이학석사학위논문

출아효모 Sec61 에 의한 기질 분류

Modes of Substrate Discrimination by the Sec61 Translocon in *Saccharomyces cerevisiae*

2014 년 02 월

서울대학교 대학원

생명과학부

김지은

Modes of Substrate Discrimination by the Sec61 Translocon in *Saccharomyces cerevisiae*

by

Ji Eun Hani Kim

Under the supervision of

Professor Hyun Ah Kim, PhD

A Thesis submitted in partial Fulfillment

Of the Requirements for the Degree of

Master of Science

February, 2014

School of Biological Sciences,

Seoul National University,

Seoul 151-747, South Korea

CONTENTS

ABSTRACT

In eukaryotes, all secretory proteins contain a degenerate, hydrophobic stretch of sequence that targets them from the cytosol to the endoplasmic reticulum. This delivery step to the ER can occur in a translation-arrested or a fully translated nascent peptide state, which is determined in part by the hydrophobicity of the signal sequence. Hydrophobic signal sequences can act also as transmembrane segments, and are embedded into the ER membrane, while weakly hydrophobic signal sequences are generally cleaved off upon translocation into the ER. Irrespective of their targeting route to the ER, their routes merge at the Sec61 translocon. This pore-forming Sec61p facilitates a lateral exit of putative transmembrane segments through its proposed lateral gate (TM2a and TM7) or ER translocation of secretory proteins. As well as the gate, it contains a central constriction ring and a luminal plug domain. Sec61p has been shown to set the hydrophobicity threshold for incoming signal sequences and to discriminate authentic signal sequences over those that are not. The main aim of this study was to examine whether or not Sec61p (more specifically, different domains of Sec61p) recognizes different types of signal sequences distinctly. Firstly, in order to find substrates with signal sequences that are distinctly recognized by Sec61p, we 1) systematically defined the threshold N-terminal length and 2) the threshold signal sequence hydrophobicity for efficient ER translocation across the Sec61 translocon taking CPY as our model protein. We found that a short N-tail length is required for translocation of CPY with weakly hydrophobic signal sequences, while those with long N-tail length were only translocated when the signal sequence was hydrophobic. Using these two types of signal sequences that may be differentially recognized by the Sec61 translocon, their translocation were tested in Sec61 mutants of the plug domain and the lateral gate. We found that signal sequences with short N-tails and weakly hydrophobic signal sequences are more sensitive to changes in plug domain residues and mutations in TM7 of the lateral gate than those with long N-tails and hydrophobic signal sequences, thereby demonstrating differential substrate recognition by Sec61.

Key words: yeast, ER, CPY, Sec61, signal sequence, N-tail, translocation

LIST OF FIGURES AND TABLES

INTRODUCTION

I.1 Protein targeting and translocation in yeast *Saccharyomyces cerevisiae*

I.1.1 Secretory pathway

In eukaryotes, all secretory proteins contain a degenerate, hydrophobic stretch of sequence that targets them from the cytosol to the endoplasmic reticulum (ER) (1), the entry into which is thought to be the first step of protein trafficking along the secretory pathway. Over 30% of all translated polypeptides enter this path in yeast (2). In the ER, proteins undergo various post-translational modifications required for their maturation, such as signal sequence cleavage, glycosylation, disulfide-bond formation, and folding, after which they exit to their final destinations in the secretory pathway.

I.1.2 Modes of protein targeting to the ER

The ER signal sequence is N-terminally positioned on secretory proteins, and once exposed from the ribosome tunnel it is recognized by a signal recognition particle (SRP). If unrecognized by SRP, usually cytosolic chaperones (Hsp40/Hsp70 etc.) bind to the mature part of the nascent chain to prevent premature folding and aggregation. Either way, that nascent chain is targeted to the ER and this delivery step can occur in a translation-arrested or a fully translated nascent peptide state, which is determined in part by the hydrophobicity of the signal sequence. Hydrophobic signal sequences, which we will term signal anchor sequences, are quickly recognized by SRP as it emerges from the tunnel, and this interaction leads to temporary translation arrest. This arrest is maintained until the SRP-nascent chain-ribosome complex docks on the Sec61 translocon, the main pore-forming unit in the ER membrane, via engagement of SRP with its ER-bound SRP receptor. Subsequent release of signal anchor sequence from SRP allows continuation of nascent chain translation, which is appropriately situated to feed through the Sec61p pore (3; review paper).

In contrast to this co-translational translocation of a subset of proteins, those with weakly hydrophobic signal sequences are not effectively engaged by SRP, and thus translation proceeds unheedingly. The fully translated nascent chains are targeted to the Sec61 translocon

in translocation-competent states by cytosolic chaperones and translocation across the pore is ribosome independent, but dependent on the Sec62/63 complex. The translocated proteins are released from the pore are pulled through by an ATP-dependent ratcheting action of a luminal chaperone, Kar2p (Bip in mammals). The signal sequences are cleaved by the signal peptidase complex, and such cleavable signal sequences will be referred to as signal peptides.

I.1.3 General features of a signal sequence

Signal sequences comprise a tripartite structure; a hydrophobic core region (H-region), and an N- and C-region. The N-region is slightly arginine-rich, and the C-region contains a signal peptidase consensus cleavage site that is cleaved off during translocation. For membrane proteins that are targeted to the ER, usually it is the first transmembrane segment that acts as a signal sequence. Signal sequences vary extremely, both in terms of length and amino acid composition (Table I-1,4-6).

Table I-1

	Substrate type	
	Co-translationally targeted	**Post-translationally targeted**
Average length	~20 amino acids	
Hydrophobicity	Hydrophobic	Relatively non-hydrophobic
Cleavage site	No	Yes (cleavage site in the C-domain)
Membrane anchorage	Single-spanning and Multi-spanning	Mostly soluble proteins, C-Tail anchorage possible
Dependence on targeting/translocation components	SRP-dependent/ Trimeric Sec61 complex and Sec63	Hsp70-dependent/ Trimeric Sec61 complex and Sec62/Sec63 complex

Table I-1 General features of signal sequences A table briefly summarizing the characteristics of signal sequences that are targeted either co- or post-translationally to the ER (4-6).

I.1.4 Sec61 translocon

Whether or not a protein employs a co- or post-translational translocation route to the ER, their routes merge at the Sec61 translocon (Figure I-1). In yeast, the Sec61 translocon is a trimeric complex consisting of Sec61p, Ssh1, and Sss1p. Sec61p is a 10 TM-spanning pore-forming protein that facilitates a lateral exit of signal sequences or signal anchor segments through its proposed lateral gate (TM2a and TM7), thereby anchoring membrane-spanning proteins to the ER membrane. Sec61p also mediates translocation of proteins with cleavable signal sequences into the ER lumen. The crystal structure of SecY, a bacterial homologue of Sec61p, shows an hour-glass shaped tunnel with a central constriction ring and a luminal plug domain composed of TM2b that block the aqueous pore (7, review paper).

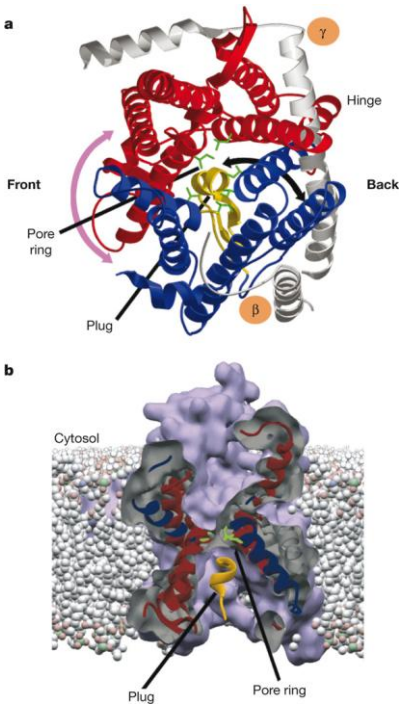


Figure I-1 Crystal structure of SecY A) The crystal structure of the SecY complex from *Methanococcus jannashii* viewed from the cytoplasmic side. The purple double-headed arrow represents how the lateral gate may open. The plug (yellow) is found blocking the aqueous pore. B) Cross-sectional view of the channel. (7)

I.2 Importance of the signal sequence in protein translocation

Studies have shown that although hydrophobicity of the signal sequence is a key determinant of pathway preference, information on translocation route also comes from the mature part of the protein (1,8). CPY, a bona fide substrate of the post-translational translocation has been shown to be targeted even when its signal sequence is deleted, and switching the signal sequence of preprocecropin A, another post-translationally translocated protein, with one that takes a co-translational translocation pathway did not diminish its capacity for post-translational translocation *in vitro* (8,9). Thus it seems that the role of signal sequence in post-translationally targeted proteins is not so significant at the stage of protein targeting to the ER, but rather their importance may lie at the stage of ER translocation.

Answering how exactly the initial engagement of signal sequences is made with the Sec61 pore and whether this engagement differs depending on the targeting route of the protein or hydrophobicity of the signal sequence are important to gain insight into the actions of the Sec61 translocon.

I.2.1 Signal sequences reside in the Sec61 pore at an early stage of translocation

Early studies have revealed the hydrophobic core region of the signal sequence and signal anchor (SA) sequences reside in the pore at an early stage of translocation (10). Not only are they inside the pore, a study by Martogolio et al. showed that they are exposed to the lipid bilayer (10).

I.2.2 Signal sequences intercalate between two TMs of Sec61p at the protein-lipid interface

Another cross-linking study using ppαF in an *in vitro* post-translational translocation system lacking Kar2p, showed the signal sequence of ppαF is cross-linked to distinct transmembrane domains of Sec61p at the early stage of translocation (11). They found the signal sequence of ppαF is cross-linked on opposite sides to either the TM2a or TM7 of Sec61p, and importantly,

that all residues in the signal sequence simultaneously cross-linked to lipids. This enables us to depict a signal sequence that is intercalated between the putative lateral gate of Sec61p at the lipid interface. Interestingly, the C-terminal portion of the helix was cross-linked to Sec62/71, a key component of the post-translational translocation machinery, which overlapped with portions of the helix that also cross-linked to TM7 of Sec61p. Their results suggest that initial signal sequence engagement at the Sec61p may require a lateral gate opened just enough to intercalate a helix, and the presence of other components of Sec61 complex such as Sec62 in close proximity to the gate, thus providing a distinctive environment for signal sequence recognition and subsequent translocation.

I.2.3 CPD A can differentiate between co- and post-translationally targeted proteins at the level of the Sec61 translocon

Recent studies have explored a differential engagement of signal sequences with the translocon. An *in vitro* translocation assay of substrates with signal sequences that mark them to the co- or post-translational translocation pathway, using a synthetic molecule called CPD A was performed (12). Translocation was exclusively inhibited for the substrates that undertake the co-translational translocation pathway. Conversely, substrates with signal sequences typical for the post-translational translocation pathway were unaffected by CPD A. When the signal sequence of the latter type of proteins was exchanged with a signal sequence that marks it to the co-translational translocation pathway, translocation was again sensitive to CPD A treatment. Since CPD A sensitivity differs between co- and post-translocated proteins, and thus distinction occurs at the level of the signal sequence, it indicates that CPD A action may likely occur at the level of the translocon. Their results, together with previous studies, imply that different types of signal sequences may engage the translocon distinctly.

I.3 Aims and experimental approach

Aims of this study were to 1) systematically determine the threshold N-terminal length for post-translationally translocated proteins and the threshold signal sequence hydrophobicity for

co-translationally translocated proteins for efficient ER translocation across the Sec61 translocon. Using two sets of CPY variants, varying in N-terminal length in front of the signal sequence and signal sequence hydrophobicity, we investigated 2) how these two types of signal sequences are recognized by the Sec61 translocon. Mutations were introduced in two domains of Sec61p, the plug domain and the lateral gate. Model constructs were tested in these Sec61 mutant strains to test our working hypothesis that distinct signal sequences are differentially engaged by the Sec61 translocon.

MATERIALS AND METHODS

M.1 Yeast strains

JRY4 strains were generated by plasmid shuffling and homologous recombination (Table M-1, 13). The pDQ1-Sec61WT vector encoding the *SEC61* gene was transformed into a *SEC61* deletion strain, RSY1293, to produce JRY3 (Table M-1, 14). JRY3 was transformed with pRS315 derivatives encoding mutant *sec61* alleles and the resulting transformants were FoA-selected (Table M-1, 15). The corresponding isogenic wild type strains were used for comparison in all cases.

Strain	Genotype
W303-1α / BWY46	*MATα, ade2, can1, his3, leu2, trp1, ura3*
BWY497	*MATa, sec62-1, ade2, can1, his3, leu2, trp1, ura3*
BWY500	*MATα, sec65-1, ade2, can1, his3, leu2, trp1, ura3*
JRY4	*MATα, sec62Δ::HIS3, ade2, can1, his3, leu2, trp1, ura3* [pRS416-Sec62-WT(URA)]
JRY4 Sec62 WT	*MATα, sec62Δ::HIS3, ade2, can1, his3, leu2, trp1, ura3* [pRS415-Sec62-WT(LEU)]
JRY4 Sec62 35DDD	*MATα, sec62Δ::HIS3, ade2, can1, his3, leu2, trp1, ura3* [pRS415-Sec62-35DDD(LEU)]
RSY1293 pH-Sec61-YCplac33	*MATα, ura3-1, leu2-3,112, his3-11,15, trp1-1, ade2-1, can1-100, sec61::HIS3* [pH6-Sec61-YCplac33(URA)]
JRY3 Sec61 WT	*MATα, ura3-1, leu2-3,112, his3-11,15, trp1-1, ade2-1, can1-100, sec61::HIS3* [pDQ1-Sec61WT(LEU)]
JRY3 Sec61 D61K	*MATα, ura3-1, leu2-3,112, his3-11,15, trp1-1, ade2-1, can1-100, sec61::HIS3* [pDQ1-Sec61D61K(LEU)]
JRY3 Sec61 L63A	*MATα, ura3-1, leu2-3,112, his3-11,15, trp1-1, ade2-1, can1-100, sec61::HIS3* [pDQ1-Sec61L63A(LEU)]
JRY3 Sec61 L63N	*MATα, ura3-1, leu2-3,112, his3-11,15, trp1-1, ade2-1, can1-100, sec61::HIS3* [pDQ1-Sec61L63N(LEU)]
JRY3 Sec61 R74E	*MATα, ura3-1, leu2-3,112, his3-11,15, trp1-1, ade2-1, can1-100, sec61::HIS3* [pDQ1-Sec61R74E(LEU)]
JRY3 Sec61 I91A	*MATα, ura3-1, leu2-3,112, his3-11,15, trp1-1, ade2-1, can1-100, sec61::HIS3* [pDQ1-Sec61I91A(LEU)]
JRY3 Sec61 Q96A	*MATα, ura3-1, leu2-3,112, his3-11,15, trp1-1, ade2-1, can1-100, sec61::HIS3* [pDQ1-Sec61Q96A(LEU)]
JRY3 Sec61 I293A	*MATα, ura3-1, leu2-3,112, his3-11,15, trp1-1, ade2-1, can1-100, sec61::HIS3* [pDQ1-Sec61I293A(LEU)]

Table M-1 Summary of yeast strains used in this study

M.2 Plasmid construction

Oligonucleotides encoding sequences against the open reading frame (ORF) of *PRC1* gene were used to amplify the gene from W3031α genomic DNA using PCR. The amplified product, now flanked with sequences complementary to the TRP marked plasmid

p424GPDHA, was subcloned into the *SmaI*-linearized p424GPDHA plasmid through homologous recombination. The C-terminal end of the subcloned gene was appended with a triple HA tag to produce p424CPYHA vector. Using this plasmid as a template, a *SmaI* restriction enzyme site was introduced between amino acid positions 2 and 3 of the ORF to produce p424CPYSma1HA vector. The N-terminal sequences of DPAPB and Sec71 were amplified from gDNA as for CPY and the PCR products were subcloned into the *SmaI*-linearized p424CPYSma1HA in a similar manner as before to produce p424DCPYHA and p424SCPYHA. All following constructs were henceforth created using these two plasmids as templates. Information for all of the oligonucleotides used in this study is available upon request to authors.

All PCR reactions were performed following the manufacturer's protocols (Toyobo, Japan). PCR products were confirmed by sizing on an agarose gel. For homologous recombination, the PCR products and the appropriate *SmaI*-linearized plasmids were transformed into W303-1α. Yeast transformants were selected on -TRP plates. Plasmids were isolated, and the DNA sequencing performed to confirm the sequence. Confirmed constructs were retransformed into BWY, JRY3 or JRY4 strains. The transformants were selected on either –TRP or –LEU–TRP plates, and subjected to further analysis.

M.3 Western blot analysis

Yeast transformants carrying either CPY or variants were grown in 5 ml of –LEU–TRP medium at 30℃ overnight, harvested by centrifugation at 3000 *g*, washed with dH_2O, resuspended in 100 μl SDS-PAGE sample buffer, heated for 15 min at 60℃, and centrifuged down. The supernatant fractions were loaded onto 8% SDS-PAGE gels and subsequently subjected to Western blotting using rabbit or mouse anti-HA antiserum (Covance, California, USA). For endoglycosidase H (Endo H) digestion, 15 μl of the whole-cell lysate was mixed

with 10.5 µl dH_2O, 3 µl Endo H buffer (800 mM sodium acetate, pH 5.8), and 1.5 µl Endo H (5 U/ml; Roche) or dH_2O for the mock treatment, and incubated at 37°C for 3 h.

M.4 Immunoprecipitation of radiolabelled proteins

Cells for pulse-labelling were either grown at 30°C till an OD_{600} (optical density at 600 nm) reached between 0.2-0.8 (JRY3 and JRY4) or at 23°C till an OD_{600} of 0.1-0.3, and switched to 37°C for an additional 4 h for protein expression in the temperature sensitive strains (BWY46, BWY497 and BWY500). Per reaction, 1.5 OD_{600} units of cells were harvested by centrifugation at 3000 *g*, washed twice with –MET medium without ammonium sulfate, and incubated at 30°C for 15 min or at 37°C for 30 min (methionine starvation). Cells were centrifuged and resuspended in 150 µl of –MET medium without ammonium sulfate and labelled with [^{35}S]-Met (50 µCi/1.5 OD_{600} units of cells) for 5 min at 25°C. Labelling was stopped by addition of 750 µl of ice-cold stop solution buffer (20 mM Tris-HCl, pH7.5, and 20 mM sodium azide). Cell pellets were harvested by centrifugation at 20,000 *g* and left at -20°C until further use. For pulse/chase experiments, radiolabelling was stopped and chased by addition of 50 µl of 200 mM cold MET per 1.5 OD_{600} units of cells for the indicated time periods. The reaction was stopped by addition of 750 µl of ice-cold stop solution buffer, centrifuged down, and cell pellets were kept frozen until use. Cell pellets were resuspended in 110 µl lysis buffer (20 mM Tris-HCl, pH7.5, 1% SDS, 1 mM DTT, 1 mM PMSF, protease inhibitor cocktail (Complete, Roche) and mixed with 100 µl of ice-cold acid-washed glass beads (Sigma). Cell suspensions were vortexed at maximum speed for 3 min. Then, samples were incubated at 65°C for 15 min and centrifuged for 5 min at 20,000 *g*. The supernatant fractions were mixed with 500 µl IP buffer (15 mM Tris-HCl, pH7.5, 0.1% SDS, 1% Triton X-100, and 150 mM NaCl), 1.3 µl anti-HA antibody, and 50 µl of prewashed protein G-agarose beads (Roche; 33% slurry in IP buffer) and rotated at 4°C overnight. The agarose beads were washed three times with IP buffer, once with ConA buffer (500 mM NaCl, 20 mM Tris-HCl, pH7.5, and 1% Triton X-100), and once with buffer C (50 mM NaCl and 10 mM Tris-HCl, pH7.5). Then the beads were incubated with 55 µl of SDS-PAGE sample

buffer at 65℃ for 15 min, centrifuged down, and the supernatant fractions were loaded onto SDS-PAGE gels. Endo H treatment was carried out as described above. Radiolabelled bands on SDS-PAGE gels were quantified using a Fuji FLA-3000 phosphoimager and the Image Reader V1.8J/Image Gauge V 3.45 software.

M.5 Bioinformatics analysis of yeast secretory and membrane proteins

The yeast signal peptide data set was downloaded from SWISS-PROT version 29 (16). This directory contains amino acid sequences of secretory proteins from *Saccharomyces cerevisiae*, which were initially filtered from the entire SWISS-PROT database using the keyword 'SIGNAL' in the feature table. Several other filtering options were implemented as described on SignalP 3.0 Server (17). The sequences obtained were run on SignalP, a prediction program for signal sequences, to extract the position of the signal sequence in the protein and their cleavage sites. Those predicted to be a non-secretory protein or any duplicates were discarded from the list. The *dG* values of the predicted signal sequences were predicted using the dG prediction program (18). The sequences of a total of 124 membrane proteins were extracted from the UniProtKB Protein Knowledgebase (19) by querying "signal anchor" AND organism:"Saccharomyces cerevisiae", and from the MIPS Comprehensive Yeast Genome Database (CYGD) (20) by querying organism:"Saccharomyces cerevisiae" AND annotation:(type:location "endoplasmic reticulum") AND "membrane protein". Any duplicates were removed manually. The position of the first transmembrane domain and its *dG* value were predicted using the dG prediction program.

M.6 Statistical analysis

All statistical analysis was performed using SPSS v.21 (21). All graph and box plots were created using this program.

RESULTS

R.1 Signal peptide proteins have short N-terminal lengths

The Signal Peptide data set comprising coding sequences of proteins predicted to contain cleavable signal sequences, *i.e.* those that take the post-translational translocation pathway, in *S. cerevisiae* was downloaded from the Signal Peptide database (SPdb v4.1) (17). Of the 309 protein sequences retrieved, the data set was filtered to a total of 232 protein sequences as described in Materials and Methods. Using a prediction program, the position of the predicted signal sequence was determined and the total number of residues preceding the signal sequence *i.e.* the N-terminal length (N-tail) was counted (17). The apparent free energy difference (ΔG_{app}), which describes how favourable or unfavourable a peptide is to be in a nonpolar environment, indirectly shows how much energy is required for this given segment to insert into the ER membrane. The lower the ΔG_{app} value (a negative value), the more hydrophobic a segment is, thus it is easier to cross a lipid bilayer, whereas the higher the ΔG_{app} value (a positive value), thus it is harder to cross a lipid bilayer. This was calculated using the Delta G predictor (18). When the ΔG_{app} value was plotted against N-terminal length, around 96% of proteins (223 out of 232) fell below 10 residues (Figure 1A). To test whether this apparent length threshold was an observation specific for secretory proteins, sequences of 124 membrane proteins localized to the ER in the *S. cerevisiae* was downloaded from UniProtKB and CYGD. Their N-terminal lengths in front of the first predicted TM segment was determined in the same manner as proteins with a cleavable signal peptide. A box plot and an independent-samples *t*-test was conducted to compare the average absolute number of amino acids in front of the predicted signal peptide, and transmembrane segments of membrane proteins, denoted as “N-terminal length” (21). The box plot shows that signal peptide proteins tend to have a significantly shorter N-terminal length (M=3, SD=6) than membrane proteins (M=79, SD=132), conditions; t(123)=-6.417, p = 0.000 (Figure 1B). Total length of proteins varies greatly between proteins, and this was not taken into consideration when comparing the absolute N-terminal lengths. However, even when the relative position of the signal peptide and TMD within the whole protein was calculated, the former showed a statistically significant shorter N-terminal length compared to membrane proteins (Figure 1C;

p=0.000). Altogether, our bioinformatics analysis suggests that signal peptide proteins have N-terminal lengths significantly shorter than those of membrane proteins.

Figure 1

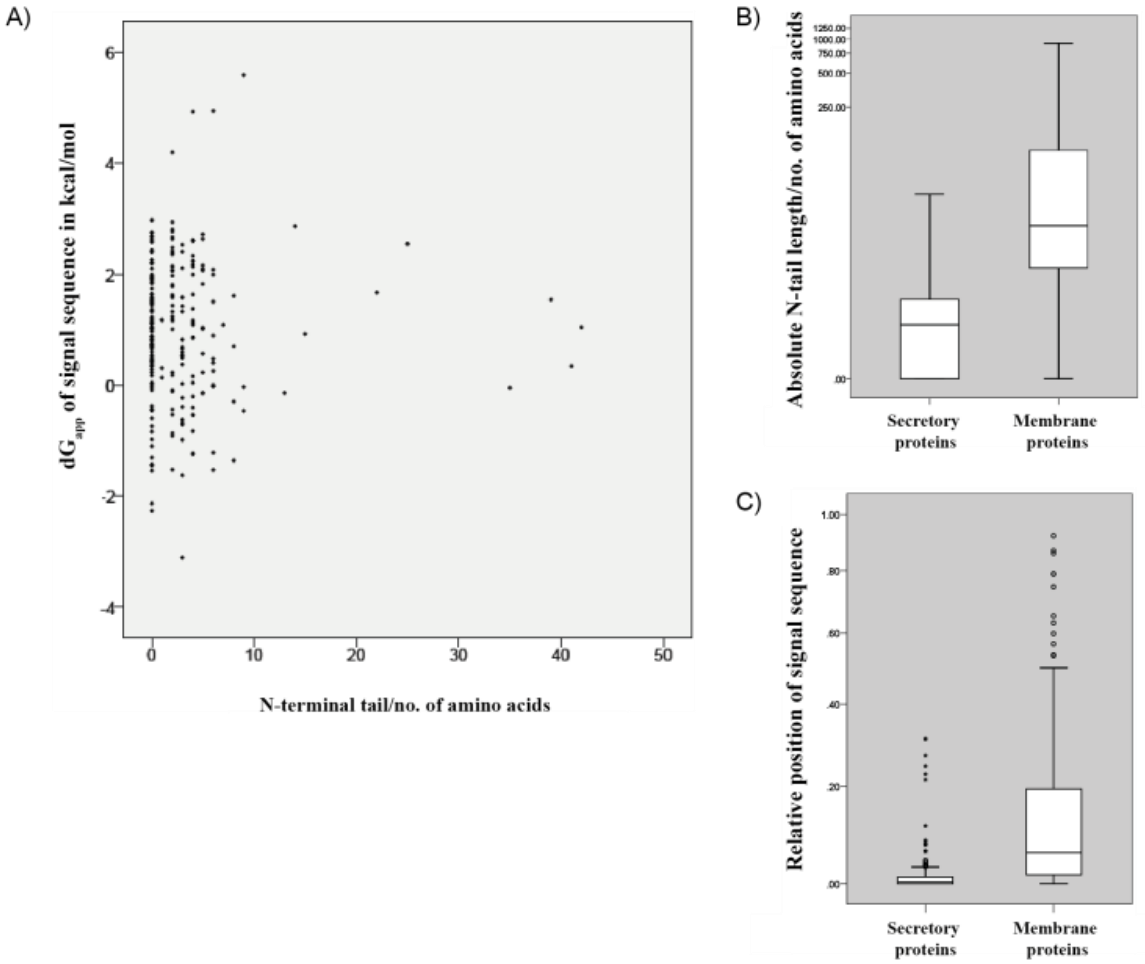


Figure 1 A) The N-terminal length against dG_{app} of 232 signal sequences of secretory proteins in yeast. B) Box plot of the absolute N-terminal lengths in front of the signal sequence and first transmembrane domain in secretory (n=232) and membrane proteins (n=132), respectively. C) Box plot of the relative positions of the signal sequence or the first transmembrane domains within the whole protein in secretory and membrane proteins, respectively.

R.2.1 N-terminal extensions inhibit translocation of CPY

To test the biological significance of the short N-terminal lengths of signal peptide proteins, we took a previously well characterized signal peptide protein with an archetypal cleavable signal sequence, carboxypeptidase Y (CPY) (1,22-23). Two types of extensions, derived from the N-terminal sequences of two different ER-resident signal anchor proteins, DPAPB and Sec71, were placed upstream of the signal sequence of CPY (Table 1). These extensions were gradually truncated, so that two different sets of CPY variants with varying lengths of N-terminal extensions were prepared (Figure 2A). When the translocation efficiencies of these constructs in the wild type (WT) strain were assessed, translocation of CPY was inhibited for the long extensions in both sets. Interestingly, the inhibitory effect of the N-terminal extensions on CPY translocation was not observed for short N-terminal length extensions (~10-11 amino acids long), and any extensions below this length was efficiently translocated. Despite the fact that the signal sequence and the mature part of the CPY are identical, this apparent N-terminal length threshold on CPY translocation was almost an all-or-none effect for the DPAPB extension set, but more gradual for the Sec71 set (Figure 2B). For example, the translocation of D(10)CPY was complete, whereas translocation of D(11)CPY, whose length is longer by a single residue, was completely inhibited. Similarly translocation is substantially blocked for S(length)CPY variants with an N-terminal extension of up to 14 residues. Efficiency of translocation is improved in S(10)CPY, but even at a shorter N-terminal extension such as S(7)CPY, translocation is still inhibited to an extent. So far, our data show translocation of CPY with an N-terminal extension of greater than ~10-11 amino acids is inhibited, in part, explains why the majority of signal peptide proteins contain less than 10 residues in the upstream of the signal peptide (Figure 1A). .

R.2.2 Length dependent inhibition of CPY translocation

To investigate in greater depth the effect of N-terminal extensions on CPY translocation, we performed pulse-chase labelling of selected variants. Translocation of CPY is complete after 5 min of radiolabelling (Figure 2C). However, for the subset of short N-terminal extensions that

were shown to be translocated efficiently when detected by Western blotting of whole-cell lysates, pulse-chase labelling revealed that their translocation is completed over time (Figure 2C and 2D). Interestingly, we found that the steepness of the translocation curve correlated to the length of the extensions (Figure 2D). Thus, our data again show that N-terminal length contribute to the translocation defect.

Table 1

Name	N-Length[a]	N-terminal sequence	Signal sequence	ΔG_{app} [b]
CPY	0		MKAFTSLLCGLGLSTTLAKAISL	1.471
D(26)CPY	26	MEGGEEEVERIPDELFDTKKKHLLDK	KAFSSLLCGLGLSTTLAKAISL	1.933
D(21)CPY	21	MEVERIPDELFDTKKKHLLDK	KAFSSLLCGLGLSTTLAKAISL	1.933
D(16)CPY	16	MPDELFDTKKKHLLDK	KAFSSLLCGLGLSTTLAKAISL	1.933
D(11)CPY	11	MDTKKKHLLDK	KAFSSLLCGLGLSTTLAKAISL	1.933
D(10)CPY	10	MTKKKHLLDK	KAFSSLLCGLGLSTTLAKAISL	1.933
D(9)CPY	9	MKKKHLLDK	KAFSSLLCGLGLSTTLAKAISL	1.933
D(8)CPY	8	MKKHLLDK	KAFSSLLCGLGLSTTLAKAISL	1.933
S(29)CPY	29	MSEFNETKFSNNSTFFETEEPIVETKSIS	KAFTSLLCGLGLSTTLAKAISL	1.471
S(26)CPY	26	MNETKFSNNSTFFETEEPIVETKSIS	KAFTSLLCGLGLSTTLAKAISL	1.471
S(19)CPY	19	MNSTFFETEEPIVETKSIS	KAFTSLLCGLGLSTTLAKAISL	1.471
S(14)CPY	14	METEEPIVETKSIS	KAFTSLLCGLGLSTTLAKAISL	1.471
S(11)CPY	11	MEPIVETKSIS	KAFTSLLCGLGLSTTLAKAISL	1.471
S(10)CPY	10	MPIVETKSIS	KAFTSLLCGLGLSTTLAKAISL	1.471
S(8)CPY	8	MPIVESIS	KAFTSLLCGLGLSTTLAKAISL	1.471
S(7)CPY	7	METKSIS	KAFTSLLCGLGLSTTLAKAISL	1.471
D(14,K)CPY	14	MKKKKTKKKHLLDK	KAFSSLLCGLGLSTTLAKAISL	1.933
D(14,D)CPY	14	MDKKKTKKKHLLDK	KAFSSLLCGLGLSTTLAKAISL	1.933
D(12,K)CPY	12	MKKTKKKHLLDK	KAFSSLLCGLGLSTTLAKAISL	1.933
D(11,A)CPY	11	MATKKKHLLDK	KAFSSLLCGLGLSTTLAKAISL	1.933
D(11,K)CPY	11	MKTKKKHLLDK	KAFSSLLCGLGLSTTLAKAISL	1.933
S(10,EE)CPY	10	MPIVEEESIS	KAFTSLLCGLGLSTTLAKAISL	1.471
D(1.66)CPY	26	MEGGEEEVERIPDELFDTKKKHLLDK	KAFSSLLC**A**LGLSTTLAKAISL	1.659
D(1.25)CPY	26	MEGGEEEVERIPDELFDTKKKHLLDK	KAFSSLLC**L**LGLSTTLAKAISL	1.253
D(0.51)CPY	26	MEGGEEEVERIPDELFDTKKKHLLDK	KAFSSLLCAL**L**LSTTLAKAISL	0.510
D(0.10)CPY	26	MEGGEEEVERIPDELFDTKKKHLLDK	KAFSSLLC**L**L**L**LSTTLAKAISL	0.098
D(-0.56)CPY	26	MEGGEEEVERIPDELFDTKKKHLLDK	**L**AFSSLLC**L**L**L**LSTTLAKAISL	-0.562
D(-2.15)CPY	26	MEGGEEEVERIPDELFDTKKKHLLDK	**LLLTL**LLC**L**L**L**LSTTLAKAISL	-2.145
S(1.66)CPY	29	MSEFNETKFSNNSTFFETEEPIVETKSIS	KAF**S**SLLC**A**LGLSTTLAKAISL	1.659
S(1.25)CPY	29	MSEFNETKFSNNSTFFETEEPIVETKSIS	KAF**S**SLLC**L**LGLSTTLAKAISL	1.253
S(0.51)CPY	29	MSEFNETKFSNNSTFFETEEPIVETKSIS	KAF**S**SLLCAL**L**LSTTLAKAISL	0.510
S(0.10)CPY	29	MSEFNETKFSNNSTFFETEEPIVETKSIS	KAF**S**SLLC**L**L**L**LSTTLAKAISL	0.098
S(-0.56)CPY	29	MSEFNETKFSNNSTFFETEEPIVETKSIS	**L**AF**S**SLLC**L**L**L**LSTTLAKAISL	-0.562
S(-2.15)CPY	29	MSEFNETKFSNNSTFFETEEPIVETKSIS	**LLL**T**L**LLC**L**L**L**LSTTLAKAISL	-2.145

a N-length, a count of the number of amino acids prior to the signal sequence.

b ΔGapp values in kcal/mol were predicted using the ΔG predictor.

c Residues altered with respect to the signal sequences of D(26)CPY and S(29)CPY are marked in bold.

Table 1 Summary of constructs used in this study

Figure 2

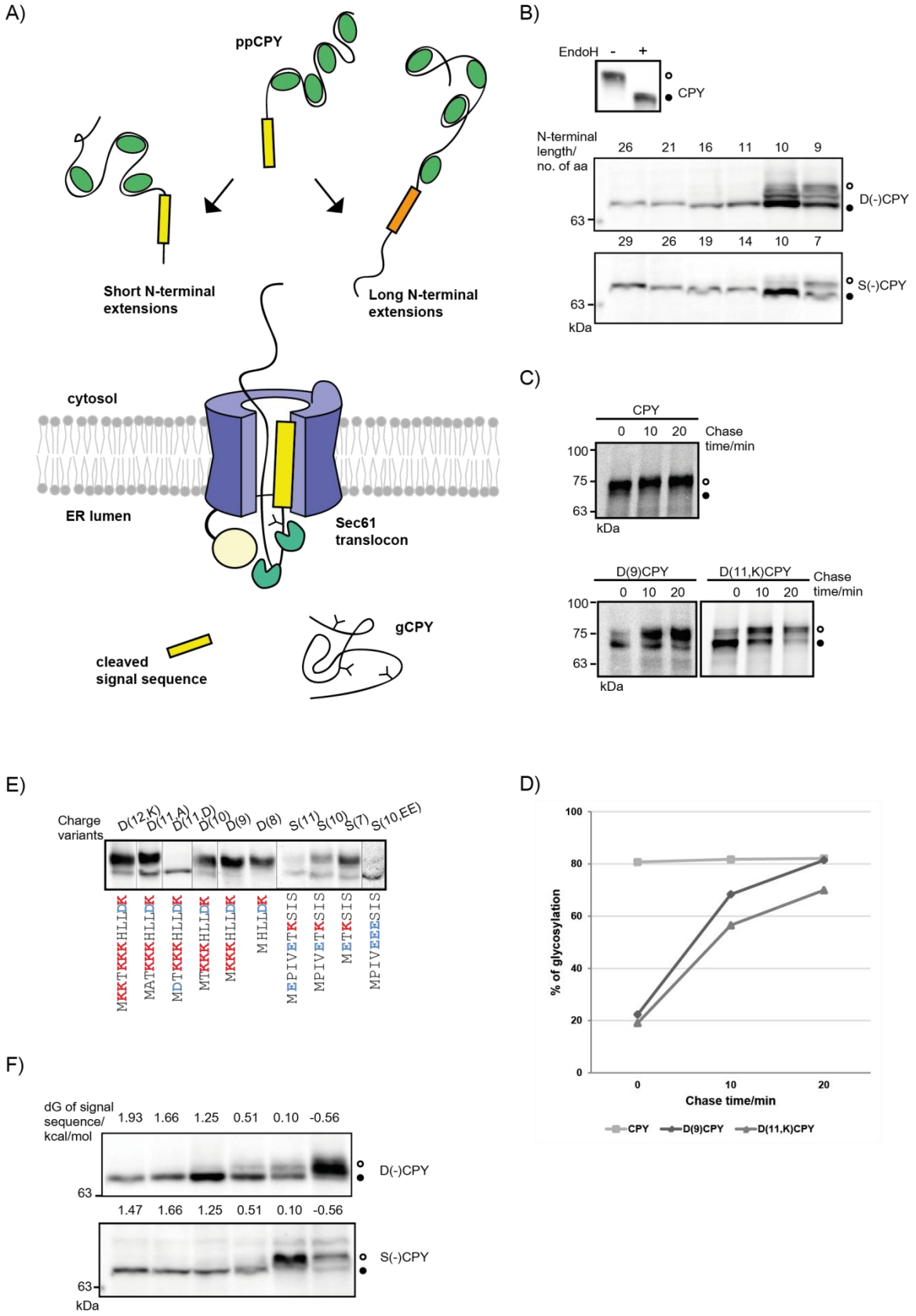

A)
ppCPY
Short N-terminal extensions
Long N-terminal extensions
cytosol
ER lumen
Sec61 translocon
gCPY
cleaved signal sequence
B)
EndoH - +
CPY
N-terminal length/ no. of aa
26 21 16 11 10 9
D(-)CPY
63
29 26 19 14 10 7
S(-)CPY
63
kDa
C)
CPY
0 10 20
Chase time/min
100
75
63
kDa
D(9)CPY
D(11,K)CPY
0 10 20 0 10 20
Chase time/min
100
75
63
kDa
E)
Charge variants
D(12,K) D(11,A) D(11,D) D(10) D(9) D(8) S(11) S(10) S(7) S(10,EE)
MKKTKKKHLLDK
MATKKKHLLDK
MDTKKKHLLDK
MTKKKHLLDK
MKKKHLLDK
MHLLDK
MEPIVETKSIS
MPIVETKSIS
METKSIS
MPIVEEESIS
D)
% of glycosylation
100
80
60
40
20
0
0 10 20
Chase time/min
CPY
D(9)CPY
D(11,K)CPY
F)
dG of signal sequence/ kcal/mol
1.93 1.66 1.25 0.51 0.10 -0.56
D(-)CPY
63
1.47 1.66 1.25 0.51 0.10 -0.56
S(-)CPY
63
kDa

Figure 2 A) Schematic drawing of the post-translational translocation of CPY across the Sec61 translocon. Two sets of CPY variants are depicted; on the left is a representation of a CPY variant with a short N-tail in front of a polar signal sequence (the same as or comparable to the signal sequence of CPY; yellow rectangle), on the right is a representation of a CPY variant with a long N-tail with a more hydrophobic signal sequence (orange rectangle) than CPY. All CPY variants are tagged with a triple HA tag at the C-terminus (not shown). Green ovals depict Hsp40/Hsp70 chaperones, and ratcheting actions of luminal chaperone Kar2 on translocating substrates is also depicted at the opening region of the luminal pore. Addition of glycans is indicated with a 'Y'. The channel-forming unit of Sec61 translocon is shown, and its plug domain is depicted as a white circle. B) JRY3 Sec61WT strains were grown in selective media and transformed with the plasmids encoding the indicated substrates. Transformants were grown in 5 ml of selective media overnight at 30℃, harvested, sampled using SDS sample buffer, heat-incubated, and loaded on 8% SDS-gels for electrophoresis. Gels were transferred to a nitrocellulose membrane, and detected by mouse or rabbit anti-HA antibody. If indicated, samples were treated with 1.5µl of Endoglycosidase H (EndoH) or mock. Glycosylation status of substrates is indicated with either open (glycosylated) or closed (not glycosylated) circles. C) Pulse-chase labelling for the indicated time-points was carried out after 1.5OD units of cells were harvested and 5 min pulse-labelled with [^{35}S]-Met. After cell lysis, radiolabelled proteins were immunoprecipitated using mouse anti-HA antibody and detected by autoradiography. D) The percentage of glycosylation form was calculated by simply taking the ratio of glycosylated proteins versus non-glycosylated proteins, and plotted against time. E) and F) Translocation assay by Western blotting was carried out as described in B) for the indicated substrates.

R.2.3 N-terminal charged residues modulate length threshold for signal peptide translocation

We observed that the extent of translocation defect varies depending on the N-terminal charged residues among the same length variants. When we compared the sequence contexts between those that were able to translocate, and those that could not, positively charged residues over negative ones were present in the former, and vice versa in the latter. Figure 2B shows translocation of D(10)CPY (extension contains 4 lysines and 1 aspartate; net charge of +3) was efficient, whereas translocation is markedly lower for S(10)CPY (extension contains 1 lysines and 1 glutamate; net charge of 0) despite having the same N-terminal length. Intriguingly, translocation of D(11)CPY (extension contains 4 lysines and 2 aspartate; net charge of +2) was inhibited substantially compared to that of D(10)CPY, indicating that minor changes in net charge have large impact on translocation efficiency. To further test the effect of charged residues on translocation, we cloned more charge variants of CPY (Table 1). Translocation of D(11)CPY was recovered to 70%~78% from 0% translocation when aspartate was replaced by an alanine [D(11,A)CPY] or a lysine [D(11,K)CPY], indicating that the presence of a negatively charged residue within the N-terminal extension is unfavourable for translocation (Figure 2E). Further, addition of positively charged residues in the N-terminus lowered the length threshold, allowing longer extensions (up to 14 residue extensions) to be translocated [D(12,K)CPY and D(14,K)CPY]. Likewise, increasing the number of negatively charged residues, from a net charge of 0 to -3, inhibited translocation of a previously translocation-competent protein [S(10)CPY and S(10,EE)CPY, respectively] (Figure 2E). Thus, our data indicate that not only is the N-terminal length important for translocation, but also the position of charged residues in the N-terminus.

R.2.4 Inhibition of CPY translocation by long N-terminal extensions is overcome by increasing signal sequence hydrophobicity

Our data indicate that having an N-terminal extension of longer than ~11-14 residues and one that contains negatively charged residues is inhibitory for translocation of CPY. Since the

data from Figure 1 show membrane proteins contain longer N-terminal lengths, we wanted to first, confirm whether increasing the hydrophobicity of the signal sequence of the longest extensions would restore translocation, and second, to determine the hydrophobicity threshold for longer N-terminal extensions. Systematic increase in signal sequence hydrophobicity of D(26)CPY and S(29)CPY led to a recovery in translocation of these proteins (Table 1, Figure 2F). Complete restoration of translocation was seen when the dG value for hydrophobicity reached -0.56 and 0.10, respectively. Therefore, inhibition of translocation brought about by long N-terminal extensions can be overcome by a more hydrophobic signal sequence. From our data, we can conclude that for proteins with less hydrophobic signal sequences, their N-terminal length should be short for efficient translocation, whereas if the N-terminus is long, a sufficiently hydrophobic signal sequence is required for translocation. Thus, we can reason based on our data that there may be differential signal sequence recognition by the Sec61 translocon.

R.3.1 Translocation of short N-extension variants follows the positive-inside-rule accommodated by the local net charge of the pore of Sec61p

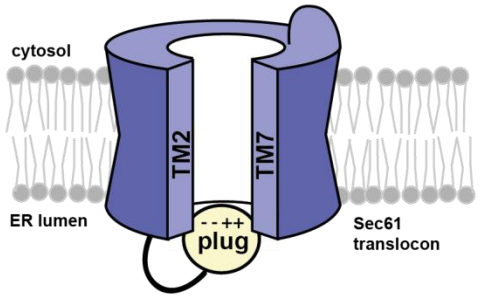
The above data imply that Sec61p adopts different mechanisms by which it recognizes signal sequences of two distinct CPY derivatives; short extensions with relatively polar signal sequences and long extensions with relatively hydrophobic signal sequences. We then hypothesized that distinct domains of Sec61p, such as the plug domain or lateral gate, may uniquely contribute to the recognition of signal sequences. To test this hypothesis, we investigated how the two sets of CPY variants are translocated in the two groups of Sec61 mutants; the plug domain mutants and the lateral gate mutants (15) (Figure 3A).

The plug domain of Sec61p (residues 52-74) is known to contribute to signal sequence orientation. It also blocks access of aqueous pore content into the ER lumen. Charged residues (R, K, D, and E) are not found in any of the 10 TM domains of Sec61p. Only charged residues of Sec61p a signal sequence may interact when it gets inside the pore are the

residues in the plug domain. Thus, we hypothesized that since translocation of extension variants of CPY are influenced profoundly by charged residues (Figure 2E), the local charge of the Sec61 pore may influence translocation.

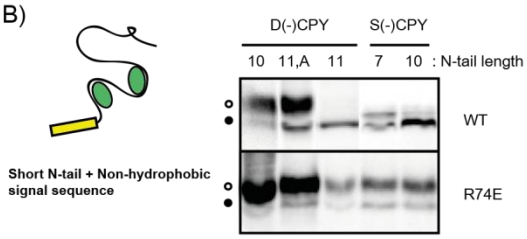
Figure 3

A)

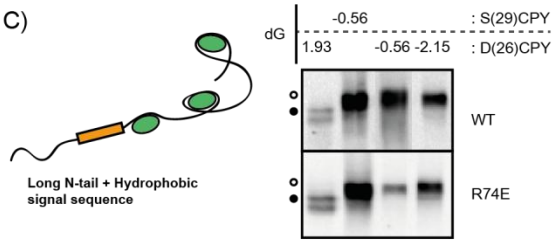


Sec61 domain	Name	Sec61 sequence	Net charge
Plug	WT	$_{52}$GIVSSETSDPLYWLRAMLASNR$_{74}$	0
	D61K	$_{52}$GIVSSETS**K**PLYWLRAMLASNR$_{74}$	+2
	L63A	$_{52}$GIVSSETSDP**A**YWLRAMLASNR$_{74}$	0
	L63N	$_{52}$GIVSSETSDP**N**YWLRAMLASNR$_{74}$	0
	R74E	$_{52}$GIVSSETSDPLYWLRAMLASN**E**$_{74}$	-2
Lateral gate (TM2)	WT	$_{80}$LGVSPIITSSMIFQFLQ$_{96}$	-
	I91A	$_{80}$LGVSPIITSSM**A**FQFLQ$_{96}$	-
	Q96A	$_{80}$LGVSPIITSSMIFQFL**A**$_{96}$	-
Lateral gate (TM7)	WT	$_{293}$IMLQSALTSNIFLISQIL$_{400}$	-
	I293A	$_{293}$**A**MLQSALTSNIFLISQIL$_{400}$	-

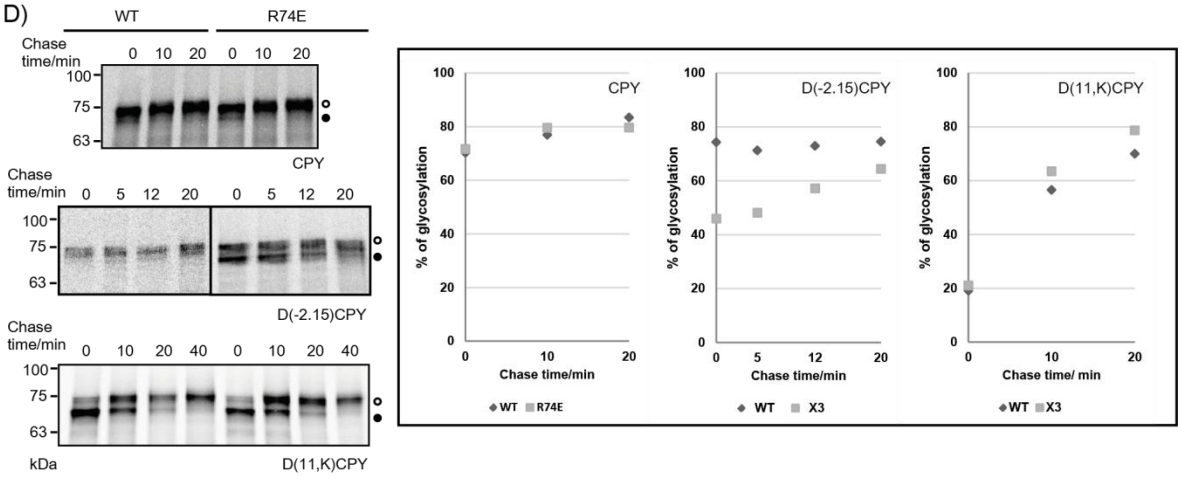
B)



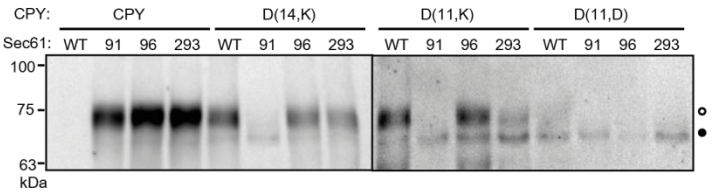
C)



D)



E)



F)

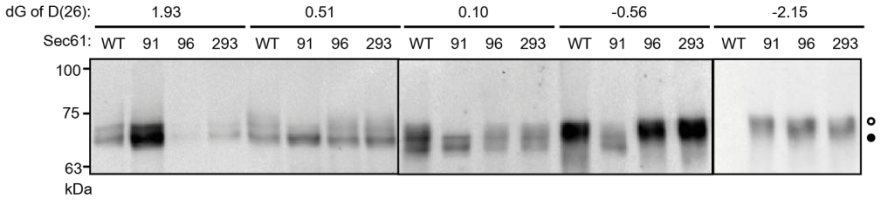


Figure 3 A) Left; a schematic drawing of Sec61p showing the two transmembrane domains (TM2a and TM7) implicated as the lateral gate, and the plug domain (white circle). Right; a table showing the Sec61 mutants, and their sequences, used in this study. The net charge (Δ no. of positive - no. of negative residues) of the plug domain is given. B) and C) Western blot analysis of the indicated CPY variants in WT or R74E strains, as described in Figure 2B. D) Pulse-chase analysis (left) of CPY (top panel), D(-2.15)CPY (middle panel), and D(11,K)CPY (bottom panel) as described in Figure 2C and quantitation of the data (right) as described in Figure 2D. E) and F) Western blot analysis of short N-tail variants (top panel) and long N-tail variants with varying signal sequence hydrophobicity (bottom panel) in WT and Sec61 lateral gate mutant (I91A, Q96A, I293A) strains. The procedure was carried out as written in Figure 2B.

In Sec61 WT, the net charge of the plug domain is zero. We expressed our CPY variants in a previously characterized 'R74E' mutant of Sec61p which has a net charge of -2 compared to the WT (24) (Figure 3A). In this mutant, CPY derivatives with short N-tail that were not translocated in the WT were translocated efficiently when analysed by Western blotting (Figure 3B, e.g. D(11)CPY, S(7 and 10)CPY) . However, those that were already efficiently tranaslocated in WT (e.g. D(10)CPY, D(11,A)CPY) were not influenced by the R74E mutation. Our data, not only indirectly show that a subset of the short N-tail variants is reliant on the plug domain for translocation, but also that the translocation defect initially seen in the WT strain is not due to untargeting but at least in part by mis-handling of the substrates by the Sec61 translocon.

Interestingly, whole-cell lysate preparation and Western blotting of the long N-extension derivatives showed that there is no effect of the R74E mutation on translocation (Figure 3C). Translocation of D(26)CPY was still inhibited by extension of the long N-terminal tail in R74E. However, when pulse-chase experiments were performed, D(-2.15)CPY, one with very hydrophobic signal sequence, was shown to translocate slowly in R74E, and gradual increase in glycosylation was seen over time compared to a complete glycosylation at 0 min chase in WT (Figure 3D). This is in comparison to what is seen when translocation of other substrates was assessed over the same time-points. If the slowing down of translocation of D(-2.15)CPY is a non-specific defect of the R74E mutation, the same should be seen for WT CPY or the short N-tail variants. However, there is not seen, as translocation pattern of these two substrates in R74E is the same as WT strain (Figure 3D). In sum, changing the net charge of the plug domain to -2, improves the translocation of only a subset of short N-tail variants, whereas translocation of long N-tail variants is minimal.

Similar experiments will be carried out using D61K, L63A, L63N strains, which have net plug domain charges of +2, 0, and 0, respectively. This is to test our working hypothesis that our two sets of CPY variants (short N-tail with polar signal sequence, long N-tail with

hydrophobic signal sequence) will show distinctive phenotype in different Sec61p mutants. Those with short N-tails with polar signal sequences may be influenced by mutations in plug domain residues in accordance to the positive-inside rule, i.e. the positive ends tend to stay in the cytosol (assuming that the Sec61p plug domain residues contribute to this). Those with long N-tails with hydrophobic signal sequences may be less sensitive to changes in the plug domain.

R.3.2 Translocation of both sets of CPY variants is inhibited in a 'closed' Sec61 lateral gate, whereas only CPY variant with short N-tail length is inhibited in an 'open' gate

Translocation of CPY derivatives with long N-terminal extensions was slower irrespective of the net charge of the plug domain (Figure 3D). This suggests that mutations in the plug domain may indirectly compromise structural integrity of the overall Sec61 structure, especially the lateral gate in an unknown way. This further tells that whilst translocation of CPY derivatives with short N-tails are more sensitive to mutations in the plug domain, CPY derivatives with long N-tails may be more sensitive to subtle changes in the lateral gate. Thus, to test this hypothesis, we tested the translocation the two sets of CPY derivatives in different lateral gate (TM2a and TM7) mutants (Figure 3A). Two types of lateral gate mutations (open and closed conformation) that have been previously characterized in our group (15) were examined. 'I91A' in TM2a has been biochemically proposed to induce a 'closed' conformation of the lateral gate, whereas 'Q96A' and 'I293A' in TM2a and TM7, respectively, have been characterized to induce an 'open' conformation (15, Figure 3A).

Our data show that whether or not our CPY variant has a short or long N-tail, its translocation is blocked in the 'closed, I91A' conformation mutant. For example, translocation of D(14,K)CPY, D(14,K)CPY, and D(-0.56)CPY is almost completely inhibited in I91A (Figure 3E and F). This indicates that both types of CPY variants require a certain degree of lateral gate opening. However, translocation of CPY and CPY variants with the most hydrophobic signal sequence was unperturbed in I91A (Figure 3E and F), suggesting these two substrates

may have enough translocative power to open the lateral gate. Next, the same constructs were expressed in Sec61p mutant strains of TM2a and TM7 that have been proposed to open the lateral gate. Protein sampling of whole-cell lysate and subsequent Western blotting showed that translocation of CPY variants with long N-tails constructs and hydrophobic signal sequences, i.e. those that are most likely to exit the lateral gate into the lipid bilayer, were not prohibited by open conformation inducing mutants (Figure 3F). Interestingly, translocation of D(11,K)CPY was inhibited specifically in a mutation in TM7, I293A (Figure 3E), despite the fact that both mutations (Q96A and I293A) have been previously shown to have a similar translocation defect of a Lep model protein (15). Thus, this implies that specific motifs may exist at the lateral gate for recognition of signal peptides with short N-tails.

R.4 Translocation of short N-terminal extensions and marginally hydrophobic long N-terminal length CPY variants are reduced in a sec62p mutant

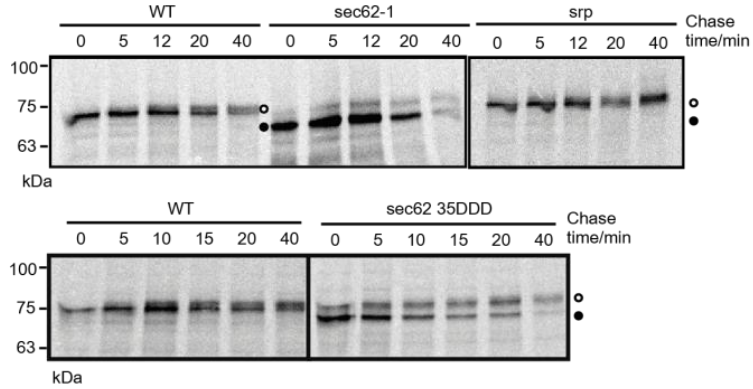
Many studies in yeast have implicated Sec62p in translocation of substrates with cleavable signal sequences. Past studies have shown hydrophobicity of the signal sequence is the main discriminatory factor for Sec62p. CPY is a bon fida substrate of the Sec62-dependent pathway. CPY was expressed in *sec62-1* and *sec62 35DDD* strains, which are both defective in Sec62 function (1,13,22). Pulse-chase labelling of CPY in these sec62 mutant strains and their isogenic WT strains revealed that in WT (isogenic WT; BWY46 and JRY3 Sec62WT, respectively), CPY translocation was complete after 5 min labelling, where 90% and 78% of CPY were already translocated by this time-point (Figure 4A-C). Conversely, in *sec62-1*, even after 20 min chase, only 20% of CPY was glycosylated, and at 40 min chase this percentage doubled (Figure 4B). The Sec62 mutation was less pronounced in *sec62 35DDD*, but nevertheless, CPY translocation occurred more slowly as only after 40 min chase did the percentage of CPY glycosylation became comparable to the percentage seen at 5 min for the WT (Figure 4C). As a control, CPY was also tested in *srp*, a strain defective in SRP function. Since CPY is a strictly post-translationally translocated protein, their translocation was unaffected in this strain (Figure 4A-B).

When our two types of variants, i.e. those containing the WT signal sequence of CPY and a short N-terminal length, and another containing more hydrophobic signal sequences with longer N-tail, were expressed in *sec62 35DDD* their dependence on Sec62p for translocation differed. For the short N-tail group, translocation was no longer supported even when the steady-state level was analysed by Western blotting, consistent with the Sec62-dependent CPY translocation seen in Figure 4A (Figure 4D). However, translocation of D(-2.15)CPY, with a signal sequence that is very hydrophobic, was independent of a functional Sec62p (Figure 4E). Interestingly, when a CPY variant had an “in-between” characteristics, i.e. one that contains an N-tail length that exceeds that threshold determined in Figure 1 and 2, and a signal sequence that is still fairly non-hydrophobic (e.g. D(0.51)CPY and D(0.10)CPY), a

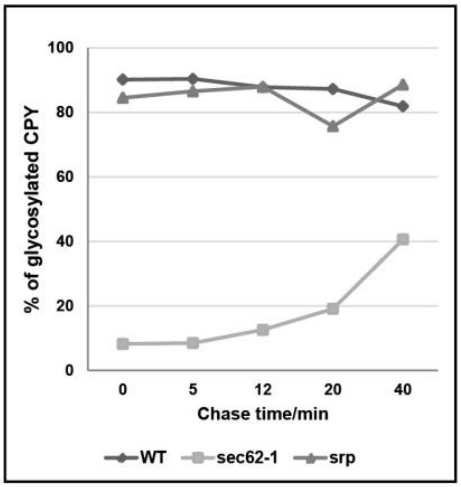
dependence on Sec62p for translocation was seen that was correlated to hydrophobicity. A much greater proportion of D(0.51)CPY (contains 7 leucines) failed to be translocated in *sec62 35DDD* compared to WT than D(0.10)CPY (contains 8 leucines) (Figure 4E). Thus, despite their long N-terminal length, which is uncharacteristic of secretory proteins that are recognized by Sec62p, translocation of substrates, especially D(0.51)CPY, are dependent on Sec62p.

Figure 4

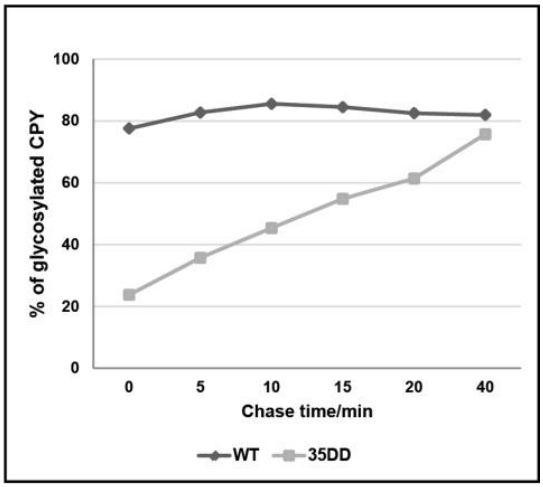
A)



B)

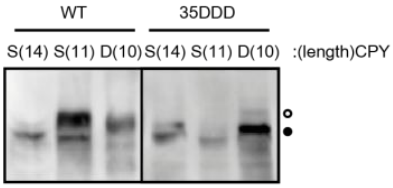


C)

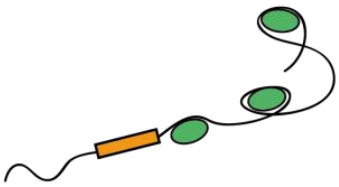


D)

Short N-tail + Non-hydrophobic signal sequence



E)



Long N-tail + Hydrophobic signal sequence

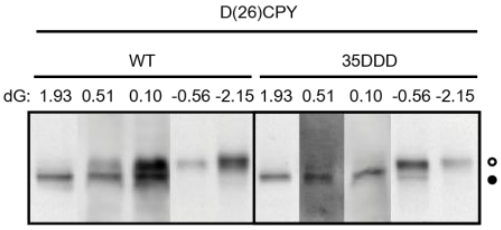


Figure 4 A) Pulse-chase experiment was carried out as in Figure 2C using transformants obtained from CPY transformed into either BWY46 (WT), BWY497 (*sec62-1*), BWY500 (*srp*), JRY4WT (WT), and JRY4Sec6235DDD (*35DDD*). B) and C) Glycosylation pattern of CPY over time was compared graphically in these strains. D) Western blot analysis as described in Figure 2B was carried out using short N-tail variants (the number of residues in the N-tail is denoted within brackets) transformed into either WT or a Sec62 defective strain (*35DDD*). E) The same was done for long N-tail variants with varying signal sequence hydrophobicities (indicated by the dG value).

DISCUSSION

The main aim of this study was to examine whether or not Sec61p (more specifically, different domains of Sec61p) recognizes different types of substrates distinctly. Our working hypothesis is that signal sequences with distinct determinants for translocation are differentially engaged by the Sec61 translocon.

In order to test this, we sought characteristics that may act as determinant factors for Sec61p recognition. To determine this, we 1) systematically defined the threshold N-terminal length and 2) the threshold signal sequence hydrophobicity for efficient ER translocation of CPY across the Sec61 translocon. Using the N-tail and hydrophobicity values we determined for CPY, we then investigated 2) how these two types of signal sequences are differentially recognized by the Sec61 translocon. Mutations were introduced in two domains of Sec61p, the plug domain and the lateral gate.

Our data show an N-terminal length-dependent inhibition of CPY (Figure 1 and 2B-D), suggesting that having a short N-terminus for signal peptide proteins is a prerequisite for efficient translocation across the Sec61 translocon. Thus, as one determinant of translocation, we proposed a short N-tail length (around 7~14 residues) in front of the signal sequence. Not only is the length a determinant, the sequence context, especially the presence of positively charged residues and the absence of negatively charged residues, promote translocation of CPY variants with short N-tails (Figure 2E). For instance, translocation of D(11)CPY was completely restored when a single aspartate residue was replaced with either a lysine or an alanine. Whereas translocation of S(10)CPY and S(7)CPY (net charge of extensions = 0) were visibly lower than D(10)CPY and D(8)CPY (net charge of extensions = +3 and +2, respectively), despite having similar or the same N-tail length. This also implies that these proteins may initially enter the pore amino-terminus first, and subsequent electrostatic interaction with plug domain residues may regulate translocation by inducing plug and/or lateral gate opening. Previous studies have already shown specific residues (R76E, R74E, E382R) in the plug domain inversely affect integration of a signal anchor protein with

inverted flanking charges, i.e. oppositely charged residues flank either ends of a signal anchor segment. They found that positively charged residues in the plug domains help N-terminally flanking positively charged residues of a signal anchor (the C-terminus contains negatively charged residues) to position on the cytosolic side to facilitate translocation of the C-terminus to the luminal side, thus, highlighting the contribution of the plug domain of Sec61 to the positive-inside rule (23, Figure 5).

Figure 5 A model to show the interaction between signal sequence and the Sec61 translocon

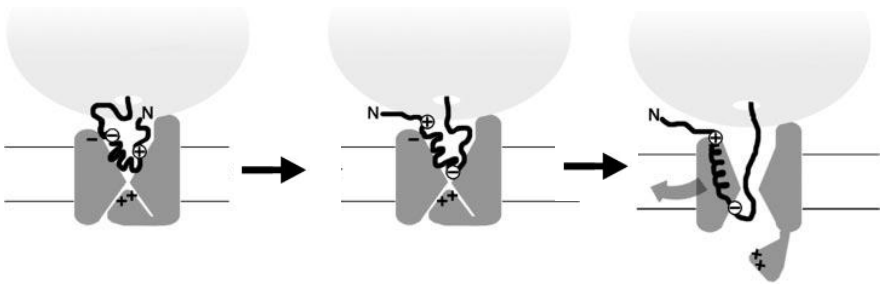


Figure 5 An electrostatic field inside the translocon is thought to contribute to signal sequence orientation according to its flanking charges (positive end faces the cytosolic side) (23).

In line with this, when our short N-tail variants were expressed in plug mutants, so far, our data also suggest the plug domain contributes to the orientation of signal sequence of CPY inside the Sec61 pore accordingly to the positive-inside rule. For example, translocation of D(11)CPY and S(10)CPY was improved in R74E mutant (Figure 2B). D(11)CPY contains an aspartate at position 2, and S(10)CPY contains a glutamate at position 5 (the first charged residues to contact the plug domain if it is assumed these proteins enter the pore head first). Therefore, it may be that the signal sequences of these mutants are now able to position correctly inside the pore of an R74E mutant (perhaps through electrostatic repulsion). However, it must be noted that translocation of proteins such as D(10)CPY and (11,K)CPY

are unaffected in R74E implying that contribution to the positive-inside rule may come from elsewhere.

Nevertheless, why such charge effect is pronounced in CPY variants (compare D(10)CPY and D(11)CPY) for example, Figure 3B) with short extensions may be that these proteins are heavily reliant on the plug domain for signal sequence reorientation (or flipping) inside the pore, and thus a short N-terminal length is required for ease of movement. Gilmore et al. (2012) have already shown specific residues in the Sec61p plug (L63, W65, and L66) contribute to setting the hydrophobicity threshold for signal sequence recognition, demonstrating the importance of the plug domain in secretory protein translocation. Substitution of apolar amino acids with polar ones caused a gain-of-function phenotype of CPY derivatives with truncated signal sequence. Taken together, it seems that both the hydrophobicity threshold and electrostatic interaction provided by the plug domain mediate translocation of secretory proteins.

Previous studies (10, 11) have shown that translocating signal sequence of secretory proteins are exposed to the lipid bilayer at the level of the Sec61 translocon, suggesting that the Sec61 lateral gate is open even during translocation of substrates that do not integrate into the lipid bilayer. Our data corroborate this in that translocation of all short N-tail variants is inhibited by I91A mutant, a Sec61p with a previously characterized ‘closed’ conformation (Figure 3E), showing that some degree of lateral access is required. Intriguingly, an ‘open’ conformation mutant only prohibited translocation when the mutation resided in the TM7 (I293A) of Sec61p, not TM2 (Q96A), showing an asymmetric reliance of translocation of a subset of our short N-tail proteins on the Sec61p lateral gate (Figure 3E, see D(12,K)CPY). Work by Plath et al. (11) found that the signal sequence of ppαF cross-links to TM2 and TM7 of Sec61p, and Sec62p. Of interest, only residues that cross-linked to TM7 also cross-linked with Sec62p. Their results imply at the lateral gate, Sec62p may be found in close proximity to TM7. Our data also show that translocation of only those with a short N-tail with polar signal sequences,

and long N-tails with moderately hydrophobic signal sequences were affected in a *sec62* defective strain (Figure 4D). This suggests that for translocation of such proteins, the signal sequence may be positioned specifically within the pore in a way that has access to Sec62p in the lipid phase through the lateral gate.

We next sought to define the hydrophobicity threshold for substrates that have an N-terminal length that exceeds this inhibitory threshold. Systematically increasing the hydrophobicity of the signal sequence of CPY with the longest extensions allowed us to determine a hydrophobicity threshold of around 0.10~-0.56 kcal/mol that allowed translocation recovery.

The way in which signal sequences can differently engage Sec61p may relate to how they enter the pore initially. Spiess et al. have shown that signal anchor proteins with N-terminal lengths greater than ~24 residues tend to have increased C-terminal translocation across the membrane than those with short N-terminal lengths, implying that a longer N-terminal length promotes hairpin formation at the translocon (24). Likewise, our CPY variants with long N-tails may enter the pore as a hairpin, which may facilitate correct orientation. The insertion as a hairpin then precludes the need for flipping of the signal sequence, which may be why moderately signal sequences with very long N-terminal lengths are able to translocate (e.g. D(0.51)CPY, D(0.10)CPY etc). Interestingly, we see that translocation of D(0.10)CPY, is still dependent on the actions of Sec62p (whatever that may be) (Figure 4E).

These data suggest that different determinants exist for substrates entering the ER. In sum, we can present a simple model when signal peptide proteins are targeted to Sec61p, they enter N-terminus first into the pore and through electrostatic interactions with plug domain residues (and other factors), the N-terminus of signal peptides are re-orientated so that the signal peptide is specifically positioned within the pore next to the lateral gate (subsequent interaction with TM7 and Sec62p may stabilize this orientation). Then, signal sequence cleavage by the signal peptidase complex may complete translocation by releasing the

substrate into the lumen with the aid of the ratcheting actions of luminal Kar2p. Conversely, proteins with hydrophobic signal sequences and a long N-terminus may enter the pore in a completely different fashion. The signal sequence may enter as a hairpin and exit into the lipid phase may be swift, as hydrophobic segments prefer to be in the lipid milieu. (Figure 6).

Figure 6

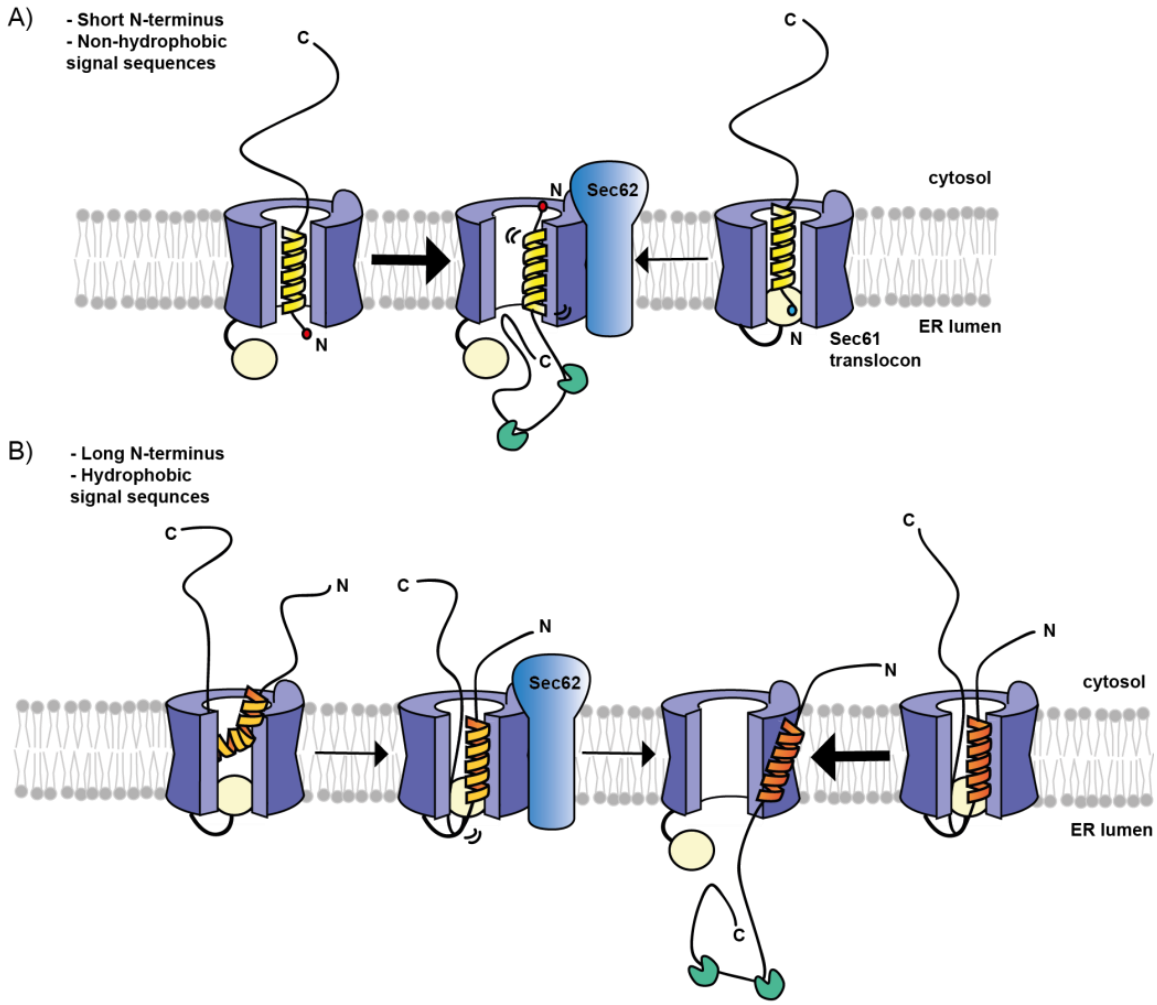


Figure 6 A) When a protein with a polar signal sequence and a short N-tail arrives at the Sec61p pore, they enter amino-terminus. Those with N-terminus enriched with positive charges (left, red circle represent a positively charged residue) are able to re-orient well, as result of electrostatic interactions with plug domain residues (and other factors), so that the N-terminus faces the cytosol (middle). Upon correct interaction with the lateral gate and Sec62, the C-terminus is released into the lumen (prior to or after signal sequence cleavage). However, those with N-terminus containing negative charges (right), are not able to re-orient meaning that translocation cannot proceed. B) Proteins with hydrophobic signal sequences and a long N-terminus may enter the pore as a hairpin and exit into the lipid phase may be swift if the transmembrane core is sufficiently hydrophobic (right, orange helix). However, those with moderately hydrophobic segments (left, light orange helix) may still enter as a hairpin, but may prefer to be in the aqueous milieu, they may still require Sec62p to aid translocation into the ER.

REFERENCES

1. Ng, D.T., Brown, J.D., and Walter, P. (1996). Signal sequences specify the targeting route to the endoplasmic reticulum membrane. The Journal of cell biology *134*, 269-278.
2. Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737-741.
3. Zimmermann R, Eyrisch S, Ahmad M, and Helms V. (2011) Protein translocation across the ER membrane. Biochim Biophys Acta. 1808(3):912-24.
4. von Heijne G. (1983) Patterns of amino acids near signal-sequence cleavage sites. Eur J Biochem *133* (1) 17-21.
5. Martoglio B. and Dobberstein B. (1998) Signal sequences: More than just greasy peptides. Trends Cell Biol *8* (10), 410-5.
6. Hegde, R.S. and Bernstein, H.D. (2006) The surprising complexity of signal sequences. Trends Biochem Sci *31*(10), 563-71.
7. Rapoport, T.A. (2007). Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. Nature *450*, 663-669.
8. Blachy-Dyson, E. and Stevens, T. H. (1987) Yeast carboxypeptidase Y can be translocated and glycosylated without its amino-terminal signal sequence. The Journal of Cell Biology *104*, 1183-1191.
9. Johnson, N., Habdenteufel, S., Theis, M., Paton, A. W., Paton, J. C., Zimmerman, R., and High, S. (2013) The signal sequence influences post-translational ER translocation at distinct stages. PLOS *8*(10), e75394.
10. Martoglio, B., Hofmann, M.W., Brunner, J., and Dobberstein, B. (1995). The protein-conducting channel in the membrane of the endoplasmic reticulum is open laterally toward the lipid bilayer. Cell *81*, 207-214.
11. Plath, K., Mothes, W., Wilkinson, B. M., Stirling, C. J., and Rapoport, T.A. (1998). Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. Cell *94*, 795-807.
12. Johnson, N., Habdenteufel, S., Theis, M., Paton, A. W., Paton, J. C., Zimmerman, R., and High, S. (2013) The signal sequence influences post-translational ER translocation at distinct stages. PLOS *8*(10), e75394.
13. Under review. Jung S. J. et al.
14. Spiess, M. et al. (1997)
15. Under review Yim C. et al.
16. SWISSPROT
17. Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature methods *8*, 785-786.
18. Hessa, T., Meindl-Beinker, N., Bernsel, A., Kim, J., Sato, Y., Lerch, M., Lundin, C., Nilsson, I., White, SH. and von Heijne, G. **(2007)** Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature.* **450**, 1026-1030.
19. Magrane M. and the UniProt consortium UniProt Knowledgebase: a hub of integrated protein data Database, 2011: bar009 (2011).
20. Güldener U, Münsterkötter M, Kastenmüller G, Strack N, van Helden J, Lemer C, Richelles J, Wodak SJ, Garcia-Martinez J, Perez-Ortin JE, Michael H, Kaps A, Talla E, Dujon B, Andre B, Souciet JL, De Montigny J, Bon E, Gaillardin C, Mewes HW. CYGD: the Comprehensive Yeast Genome Database. *Nucleic Acids Research Jan 1;33 Database issue:D364-8 (2005)*
21. IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.
22. Trueman SF, Mandon EC, Gilmore R. (2012) A gating motif in the translocation channel sets the hydrophobicity threshold for signal sequence function. J Cell Biol. 10;199(6):907-18.
23. Junne T, Schwede T, Goder V, and Spiess M. (2007) Mutations in the Sec61p channel affecting signal sequence recognition and membrane protein topology. J Biol Chem 282(45):33201-9.
24. Kocik, L., Junne, T., and Spiess, M. (2012). Orientation of internal signal-anchor sequences at the Sec61 translocon. Journal of molecular biology *424*, 368-378.

국문초록

Saccharomyces cerevisiae 의 세포질에서 생산된 30%의 단백질은 소포체로 들어가며, 이러한 수송 과정은 이 단백질들이 가진 고유의 기능을 수행하는 데 있어 필수적이다. 지금까지 알려진 바에 따르면, 이 단백질들이 세포질에서 소포체로 이동하는 경로는 크게 두 가지로 나뉘어 진다. 첫번째는 단백질이 번역과 동시에 소포체로 이동하는 경로이며, 두번째는 번역이 종료된 후, 소포체로 이동하는 경로이다. 그러나 경로와 상관 없이, 소포체로 수송된 단백질들은 Sec61 이라는 채널 단백질을 통해 소포체 내강으로 전좌된다. 이 Sec61 은 단백질이 통과할 수 있는 통로를 형성하고, 소수성 막관통 영영의 삽입을 위한 측면 통로를 가지고 있다.

이 연구의 목적은 Sec61 이 서로 다른 종류의 신호서열을 구분하는지에 대해 조사하는 것이다. 첫째로, Sec61 이 구분하는 신호서열을 가진 단백질들을 찾기 위해서, 모델 단백질인 CPY 를 이용하여 효율적인 전좌에 필요한 아미노 말단의 길이와 소수성의 역치를 알아보았다. 이를 통해 약한 소수성을 띠는 신호서열을 가진 CPY 의 전좌를 위해서는 아미노 말단의 길이가 짧아야 하며, 아미노 말단의 길이가 긴 경우에는 매우 소수성인 신호서열이 필요함을 관찰하였다. 나아가 이러한 두 종류의 신호서열이 Sec61 에 의해 다르게 인식되는지를 알아보기 위하여 이들을 Sec61 plug 도메인 변이주와 측면 통로 변이주에 발현시켰다. 그 결과 짧은 아미노 말단과 약한 소수성 신호 서열을 가진 단백질의 전좌는, 긴 아미노 말단과

강한 소수성 신호 서열을 가진 단백질보다 plug 도메인과 TM7 에 변이에 의해 더 크게 영향을 받았으며 따라서 Sec61 이 서로 다른 기질을 상이한 기작을 통해 인식함을 발견하였다.

주요 단어: 출아효모, 소포체, CPY, Sec61, 신호 서열, 아미노 말단 길이, 전좌