sec61 changing to ‘open’ state when electric repulsion propagate through it [9], [12]. Ribosome and ribosome nascent-chain complex (RNC) share this interaction site [15].

Sec72p is cytosolic peripheral protein. The N-terminus of Sec72p and C-terminus of Sec71p interact. The C-terminal domain of Sec72p is consist of 6 helical bundles called tetratricopeptide repeat (TPR). This domain interacted the cytosolic chaperons; Ssa1 and Ssb1. Ssa1 comes in post-translocational targeting. Ssb1 is characterized in that combined with the ribosome. From this, subunits of traditional ‘post-translocon’, Sec71p and Sec72p might potentially be related to both of post- and co-translocations. [16]. It is also discovered that Sec71 deletion strain failed to survive at 37°C. However, Sec71/72 are not essential at 30°C, unlike Sec62p and Sec63p are. Many previous studies have explained the properties and structure of Sec71/72p, but studies about the function of Sec71/72 p are still elusive in many points.

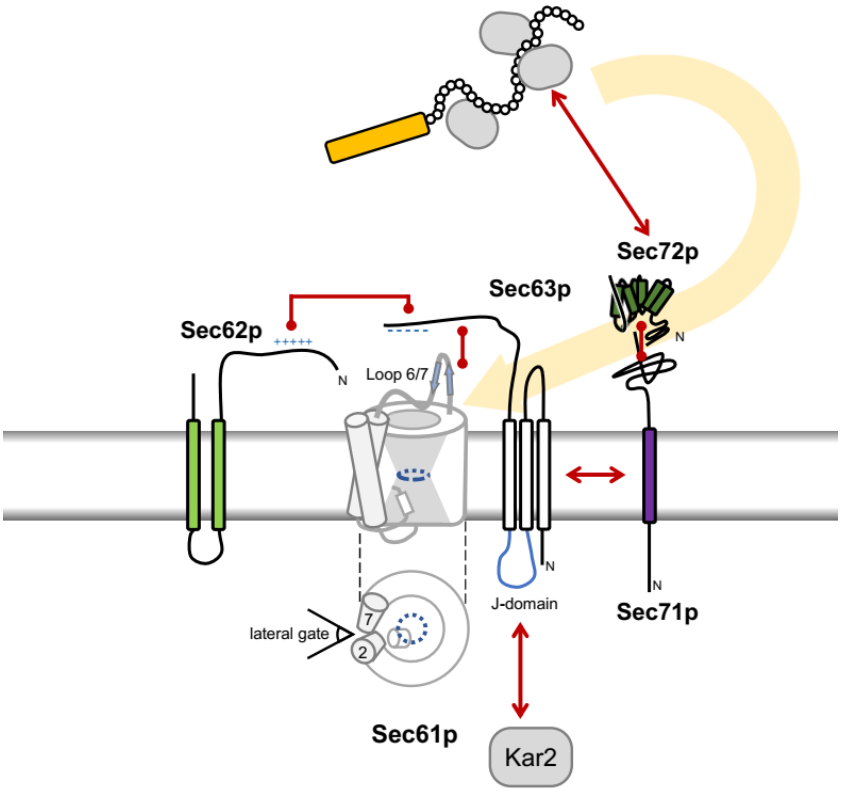


Figure 2 Schematics of yeast SEC Translocon Subunits

Post-translocon SEC heptamer has four auxiliary proteins. Sec61 complex (white and grey) in open state is facing Sec62p (yellow green). Sec63 (white) is positioned behind the hinge of Sec61p and confines Sec62p positioning at the cytosolic side. Sec71p (purple) and Sec72p (green) form tight binding. Blue character (+, -) or lines are functionally notable residue or domain. Circle-end red line remarks interaction involving site information. Arrow redline denotes interaction, but exact binding sites are unknown. Dark grey particles are the other interaction partners of SEC translocon.

I.1.3 Two routes into the ER: conventional distinction

ER protein targeting path contains roughly two steps that are guiding of a nascent chain and docking at the Sec translocon. At first, the proteins are delivered nearby to the ER surface. Cytosolic interaction partners, like SRP, SND, GET, or cytosolic chaperone network, recognize a targeting signal and drag ribosome-nascent chain along with a precursor protein to the SEC complex. Next, the translocation substrate and the pore of SEC complex interact with to initiate a translocation.

Nevertheless, targeting signals vary in its sequence and contain three-regions; N-, H- and C-region [17] (Fig.3E). N-region is usually positively charged. H-region is a short hydrophobic sequence and has a role of SRP binding. It is in the middle of the signal sequence, and leucine is primarily abundant. C-region has cleavable sites at the C-terminal end of the signal sequence (Fig.3A). Signal sequences are not always cleaved out after targeting. Generally, signal sequences are classified into two types, cleavable signal sequences and signal anchors, the latter is not cleaved and acts as a TM. In this paper, the term 'signal sequence' includes both types.

Whether translocation is coupled with translation or not is conventional division (Fig.3A, B) and referred to as co- or post-translocation [18]–[20]. The main difference between those two routes is whether a targeting depends on the signal recognition particle (SRP) or not. And the other researches have reported that many small secretory proteins are handled by post-translocation, so the length of the mutual part is related [21]. Two pathway is different at mainly working subunits of SEC complex [21]–[23]. For instance, the SEC translocon heptamer consisting of Sec61 trimer, Sec62p, Sec63p and Sec71/72p is called 'post-translocon.'

However, several recent studies [23] have reputed co- and post- distinction. It was reported that SRP enhanced post-translocation [24]. Also translating ribosome stabilized SEC complex within Sec62/63p subcomplex during *in vitro* translocation [25], and RNC-Sec63p crosslinked during co-translocation [26].

Although the common distinction between the co- and post-translocation is considered useful, more rigid identification is required.

I.1.4 The orientation of a signal sequence

The orientation of a signal sequence is a dominant topological factor of membrane proteins and is important to soluble proteins as an upstream step of a signal cleavage. Previous topology studies using signal anchor proteins have suggested that Sec62/63p complex takes a significant role in determining the signal sequence orientation [27]. Positively charged flanking regions of the signal sequence induces signal sequences to have a certain preferred position [28]. N-length and N-terminal N-glycosylation sites also affect [29].

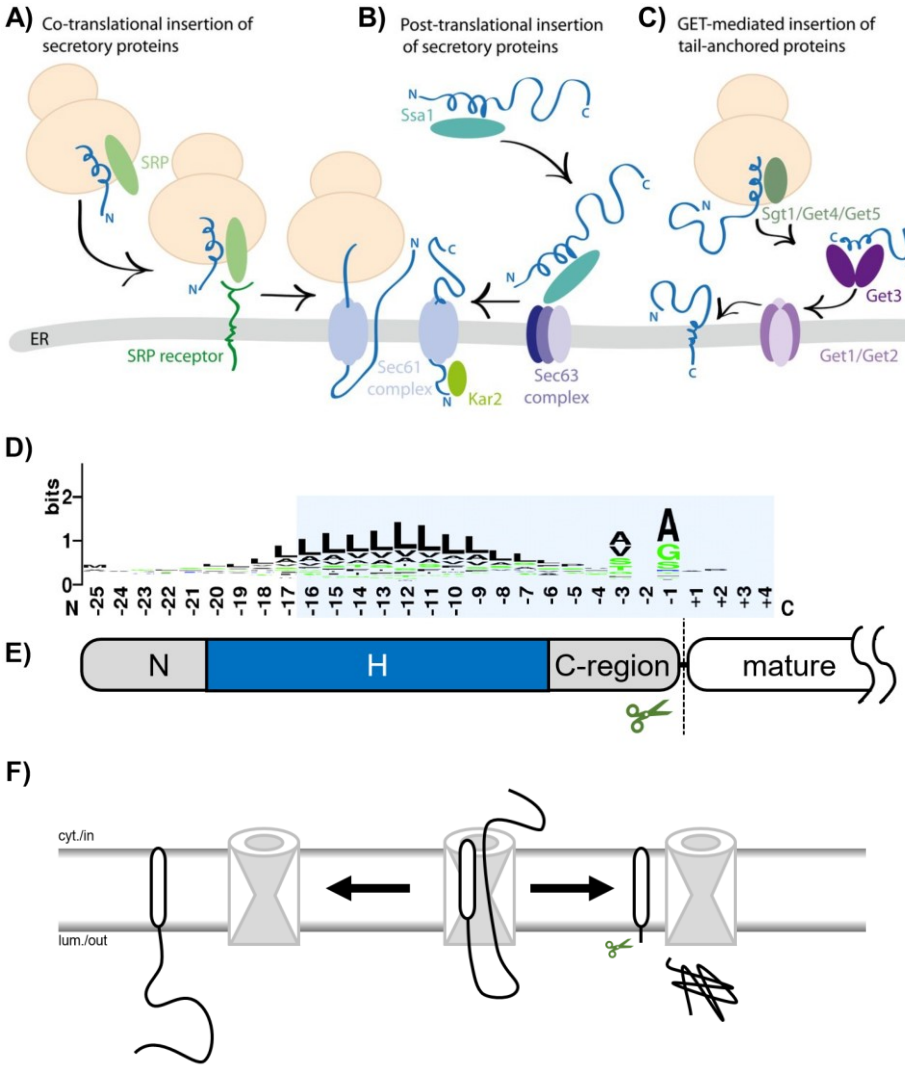


Figure 3. The Role and the Feature of Signal Sequence

(A)-(C) ER targeting mechanisms (retrieved from [30]). (D) Variation of natural signal sequences. Each letter shows the amino acid abundance of each position of the signal sequence. Positions are a distance from the cleavage site as the origin (retrieved from [31]). (E) The signal sequence is composed of 3 parts. A scissor sign and a dotted line mark cleavage site. (F) The signal anchor (left) and cleavable signal sequence (right) finally have $N_{cyt} - C_{lum}$ topology.

I.1.5 The signal sequence docking models: a head-in and inversion and a looped conformation

The final orientation of the cleavable signal sequence and signal anchor are clear (Fig.3F). However, the question of how they arrived that pose remains elusive.

When the mature part was shorter, the efficiency of translocation more decreased for human small secretory proteins. This result was interpreted as their signal sequence preferred Nlum-Ccyt at first [32]. And the positive charge is abundant on N-terminal of the signal sequence and this charge distribution is critical to translocation of mammalian small secretory pre-proteins. When signal sequence loses its N-terminal positive charge, preprotein has ER proximity. And signal cleavage and glycosylation occurred. However, translocation across the ER membrane was inhibited [33]. Therefore, these suggest that post-translocational substrate initially get a N_{lum} position but somehow is reoriented by Sec62/63. On the other hand, the structure of the Sec translocon with a bound signal sequence shows that the signal sequence headed cytosol when SEC translocon was in open state [11]. Now there are two hypotheses about the initial engagement of the signal sequence at the Sec61 complex: 'head-in and inversion' model and 'looped conformation' model.

In the head-in-inversion model, a signal sequence passes the pore of translocon N-terminal head-first, then is flipped to change orientation (Fig.4A). Oppositely, in the looped conformation, N-terminal part is folded and engaged translocon in hairpin (Fig.4B).

I.1.6 Remaining problems

Despite that the structure of the translocon is available, the underlying mechanism of protein translocation is still not revealed in many ways. In particular, how diverse features of signal sequences engage and initiate protein translocation through the same translocon remains as an unresolved question. This study focused on to characterize features of signal sequences that use the head-in or the looped conformation docking modes (Fig.4), and to determine which translocon components mediate the two processes.

1.2. Purpose of Research

My aims were to 1) investigate the interplay between signal sequence features (hydrophobicity and N-length), the translocon docking mode (head-in or looped conformation mode) and translocation pathways, and 2) to determine which types of signal sequences require Sec71 and Sec72 of the SEC complex for efficient protein translocation.

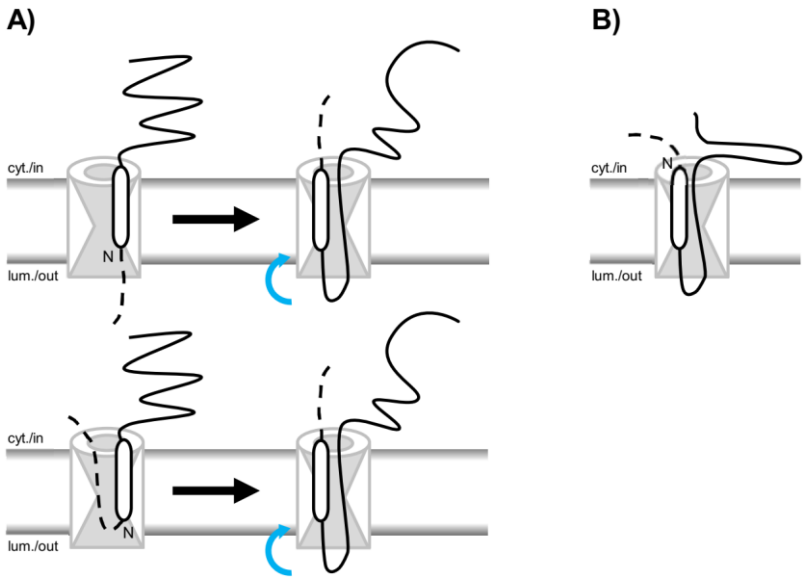


Figure 4. Two initial engagement models of substrate docking at the Sec61 complex

(A) Schematics of a head-in and inversion model. This model suggests N-terminus of the substrate is directly stuck in Sec61 pore, then re-orientated to C-terminal translocation. (B) Schematics of a looped conformation model.

METHODS

M.1 Yeast strains

I used seven yeast strains (Table 1). All strains were derived from W303-1α, which is regarded as a wild type.

BWY500 strain contains mutations in one of the SRP encoding genes (*SEC65)*. BWY500 (*sec65-1*) exhibits temperature-sensitive phenotypes that show a conditional growth defect at the nonpermissive temperature (37℃). However, BWY500 adapts to high temperature if grown for more than 4 hours at the nonpermissive temperature, so I used 15 min – 3hr incubation at 37℃ [34].

JRY4, ΔSEC71, ΔSEC72 strains are from the previous study [36].

Table 1. List of *Saccharomyces Cerevisiae* strains

Strain	Genotype	Source
W303-1α / BWY46	*MATα, ade2 his3 leu2 ura3 trp1 can1*	
BWY500	*MATα, sec65-1 ade2 his3 leu2 ura3 trp1 can1*	
JRY4	*MATα, sec62Δ::HIS3 ade2 his3 leu2 ura3 trp1 can1*	[35]
JRY4 Sec62 wt	*MATα, sec62Δ::HIS3 ade2 his3 leu2 ura3 trp1 can1* [pRS416 SEC62 wt (URA)*]*	[35]
JRY4 Sec62 35DDD	*MATα, sec62Δ::HIS3 ade2 his3 leu2 ura3 trp1 can1* [pRS415 SEC62 35DDD (LEU)]	[35]
ΔSEC71	*MATα, sec71::HIS3 ade2 his3 leu2 ura3 trp1 can1*	[36]
ΔSEC72	*MATα, sec72::HIS3 ade2 his3 leu2 ura3 trp1 can1*	[36]

M.2 Plasmid construction

I generated pRS424 plasmids containing CPY mutants by site-directed mutagenesis PCR.

A primer pair coding each end of target ORF were mixed with genomic DNA extracted from yeast. After amplification, PCR fragment and linearized vector were ligated through homologous recombination in W303-1α or through Gibson assembly assay. Target ORF was a whole protein-coding part for PDI1 and EXG1 and first 120 bases for PUG1, KTR2, GPI12, and KTR7

M.3 Prediction of N-length and hydrophobicity of signal sequences

Model translocation substrates are on the list (Table 2). Due to each model protein has modified N-region and signal sequence differ from CPY, a signal sequence position and hydrophobicity were calculated using ΔG predictor (http://dgpred.cbr.su.se/) [37], [38].

ΔG predictor is used to searching for the first hydrophobic segment. Most of other signal prediction algorithms, like SignalP, focus on the very N-terminal part due to being based on Deep learning from natural sequences. The signal sequence that starts after the third position is often not predicted. Because of the long N-region of our model protein, ΔG predictor was used instead of signal predictors.

M.4 Prediction of N-length and hydrophobicity of signal sequences

To predict a secondary structure of N-region and signal sequence for A10:CPY, I used the program YASPIN (http://www.ibi.vu.nl/programs/yaspinwww/).

M.5 Western blot

I harvested yeast in OD_{600} range between 0.5-1.0, and resuspended cell pellets in 200μL sample buffer [50mM DTT, 50mM Tris-HCl(pH7.6), 5% w/v SDS, 5% glycerol, 50mM EDTA (pH 8.0), 1mM PMSF, Bromophenol blue, Protease Inhibitor Cocktail] per 5 OD unit cells. The mixture was boiled for 5 min at 95℃ and was centrifuged maximum RPM in a table-top microcentrifuge. On the acrylamide gel, 10-15μL supernatant was loaded, and the gel was run at 75-120V voltage range. Other details are the same as previously described [27].

M.6 Autoradiographic pulse-labelling and chase of proteins

Yeast was incubated up to OD_{600} 0.3 to 0.8 at the permissive temperature (30℃, but 30℃ for BWY500) and harvested. Then cell pellets were washed by distilled water and resuspended with 1mL of -Met media. It was incubated in 30℃ (37℃ for BWY500) for 15 – 30 minutes for starvation. The starved cells were rewashed and resuspended with 150μL of -Met media per 1.5 OD unit cells. Proteins were [^{35}S-Met] radiolabeled at 30℃ shaking incubation for 5 minutes. For pulse-chase, cold met media was added after pulse-labelling. The mixture was incubated and sampled at certain time points.

To end labelling, 750μL of pre-cooled buffer A [100 mM sodium acetate, 20 mM Tris-HCl (pH 7.5)] was added to terminate protein synthesis. I added ice-cold glass beads and 100μL lysis buffer to yeast cells and lysed through 2 min vortexing. The

supernatant of cell lysates was incubated with protein G agarose beads and antibody for immunoprecipitated in IP buffer [15 mM Tris-HCl (pH7.5), 0.1% SDS, 1 mM DTT, 1 mM PMSF, 40 μM protein inhibitor cocktail]. Sample preparation from IP beads was same as with the Western blot procedure.

Gel electrophoresis process was identical as done in Western blot. Phospho-imager (FLA7000) was used to visualize radiolabeled proteins. Gel pictures were analyzed using Image Lab 6.0 software.

Table 2. List of model CPYs

Name	N-length*	ΔG* [Kcal/mol]	N-terminal sequence**	Source
CPY	0	1.299	MKAFTSLLCGLGLSTTLAKAISLQRPL	[36]
D1:CPY	2	1.933	MLKAFSSLLCGLGLSTTLAKAISLQRPL	[36]
D4:CPY	5	1.933	MLLDKKAFSSLLCGLGLSTTLAKAISLQRPL	[36]
D7:CPY	8	1.933	MKKHLLDKKAFSSLLCGLGLSTTLAKAISLQRPL	[36]
D8:CPY	9	1.933	MKKKHLLDKKAFSSLLCGLGLSTTLAKAISLQRPL	[36]
D9:CPY	10	1.933	MTKKKHLLDKKAFSSLLCGLGLSTTLAKAISLQRPL	[36]
D10:CPY	11	1.933	MDTKKKHLLDKKAFSSLLCGLGLSTTLAKAISLQRPL	[36]
D4:(1.9) CPY	5	1.933	MLLDKKAFSSLLCGLGLSTTLAKAISLQRPL	[36]
D24:(-0.5) CPY	22	-0.950	MEGGEEEEVERIPDELFDTKKKHLLDKLAFSSLLCLLLLSTTLAKAISLQRPL	[36]
D24:(-2.6) CPY	22	-2.615	MEGGEEEEVERIPDELFDTKKKHLLDKLLLTLLLCLLLLSTTLAKAISLQRPL	[36]
S6:(1.5) CPY	8	1.465	METKSISKAFTSLLCGLGLSTTLAKAISLQRPL	[36]
S7:(1.5) CPY	9	1.465	MPIVESISKAFTSLLCGLGLSTTLAKAISLQRPL	[36]
S9:(1.5) CPY	11	1.465	MPIVETKSISKAFTSLLCGLGLSTTLAKAISLQRPL	[36]
S10:(1.5) CPY	12	1.465	MEPIVETKSISKAFTSLLCGLGLSTTLAKAISLQRPL	[36]
S30:(0.5)CPY	30	0.506	MSEFNETKFSNNSTFFETEEPIVETKSISKAFSSLLCALLLSTTLAKAISLQRPL	[36]
S29:(0.0)CPY	29	0.098	MSEFNETKFSNNSTFFETEEPIVETKSISKAFSSLLCLLLLSTTLAKAISLQRPL	[36]
S27:(-0.9)CPY	27	-0.901	MSEFNETKFSNNSTFFETEEPIVETKSISLAFSSLLCLLLLSTTLAKAISLQRPL	[36]
A5	5	1.326	MAAAAAKAFTSLLCGLGLSTTLAKAISLQRPL	
A6	6	1.326	MAAAAAAKAFTSLLCGLGLSTTLAKAISLQRPL	
A7	7	1.326	MAAAAAAAKAFTSLLCGLGLSTTLAKAISLQRPL	
A8	8	1.326	MAAAAAAAAKAFTSLLCGLGLSTTLAKAISLQRPL	
A10	10	1.326	MAAAAAAAAAAKAFTSLLCGLGLSTTLAKAISLQRPL	
A14	14	1.326	MAAAAAAAAAAAAAAKAFTSLLCGLGLSTTLAKAISLQRPL	
KA5	6	1.326	MKAAAAAKAFTSLLCGLGLSTTLAKAISLQRPL	
KA10	11	1.326	MKAAAAAAAAAAKAFTSLLCGLGLSTTLAKAISLQRPL	
A5KA5	11	1.326	MAAAAAKAAAAAKAFTSLLCGLGLSTTLAKAISLQRPL	
A5K	8	1.465	MAAAAAKKAFTSLLCGLGLSTTLAKAISLQRPL	
A10K	13	1.465	MAAAAAAAAAAKKAFTSLLCGLGLSTTLAKAISLQRPL	
LLG CPY	0	0.546	MKAFTSLLCLLGLSTTLAKAISLQRPL	[36]
LLL CPY	0	-0.551	MKAFTSLLCLLLLSTTLAKAISLQRPL	[36]
A10:LLL	9	-0.568	MAAAAAAAAAKAFTSLLCLLLLSTTLAKAISLQRPL	[36]
A14:LLL	13	-0.568	MAAAAAAAAAAAAKAFTSLLCLLLLSTTLAKAISLQRPL	
A7:(ΔK)LLL	2	-2.397	MAAAAAAALAFTSLLCLLLLSTTLAKAISLQRPL	
A9:(ΔK)LLL	4	-2.397	MAAAAAAAAALAFTSLLCLLLLSTTLAKAISLQRPL	
A10:(ΔK)LLL	6	-2.397	MAAAAAAAAAALAFTSLLCLLLLSTTLAKAISLQRPL	
A14:(ΔK)LLL	10	-2.397	MAAAAAAAAAAAAAALAFTSLLCLLLLSTTLAKAISLQRPL	
PDI1	0	1.893	MKFSAGAVLSWSSLLLASSVFAQQEAVAP	
A10:PDI1	9	1.843	MAAAAAAAAAKFSAGAVLSWSSLLLASSVFAQQEAVAP	
EXG1	0	1.326	MLSLKTLLCTLLTVSSVLATPVPARDPSS	
A10: EXG1	0	0.551	MAAAAAAAAALSLKTLLCTLLTVSSVLATPVPARDPSS	

*** &** The first TM predicted sequence was regarded as a signal sequence.**

RESULTS

R.1 The extension of N-region inhibited SRP independent secretory proteins but not to SRP dependent hydrophobic signal sequences

At first, I focused on the relation between N-length and hydrophobicity of signal sequence.

The term 'N-length' is the length of N-region that starts at N-terminus and ends at upstream of the hydrophobic core of signal sequence. The hydrophobicity of a polypeptide is Gibbs free energy occurs when it resolves to a biological lipid membrane. The value goes higher, then the probability of membrane insertion of peptide gets lower. The prediction program marks the first 'comparatively hydrophobic region' of which length is within around nineteen. Because this region is regarded as a signal sequence, the proceeding part is as N-region.

The previous study reported that hydrophobic natural signal sequences tend to have longer average N-region [36]. I grouped natural signal sequences into two; hydrophilic ($\Delta G>0$) and hydrophobic ($\Delta G<0$). Then I counted the natural signal sequences having certain N-length. The 85% of hydrophilic signal sequences had N-length shorter than ten (Fig.5A, yellow) whereas hydrophobic signal sequences showed a long-tail distribution. The 85% offset was twenty-eight for hydrophobic signal sequences (Fig.5A, purple).

In the same study, the authors used model substrates having various N-region and signal sequences hydrophobicity (Table 2, green highlight). I mapped N-length and hydrophobicity or these substrates (Fig.5B). The colored dots remarks proteins of which translocation in an SRP mutant that is BWY500 (*sec65-1*) were under half of in wild type yeast. It showed that most of the hydrophobic signal sequences are SRP dependent.

Here I hypothesized SRP dependent translocation might be not affected by long N-region. To determine the effects of N-length in hydrophobic and less hydrophobic signal sequences, I prepared a set of model proteins based on CPY(PRC1) [36].

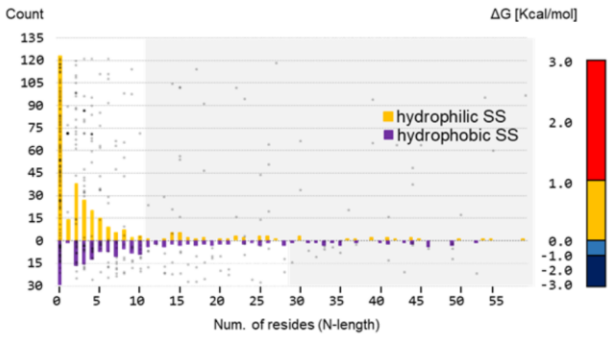
CPY signal sequence contains no N-region, that means it is present at the N-terminus. It is a well-known model protein of post-translocation [19], [29], [39]. I added ten alanine amino acids at the N-terminus of wild type CPY and LLL CPY mutant (Table 1, Fig.5E). LLL CPY mutant has point mutations at position 10 and 12, Glycine to Leucine, increasing its hydrophobicity.

CPY derivatives have 0, 10 and 14 N-length. The hydrophobicity of the CPY signal

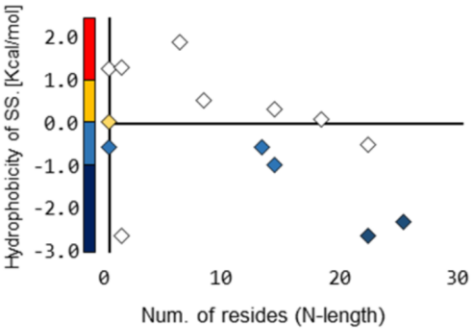
sequence is 1.933, and of the LLL CPY signal sequence is -2.615 Kcal/mol (Fig. 5C). The C-terminal translocation of model substrates was measured as band intensity ratio of glycosylated and unglycosylated forms from electrophoresis. A glycosylated version of protein increases its molecular weight of about 2 kDa per glycan. Thus, two types behave differently on their gel migration.

The translocation efficiency of CPY variants was assessed in wild type cell, 35DDD Sec62 mutant and SRP mutant BWY500 (*sec65-1*). The translocation of wild type CPY was completely disrupted by long N-region made of 10 and 14 alanine stretch whereas translocation of A10:LLL and A14:LLL CPY were shown to be efficient (Fig.5D). The CPY signal sequence is hydrophilic, and its translocation depends on Sec62 but not on SRP. Hydrophobic LLL CPY showed Sec62 independent, and SRP dependent C-terminal translocation (Fig.5E). The Sec62 and SRP dependence of A10:CPY was not able to know from 5 min pulse-labelling because translocation of A10:CPY in wild type was zero. To test the dependence of A10:CPY, I expressed A10:CPY and did 0, 5, 10 min pulse-chase in three yeast strain (Fig.5F). The portion of glycosylated from gradually increased in W303-1α and BWY500 (sec65-1). In 35DDD Sec62 mutant, initial translocation rate was 4.8%, and only 3% point increased after 10 min.

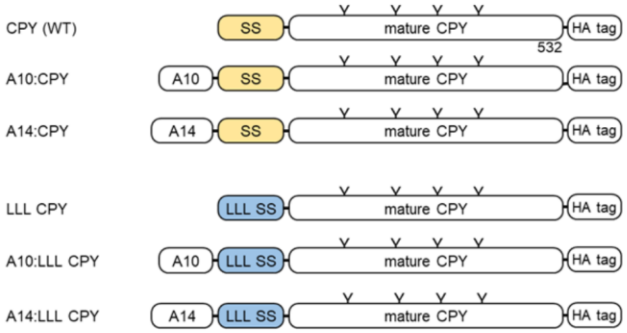
A)


Count
ΔG [Kcal/mol]
hydrophilic SS
hydrophobic SS
Num. of resides (N-length)

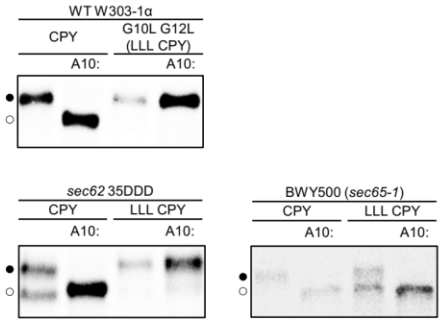
B)


Hydrophobicity of SS. [Kcal/mol]
Num. of resides (N-length)

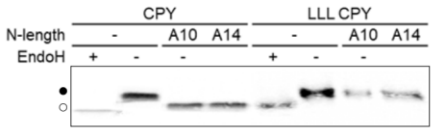
C)


CPY (WT)
SS
mature CPY
HA tag
532
A10:CPY
A10
A14:CPY
A14
LLL CPY
LLL SS
A10:LLL CPY
A14:LLL CPY

E)


WT W303-1α
CPY
G10L G12L (LLL CPY)
A10:
sec62 35DDD
BWY500 (sec65-1)
LLL CPY

D)


CPY
LLL CPY
N-length
EndoH
A10
A14

F)

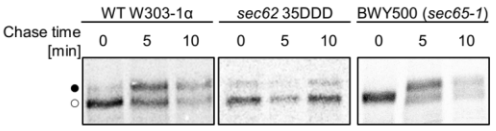

WT W303-1α
sec62 35DDD
BWY500 (sec65-1)
Chase time [min]
0
5
10

Figure 5. N-length and hydrophobicity of various signal sequences

(A) The bar graphs show a number of natural proteins of the yeast of which N-region is consist of a certain number of residues. The yellow bar is for hydrophilic signal, and purple bar is for hydrophobic signal sequences. The scatter chart of grey dots is hydrophobicity map to N-length. The hydrophobicity ranges of the signal sequence are noted at the right axis. (B) The CPY variants used in previous studies are plotted by N-length and hydrophobicity of signals sequences. From the same study, CPY variants that showed SRP dependent translocation are colored by their hydrophobicity level. The color bar is shown beside the hydrophobicity axis. (C) CPY series and LLL CPY series are schematically drawn. Extended N-region, signal sequences region, mature part and HA tag are separated. The V marks are glycosylation sites. (D) The gel of 5 min pulse-labelling of CPY and LLL CPY are aligned. A black dot indicated the size of fully glycosylated of CPY, and a hollow indicates the size of unglycosylated or deglycosylase form. (E) The CPY, A10:CPY, LLL CPY and A10:LLL CPY are compared. Background strains were written above each gel. (F) The result of A10:CPY chase is represented. The chase time and background strain were noted. (D)-(F) Every electrophoresis was run in 7% acrylamide gel.

R.2 The length of N-region is critical to inhibit the translocation of SRP independent CPY signal sequences than the charge, secondary structure, and mature part of preproteins.

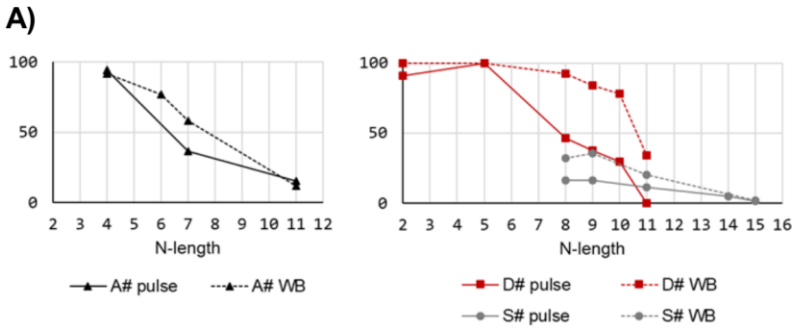
The inhibitory effect of extended N-region on the SRP independent CPY signal sequence was shown by A10:CPY length over ten. In this section, I investigated which factor of A10 caused translocation defect.

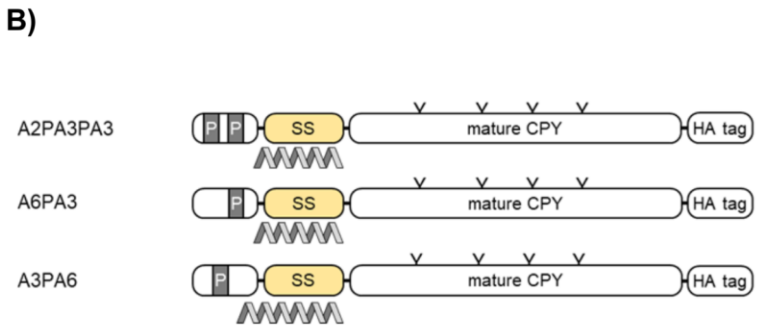
I modified N-region of which the length, the origin of sequence, and secondary structure. There were three origins of extended N-region of CPY derivatives. The N-region of A# CPY series is simple alanine repeating. The N-region of D# CPY and S# CPY were derived from DAP2 cytosolic domain and Sec71 luminal domain (Table 2). The translocation efficiency and the rate of substrates were assessed by 5 min pulse-labelling (Fig.6A).

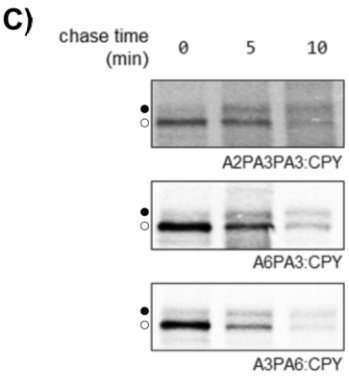
The translocations of A#, D#, and S# CPYs were inhibited depend on N-length and defect linearly increased. At A# CPY, the translocation defect by pulse-labelling was shown from length 6 and was decreased to 12% than the original CPY. The inhibition of translocation of D# CPY started at length around N-length 8 then hit zero at 11. The ratio of the translocated form of S6:CPY was half fold than A6:CPY and one-third than D7:CPY. However, it went to zero at N-length 15.

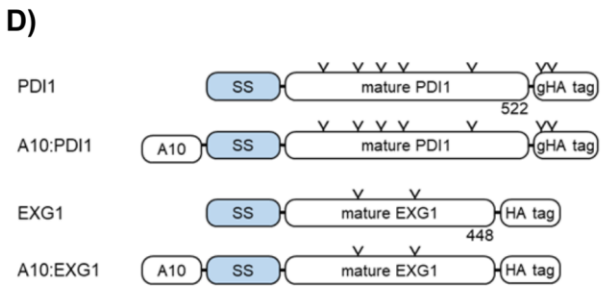
The alanine stretch was the most inhibitory among A# CPY, D# CPY and S# CPY. I attended to that alanine easily formed helical structure than the other two. Thus, one or two prolines were inserted to alanine stretch to break α-helix involving whole N-region and signal sequence. However, translocation of A2PA3PA3, A6A3 and A3PA6:CPY was not improved against A10:CPY (Fig.6B, C).

To test whether this inhibitory effect of long N-region in protein translocation is found in other proteins, PDI1 and EXG1, two other yeast proteins were additionally prepared, and their translocation was assessed. PDI1 and EXG1 naturally have 0 N-length and mildly hydrophilic signal sequence like CPY (Table 2, Fig.6D). Original PDI1 showed size shift at 20 min chase. That means, translocation was slow but occurred slowly. The 5 min pulse-labelling result was different between EXG1 and A10:EXG1. A10:EXG1 showed less glycosylated from (Fig.6E). Therefore, the alanine N-region inhibited translocation of PDI1 and EXG1 as it did to CPY.


A)
100
50
0
2 3 4 5 6 7 8 9 10 11 12
N-length
A# pulse
A# WB
100
50
0
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
N-length
D# pulse
D# WB
S# pulse
S# WB


B)
A2PA3PA3
A6PA3
A3PA6
P
SS
mature CPY
HA tag


C)
chase time (min)
0
5
10
A2PA3PA3:CPY
A6PA3:CPY
A3PA6:CPY


D)
PDI1
SS
mature PDI1
gHA tag
522
A10:PDI1
A10
EXG1
mature EXG1
HA tag
448
A10:EXG1

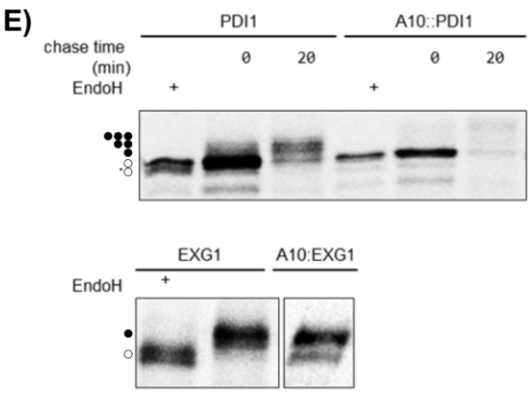

E)
PDI1
A10::PDI1
chase time (min)
0
20
0
20
EndoH
+
+
EXG1
A10:EXG1
EndoH
+

Figure 6. Effect of long N-regions to translocation

(A) The translocation efficiency of A# CPY (left), D# CPY (right, red) and S# CPY (right, grey) are mapped. The Western blot results are in the dotted line, and the 5 min pulse-labelling results are in solid line. (B) Schematics of the helix-break versions of A10:CPY are shown. The grey bands indicate the positions of prolines in N-region. The grey-light grey zigzags are marks of secondary structure prediction result of N-regions and signal sequences. (C) The translocation of the helix-break version described in B is shown. The chase time is 0, 5, 10 min and indicated above the gel. (D) The translocation test substrate derived from protein CPY-like natural protein, PDI and EXG1 are parted into three sections. The PDI1 tagged glycosylatable HA tag. (E) The top is the pulse-chase result of PDI1 and A10:PDI1. The bottom is 5 min pulse-labelling result of EXG1 and A10:EXG1. The electrophoresis of EXG1 and A10:EXG1 is on 10% acrylamide gel.

R.3 The charge distribution along N-region modulates the temporal translocation rate than the initial engagement of SRP independent signal sequences

One of the characteristics of alanine stretch relative to N-region of D# and S# CPY is 'chargeless.' (Fig.6A). These data suggested there is an effect of charges to translocation of the CPY signal sequence.

To investigate the effect of charges on translocation, I introduced a single lysine into alanine stretch, which is the N-region of A10:CPY. There are three derivatives have different distance from lysine to the starting point of signal sequence; KA10, A10K and A5KA5:CPY (Fig.7A). At the time point, 0 min of pulse-chase, A10K and KA10:CPY initially translocated similar with A10:CPY. Translocation ratios went higher when the time arrived, are different among the three models CPYs (Fig.7C). However, there were delicate gaps depended on the charge position. When the lysine was close to signal sequence starting point, that translocation rate might increase.

For shortening N-length, KA5 and A5K:CPY were translocated better than A6:CPY. It seems that a single positive charged residue could hasten signal re-orientation by a positive-inside-rule, so 5 min pulse-labelling could not provide enough time-resolution for them. Meanwhile, A5KA5:CPY has common N-region with KA5. The translocation of A5KA5:CPY was about zero in any time points (Fig.7D).

Here, I showed that N-loop formation model is not suitable to explain the translocation of CPY signal sequence (Fig.7E).

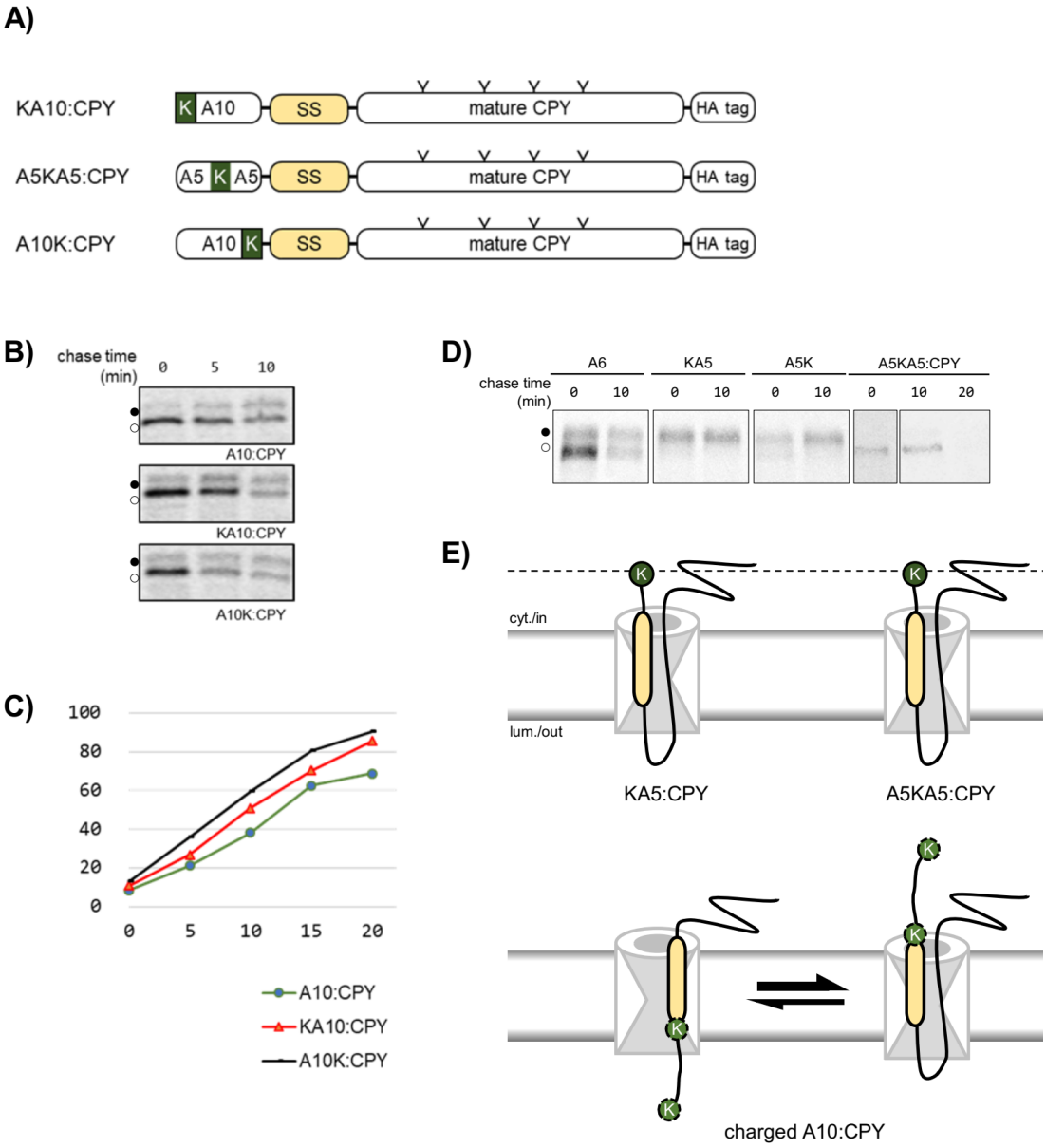

A)
KA10:CPY
K
A10
SS
mature CPY
HA tag
A5KA5:CPY
A5
K
A5
SS
mature CPY
HA tag
A10K:CPY
A10
K
SS
mature CPY
HA tag
B)
chase time (min)
0
5
10
A10:CPY
KA10:CPY
A10K:CPY
C)
100
80
60
40
20
0
0
5
10
15
20
A10:CPY
KA10:CPY
A10K:CPY
D)
A6
KA5
A5K
A5KA5:CPY
chase time (min)
0
10
0
10
0
10
0
10
20
E)
cyt./in
lum./out
K
KA5:CPY
A5KA5:CPY
charged A10:CPY

Figure 7. The translocation of charge introduced A# CPYs

(A) Schematics of A# CPY variants reflect the position of lysine in deep green. (B) Pulse-chase data at three time points are enumerated. (C) The plot is a quantification of B. Glycosylation rates were compared. (D) Two pints pulse-chase of charged A# CPY having 6 N-length and A5KA5:CPY. (E) Model suitability of KA5:CPY and A5KA5 are graphically are expressed. N-loop formation model (top) and head-in-inversion model (bottom) are drawn. Dotted line means KA5 and A5KA5:CPY share the part of N-terminal sequences. Light green circles are the position of lysine at KA10 and A10K.

R.4 Translocation of weakly positive charge-biased signal sequences required Sec62, Sec71/72 even if they were SRP dependent

Signal sequences have the feature that the positive charge is abundant at right upstream of H-region. Because the charge-bias of the signal sequence is related to re-orientation of head-in and inversion substrates, I tested the effect of charge at SRP dependent substrates.

I focused on the lysine at the second position of a CPY. Exchange lysine to leucine makes the signal sequence more hydrophobic and reverses charge gradient on signal sequence (Fig.8A). These substrates were named (ΔK)LLL CPY. Prolonged N-regions of (ΔK)LLL CPY are ten and fourteen alanine. Against to LLL CPY versions, (ΔK)LLL CPY having same N-length depended on Sec62 and SRP both to translocation (Fig.8B, C). Moreover, Sec71/72p dependence also appeared at A10 and A14:(ΔK)LLL CPYs (Fig.8D, E).

Because of the translocation of A14:(ΔK)LLL was defective than A10:(ΔK)LLL CPY, it was tested whether inhibitory effect to (ΔK)LLL CPY depended on N-length or did not. When A7, A9, A10 and A14:(ΔK)LLL CPYs were expressed in ΔSEC72, the translocation defect was gradually increased as N-length increased (Fig. 8F). Same test in ΔSEC71, four length derivatives were not translocated (not printed).

These behaviors were similar with the results of SRP independent substrates: the Sec62 dependence and the length dependent inhibition.

Then I tested whether the charge dependent feature that difference between LLL CPYs and (ΔK)LLL CPYs was general phenomena to SRP dependent signal sequences or not. I found 26 natural yeast proteins of which the signal sequence was hydrophobic, and the length of N-region was in range ten to twelve. I picked 4 proteins and merged their N-terminal 40 amino acids in the CPY model substrate instead of an original CPY signal sequence (Table 3, Fig.9 A).

Because derived from natural signal sequences, the N- region of these chimera proteins are diverse in charge distribution in N-terminal parts. I counted the net charge of ‘upstream’ and ‘downstream’ regions. The ‘upstream’ region starts from the N-terminus to the middle of the signal sequences, and the ‘downstream’ region started from right next upstream to the 20th flanking residue after end of signal sequences. The difference of net charge of two regions are called Δq.

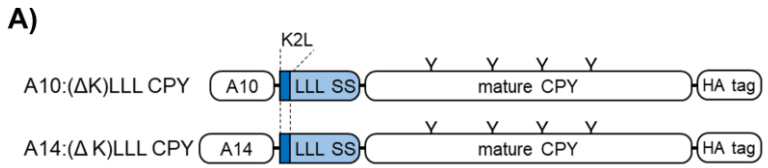
The PUG1(40):CPYm has the smallest Δq value. And the translocation of

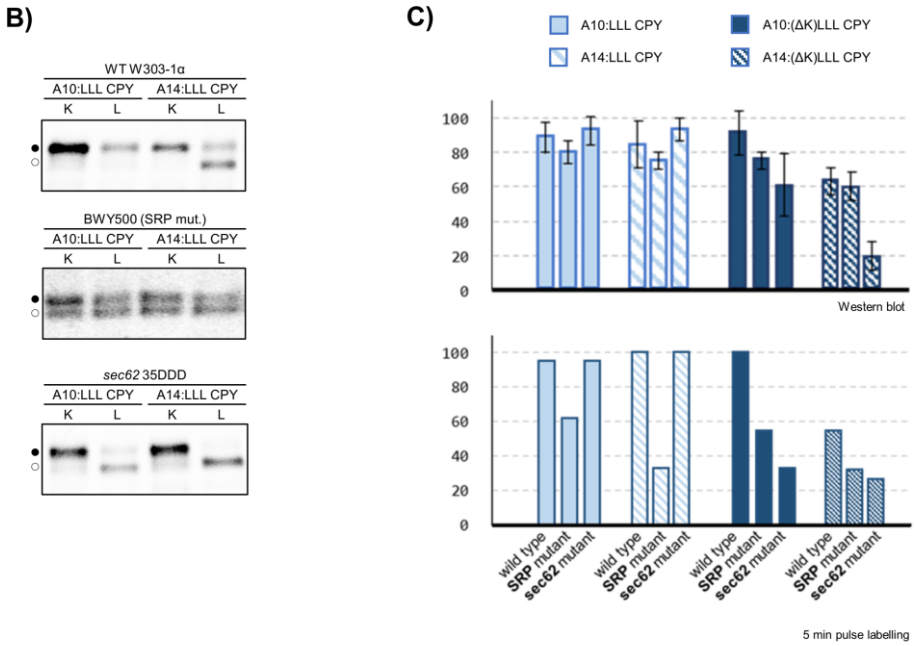
PUG1:CPYm was significantly defected in wild type, Sec62 mutant strain, Sec71 and Sec72 deletion strains (Fig.9B). The translocation of KTR2:CPYm, GPI12:CPYm and KTR7:CPYm were only inhibited in ΔSEC71.

Table 3. List of the chimaera CPYs

Protein name	N-length	ΔG [Kcal/mol]	Δq	N-length Sequence	Signal sequence	C-terminal flanking 20 residues
PUG1(40):CPY$_m$	17	-2.876	1	MSTTDSGFVLYHYTPSK	AAAIVFVVLFIIMTVIFAVQTLY	SLQRPLGLDKDVLLQAAEKF
KTR2(40):CPY$_m$	10	-1.987	2	MQICKVFLTQ	VKKLLFVSLLFCLIAQTCWLALV	LALVPYQRQLSSLQRPLGLD
GPI12(40):CPY$_m$	13	-0.891	5	MKMLRRTKVNFSK	LLYKITKLAIVLTILYIYFT	PKIVSRNSLQRPLGLDKDVL
KTR7(40):CPY$_m$	20	-2.133	7	MAIRLNPKVRRFLLDKCRQK	RYGFLFLGCIFAILYCMGTWSL	QRPLGLDKDVLLQAAEKFGL

'Δq' is the difference between the N-region charge count and C-region charge count and is regarded as a parameter of charge gradient through signal sequences. Positive charges (K, R, H) are weight as +1 and negatives (D, E) are -1.


A)
K2L
A10:(ΔK)LLL CPY
A10
LLL SS
mature CPY
HA tag
A14:(Δ K)LLL CPY
A14
LLL SS
mature CPY
HA tag


B)
WT W303-1α
A10:LLL CPY
A14:LLL CPY
K
L
K
L
BWY500 (SRP mut.)
A10:LLL CPY
A14:LLL CPY
K
L
K
L
sec62 35DDD
A10:LLL CPY
A14:LLL CPY
K
L
K
L
C)
A10:LLL CPY
A10:(ΔK)LLL CPY
A14:LLL CPY
A14:(ΔK)LLL CPY
100
80
60
40
20
0
Western blot
100
80
60
40
20
0
wild type
SRP mutant
sec62 mutant
wild type
SRP mutant
sec62 mutant
wild type
SRP mutant
sec62 mutant
wild type
SRP mutant
sec62 mutant
5 min pulse labelling

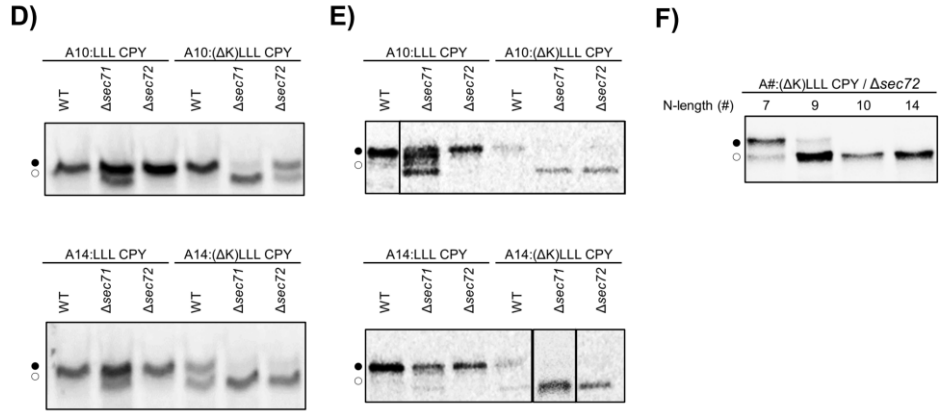

D)
A10:LLL CPY
A10:(ΔK)LLL CPY
WT
Δsec71
Δsec72
WT
Δsec71
Δsec72
A14:LLL CPY
A14:(ΔK)LLL CPY
WT
Δsec71
Δsec72
WT
Δsec71
Δsec72
E)
A10:LLL CPY
A10:(ΔK)LLL CPY
WT
Δsec71
Δsec72
WT
Δsec71
Δsec72
A14:LLL CPY
A14:(ΔK)LLL CPY
WT
Δsec71
Δsec72
WT
Δsec71
Δsec72
F)
A#:(ΔK)LLL CPY / Δsec72
N-length (#)
7
9
10
14

Figure 8. The effect of deletion positive charge from N-terminal of the signal sequence

(A) The schematics of model substrates show forms of A10:LLL CPY to A14:(ΔK)LLL CPY. K2L points the mutation on the second amino acid. (B) The gel data shows the comparisons between LLL CPYs and (ΔK)LLL CPYs. Position 2 was remarked on the column label. (C) Plot represented quantification results of Western blot (top) and 5 min pulse-labelling (bottom). (D) The gel data of Western blot gel shows translocational efficiencies of LLL CPY and (ΔK)LLLCPY in wild type and mutant strains. Δ71 is Δ SEC71 mutants and vice vera to Δ72. (E) The gel data of 5 min pulse-labelling is marked as similar as D. (F) Length variants of (ΔK)LLL CPY were expressed in ΔSEC72 strains.

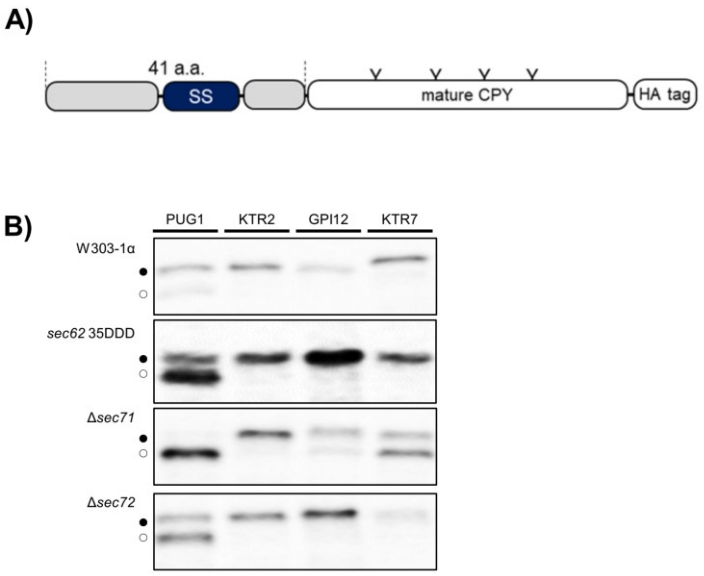


Figure 9. Sec71/72 dependence of chimera CPY constructs consist of 40 amino acids derived from natural long N-region and hydrophobic signal sequences

(A) The schematics of four chimera proteins are presented. The N-terminal 41 amino acids include start methionine are drawn in three parts. The hydrophobic natural signal sequence is colored in navy. The grey areas are up and down flanking region. The 40 amino acids were introduced instead of CPY signal sequence. (B) The 5 min pulse-labelling results are aligned in order of strains; wild type, 35DDD Sec62 mutant, ΔSEC71 and ΔSEC72. The names of chimera proteins are abbreviated into the protein names that is origin of the first 40 residues.

DISCUSSION

The protein targeting and translocation into the ER are a vital cellular process. Thus, the secretory pathway has been important to understand organellar protein genesis. However, the very early stage of a substrate and a translocon lacks dynamic information because it is structurally instable and too fast.

In this paper, the existing two docking models were examined focusing on the hydrophobicity of signal sequences. I constructed various CPY derived model substrates of which translocation efficiency and SEC subunit dependence are different.

My working hypothesis is that the docking process depends on whether translocation is coupled with translation or not. The post-translocational substrate follows head-in and inversion only; on the other hand, co-translocational substrate follows two modes alternatively (Fig.10). If N-region gets longer, the flipping of a signal sequence inside pore might be blocked or get slower. Thus, to expending N-region can decrease the translocational efficiency of a substrate following head-in and inversion process. And a looped conformation might not be disrupted it translocation by long N-region, because re-orientation process is not necessary to it.

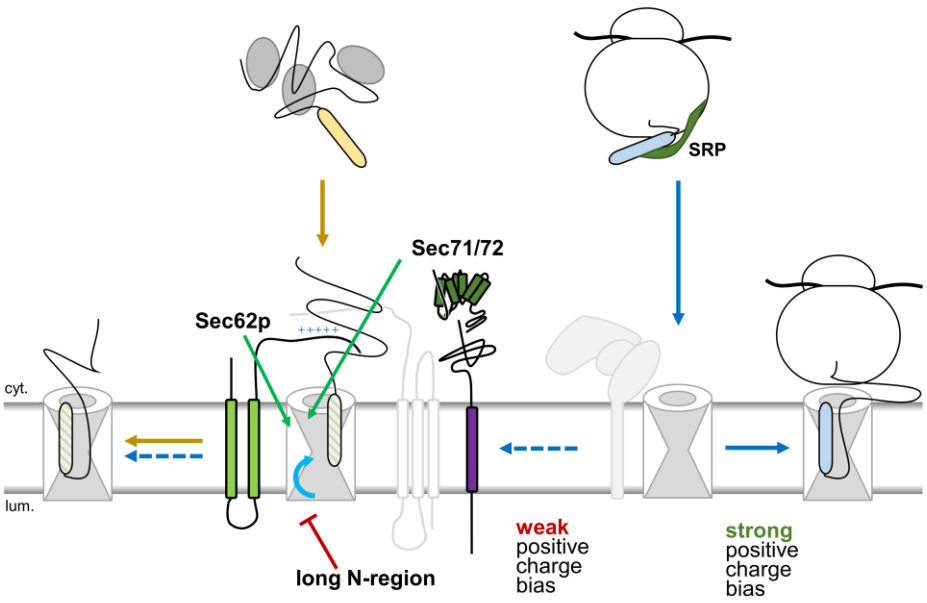


Figure 10. The translation-translocation coupling and the initial engagement between a substrate and a translocon

However, it was reported that a reside of the second position is a cytosolic target and acetylation efficiency goes different residue by residue by translocation assay of CPY mutations. The acetylation of N-terminus blocks the translocation. Thus, to investigate the inhibitory effect of N-length of translocation, aspartate and glutamate are not appropriate residues to second residue [40].

The translocation of original CPY signal sequence was affected by long N-region, and this implies that the CPY signal sequence follows head-in and inversion mode. By positive-inside-rule, positively charged residue tends to localize at cytosol at signal sequence orientation. When N-region contained many some residues, translocation level and the rate increased. From comparison among A10, KA10 and A10K:CPY, the distance of positive charges and start position of a signal sequence is potentially significant to re-orient a signal sequence.

The translocation of hydrophobic signal sequence of LLL CPY series depended on SRP and were not inhibited by long N-region. The signals sequence CPY and LLL CPY contain two positive charges that each are closed to N-terminus or C-terminus of signal sequence. When I exchanged N-terminal lysine at second position to leucine, A10:(ΔK)LLL CPY still fully translocated in wild type cell despite of long N-region. And (ΔK)LLL CPY series were shown SRP, Sec62, and Sec71/72 dependence. The translocational efficiency of (ΔK)LLL CPY at ΔSEC72 decreased as N-length increased.

Here, it could be suggested that the hydrophobic and SRP dependent signal sequence follows looped conformation. Although an orientational signal of signal sequence by charge-bias gets weak, the translocation through SEC heptamer follows looped conformation. Oppositely, if one of the Sec62, Sec71 and Sec72 is malfunctional, signal sequence with weak charge-bias follows head-in and inversion.

Furthermore, Sec71/72 might function as fastening N-terminus to cytosol at initial engagement of loop formation or after flipping of head-in and inversion mode because it was reported that translocation phenotype of ΔSEC72 was similar to the phenotype of 35DDD Sec62 mutant strains were similar [36]. The Sec62p helps re-orientation of [27], thus function of Sec71/72p could be inferred.

The co- and post-translation of ER targeting are a conventionally useful concept to study the secretory pathway. However, there are substrate dependent on both of SRP and Sec62. This result was conflict to the argument that co- and post-translocational

pathway are exclusive.

This study assumed that the defective effect of signal sequence re-orientation is dominant than of the delivery failure from cytosol to ER surface. However, the how much long N-region disrupt ER proximity of SRP independent substrates was not well considered here. I will insert a glycosylation site to N-terminus of A10:CPY and A14:(ΔK)LLL CPY to test the efficiency of initial engagement.

References

[1] G. Blobel and B. Dobberstein, “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.*, vol. 67, no. 3, pp. 835–851, Dec. 1975.

[2] P. Novick, S. Ferro, and R. Schekman, “Order of events in the yeast secretory pathway,” *Cell*, vol. 25, no. 2, pp. 461–469, Aug. 1981.

[3] E. Wallin and G. V. Heijne, “Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms,” *Protein Sci.*, vol. 7, no. 4, pp. 1029–1038, Apr. 1998.

[4] W.-K. Huh *et al.*, “Global analysis of protein localization in budding yeast,” *Nature*, vol. 425, no. 6959, pp. 686–691, Oct. 2003.

[5] Y. Chen *et al.*, “SPD—a web-based secreted protein database,” *Nucleic Acids Res.*, vol. 33, no. suppl_1, pp. D169–D173, Jan. 2005.

[6] J. Choi, J. Park, D. Kim, K. Jung, S. Kang, and Y.-H. Lee, “Fungal Secretome Database: Integrated platform for annotation of fungal secretomes,” *BMC Genomics*, vol. 11, no. 1, p. 105, Feb. 2010.

[7] M. Pohlschröder, K. Dilks, N. J. Hand, and R. Wesley Rose, “Translocation of proteins across archaeal cytoplasmic membranes,” *FEMS Microbiol. Rev.*, vol. 28, no. 1, pp. 3–24, Feb. 2004.

[8] T. Cranford-Smith and D. Huber, “The way is the goal: how SecA transports proteins across the cytoplasmic membrane in bacteria,” *FEMS Microbiol. Lett.*, vol. 365, no. 11, Jun. 2018.

[9] T. Becker *et al.*, “Structure of Monomeric Yeast and Mammalian Sec61 Complexes Interacting with the Translating Ribosome,” *Science*, vol. 326, no. 5958, pp. 1369–1373, Dec. 2009.

[10] R. M. Voorhees and R. S. Hegde, “Toward a structural understanding of co-translational protein translocation,” *Curr. Opin. Cell Biol.*, vol. 41, pp. 91–99, Aug. 2016.

[11] R. M. Voorhees and R. S. Hegde, “Structure of the Sec61 channel opened by a signal sequence,” *Science*, vol. 351, no. 6268, pp. 88–91, Jan. 2016.

[12] S. Itskanov and E. Park, “Structure of the posttranslational Sec protein-translocation channel complex from yeast,” *Science*, p. eaav6740, Dec. 2018.

[13] C. Viotti, “ER to Golgi-Dependent Protein Secretion: The Conventional Pathway,” in *Unconventional Protein Secretion: Methods and Protocols*, A. Pompa and F. De Marchis, Eds. New York, NY: Springer New York, 2016, pp. 3–29.

[14] M. Gogala *et al.*, “Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion,” *Nature*, vol. 506, no. 7486, pp. 107–110, Feb. 2014.

[15] M. D. Potter and C. V. Nicchitta, “Endoplasmic Reticulum-bound Ribosomes Reside in Stable Association with the Translocon following Termination of Protein Synthesis,” *J. Biol. Chem.*, vol. 277, no. 26, pp. 23314–23320, Jun. 2002.

[16] A. Tripathi, E. C. Mandon, R. Gilmore, and T. A. Rapoport, “Two alternative binding mechanisms connect the protein translocation Sec71-Sec72 complex with heat shock

proteins," *J. Biol. Chem.*, vol. 292, no. 19, pp. 8007–8018, May 2017.

[17] G. V. Heijne, "Patterns of Amino Acids near Signal-Sequence Cleavage Sites," *Eur. J. Biochem.*, vol. 133, no. 1, pp. 17–21, 1983.

[18] B. C. Hann and P. Walter, "The signal recognition particle in S. cerevisiae," *Cell*, vol. 67, no. 1, pp. 131–144, Oct. 1991.

[19] D. T. Ng, J. D. Brown, and P. Walter, "Signal sequences specify the targeting route to the endoplasmic reticulum membrane.," *J. Cell Biol.*, vol. 134, no. 2, pp. 269–278, Jul. 1996.

[20] C. Rothe and L. Lehle, "Sorting of invertase signal peptide mutants in yeast dependent and independent on the signal-recognition particle," *Eur. J. Biochem.*, vol. 252, no. 1, pp. 16–24, Dec. 2001.

[21] A. K. K. Lakkaraju, R. Thankappan, C. Mary, J. L. Garrison, J. Taunton, and K. Strub, "Efficient secretion of small proteins in mammalian cells relies on Sec62-dependent posttranslational translocation," *Mol. Biol. Cell*, vol. 23, no. 14, pp. 2712–2722, May 2012.

[22] Y. Harada, H. Li, J. S. Wall, H. Li, and W. J. Lennarz, "Structural Studies and the Assembly of the Heptameric Post-translational Translocon Complex," *J. Biol. Chem.*, vol. 286, no. 4, pp. 2956–2965, Jan. 2011.

[23] N. Johnson *et al.*, "The Signal Sequence Influences Post-Translational ER Translocation at Distinct Stages," *PLOS ONE*, vol. 8, no. 10, p. e75394, 9 2013.

[24] B. M. Abell, M. R. Pool, O. Schlenker, I. Sinning, and S. High, "Signal recognition particle mediates post-translational targeting in eukaryotes," *EMBO J.*, vol. 23, no. 14, pp. 2755–2764, Jul. 2004.

[25] B. J. Conti, P. K. Devaraneni, Z. Yang, L. L. David, and W. R. Skach, "Cotranslational Stabilization of Sec62/63 within the ER Sec61 Translocon Is Controlled by Distinct Substrate-Driven Translocation Events," *Mol. Cell*, vol. 58, no. 2, pp. 269–283, Apr. 2015.

[26] C. H. Jan, C. C. Williams, and J. S. Weissman, "Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling," *Science*, vol. 346, no. 6210, pp. 1257521–1257521, Nov. 2014.

[27] J. H. Reithinger, J. E. H. Kim, and H. Kim, "Sec62 protein mediates membrane insertion and orientation of moderately hydrophobic signal anchor proteins in the endoplasmic reticulum (ER)," *J. Biol. Chem.*, vol. 288, no. 25, pp. 18058–18067, Jun. 2013.

[28] A. Elazar, J. J. Weinstein, J. Prilusky, and S. J. Fleishman, "Interplay between hydrophobicity and the positive-inside rule in determining membrane-protein topology," *Proc. Natl. Acad. Sci.*, vol. 113, no. 37, pp. 10340–10345, Sep. 2016.

[29] S. F. Trueman, E. C. Mandon, and R. Gilmore, "A gating motif in the translocation channel sets the hydrophobicity threshold for signal sequence function," *J Cell Biol*, vol. 199, no. 6, pp. 907–918, Dec. 2012.

[30] C. K. Barlowe and E. A. Miller, "Secretory Protein Biogenesis and Traffic in the Early Secretory Pathway," *Genetics*, vol. 193, no. 2, pp. 383–410, Feb. 2013.

[31] K. Hiller, A. Grote, M. Scheer, R. Münch, and D. Jahn, "PrediSi: prediction of signal peptides and their cleavage positions," *Nucleic Acids Res.*, vol. 32, no. suppl_2, pp. W375–W379, Jul. 2004.

[32] V. Goder and M. Spiess, “Molecular mechanism of signal sequence orientation in the endoplasmic reticulum,” *EMBO J.*, vol. 22, no. 14, pp. 3645–3653, Jul. 2003.

[33] H. Guo *et al.*, “Positive charge in the n-region of the signal peptide contributes to efficient post-translational translocation of small secretory preproteins,” *J. Biol. Chem.*, p. jbc.RA117.000922, Dec. 2017.

[34] J. D. Brown, D. T. Ng, S. C. Ogg, and P. Walter, “Targeting pathways to the endoplasmic reticulum membrane,” *Cold Spring Harb. Symp. Quant. Biol.*, vol. 60, pp. 23–30, 1995.

[35] S. Jung, Y. Jung, and H. Kim, “Proper insertion and topogenesis of membrane proteins in the ER depend on Sec63,” *Biochim. Biophys. Acta BBA - Gen. Subj.*, vol. 1863, no. 9, pp. 1371–1380, Sep. 2019.

[36] C. Yim, S. Jung, J. E. H. Kim, Y. Jung, S. D. Jeong, and H. Kim, “Profiling of signal sequence characteristics and requirement of different translocation components,” *Biochim. Biophys. Acta BBA - Mol. Cell Res.*, vol. 1865, no. 11, Part A, pp. 1640–1648, Nov. 2018.

[37] T. Hessa *et al.*, “Recognition of transmembrane helices by the endoplasmic reticulum translocon,” *Nature*, vol. 433, no. 7024, pp. 377–381, 2005.

[38] T. Hessa *et al.*, “Molecular code for transmembrane-helix recognition by the Sec61 translocon,” *Nature*, vol. 450, no. 7172, pp. 1026–1030, Dec. 2007.

[39] T. Junne, T. Schwede, V. Goder, and M. Spiess, “Mutations in the Sec61p Channel Affecting Signal Sequence Recognition and Membrane Protein Topology,” *J. Biol. Chem.*, vol. 282, no. 45, pp. 33201–33209, Nov. 2007.

[40] G. M. A. Forte, M. R. Pool, and C. J. Stirling, “N-Terminal Acetylation Inhibits Protein Targeting to the Endoplasmic Reticulum,” *PLOS Biol.*, vol. 9, no. 5, p. e1001073, 31 2011.

초록

출아효모의 막 단백질과 분비 단백질은 분비경로의 첫 단계로 소포체로 향하며 이는 전체 효모 단백질체의 약 30%를 자리한다. 따라서 소포체로의 전좌를 구성하는 세 단계; 세포질에서 합성된 단백질이 소포체로 근접하는 단계, 단백질의 아미노 말단이 전위효소 복합체 SEC과 상호작용을 시작하는 단계, 전좌가 본격적으로 진행되는 단계 각각을 이해할 필요가 있다.

이 연구는 전좌 극초기에 전좌효소와 전좌기질이 상호작용을 시작하는 과정에 초점을 맞추고 있다. 앞선 연구로부터 직진-역전 모델과 고리 모형 두 가지가 제시되어왔다. 따라서 이번 연구의 목적은 두 모델이 적합한지 확인하는 것이다. 이를 위해 각각의 모델을 따르는 기질을 분류하여 비교하고 있다. 모델 기질은 CPY의 아미노 말단의 길이, 신호서열 소수성을 변형하여 준비하였다.

이를 통해 Sec62에 의존적이고 SRP에 비의존적인 신호서열은 아미노 말단의 길이가 증가할수록 전좌가 느려지며, 양전하에 의해 전좌 속도가 회복되는 것을 관찰하였다. SRP 의존적인 신호서열의 전좌는 아미노 말단의 길이에 무관하였다. 그러나 아미노 말단의 양전하를 없애자 Sec62, Sec71/72에 의존적으로 변했으며, 아미노 말단의 길이에도 영향을 받았다. 따라서 번역후전좌의 기질은 아미노 말단 길이의 영향을 받기 쉬운 직진-역전 모델을 따르고, 번역중전좌의 기질은 전하분포에 따라 신호서열의 초기 방향이 결정된다고 해석하였다.

나아가 CPY 신호서열 뿐 아니라 다른 단백질의 신호서열에 대해서도 긴 아미노 말단과 신호서열 주변의 전하분포가 전좌에 미치는 영향이 보편적임을 보았다.

주요단어: 출아효모, 분비경로, 신호서열, 전좌, Sec71, Sec72