이학석사 학위논문

Studies on the roles of Spc1 and Spc2 in substrate selection for signal peptidase

신호서열절단효소 기질 선택에서의 Spc1과 Spc2의 역할에 관한 연구

2020년 8월

서울대학교 대학원

생명과학부

정 연 지

신호서열절단효소 기질 선택에서의 Spc1과 Spc2의 역할에 관한 연구

지도교수 김 현 아

이 논문을 이학석사 학위논문으로 제출함

2020년 6월

서울대학교 대학원

생명과학부

정 연 지

정연지의 이학석사 학위논문을 인준함

2020년 6월

위 원 장 조 형 태 (인)

부위원장 김 현 아 (인)

위 원 한 광 직 (인)

Studies on the roles of Spc1 and Spc2 in substrate selection for signal peptidase

by

Yeonji Chung

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

August 2020

School of Biological Sciences

Seoul National University

Abstract

Proteins that are targeted to the endoplasmic reticulum (ER) membrane in eukaryotes or plasma membrane in prokaryotes mostly have an N-terminal signal sequence. The signal sequence contains a cleavage site at its C-terminus, where evolutionarily conserved signal peptidase (SPase) cleaves a signal sequence (signal peptide) during or after translocation of the precursor across the membrane. Uncleaved signal sequence integrates into the membrane, which then becomes a transmembrane segment of a membrane protein. Due to heterogeneity of signal sequences, how SPase selects substrates remains elusive. In eukaryotic cells, SPase resides in the ER membrane in a complex with other integral membrane components. Yeast signal peptidase complex (SPC) is composed of Sec11, Spc1, Spc2, and Spc3, among which Sec11 is a catalytic subunit, but the roles of the other subunits are poorly understood.

In this thesis, to investigate the roles of Spc1 and Spc2, first, I established an *in vivo* cleavage assay with a set of test proteins carrying signal sequences varied in their hydrophobicity and N-length. With this, I defined the range of SPase substrates. My data show that SPase favors short N-length and less hydrophobic signal sequences for processing. Next, the substrate spectrum in *spc1*Δ and *spc2*Δ cells was determined and compared to that in the wild-type cell. In *spc1*Δ cells, longer N-length signal sequences that were membrane anchored in the wild-type cell became more susceptible

to cleavage by SPase without Spc1. In *spc2*Δ cells, shorter N-length signal sequences became less efficiently cleaved, while longer N-length signal sequences became more efficiently cleaved by SPase, indicating that the distinction between cleavable versus uncleavable signal sequences is weakened in the absence of Spc2. These results show that cleavage of membrane anchored signal sequences is enhanced by SPase lacking Spc1 or Spc2. Mutagenesis experiments revealed that the usage of canonical cleavage sites in CPY variants remained the same in the absence of Spc1 or Spc2. These results suggest that Spc1 and Spc2 are involved in regulating substrate selection for SPase, and implicate their roles in membrane protein biogenesis.

Keywords: Signal sequence, Signal peptidase, Spc1, Spc2, Membrane protein

Student Number: 2018-22166

Table of Contents

Introduction

Proteins synthesized in living cells should be localized at the right place through specific biosynthetic pathways to function properly. The signal sequence plays important roles for biogenesis of secretory and membrane proteins, in that it first targets precursors to the entry site of protein trafficking during or after protein synthesis. In eukaryotic cells, proteins targeted to the endoplasmic reticulum (ER) membrane translocate across the membrane through the Sec61 channel, and when the C-terminus of the signal sequence is exposed to the ER lumen, signal peptidase (SPase) recognizes and cleaves the signal sequence cleavage site co- or post-translocationally. SPase substrates should be translocated in a $N_{cytosol}$-C_{lumen} membrane orientation for a cleavage site to be recognized by SPase from the ER luminal side. Uncleaved signal sequence is inserted into the membrane, becoming a transmembrane segment of a membrane protein (Fig. 1).

The signal sequence is so diverse, and this diversity of the signal sequence confers complexity in predicting which one will be processed by SPase. Nevertheless, once SPase selects substrates, it processes cleavage sites in a highly precise manner. Hence, whether signal sequences are to be cleaved or not is an important problem, since it can determine the fate of secretory and membrane proteins.

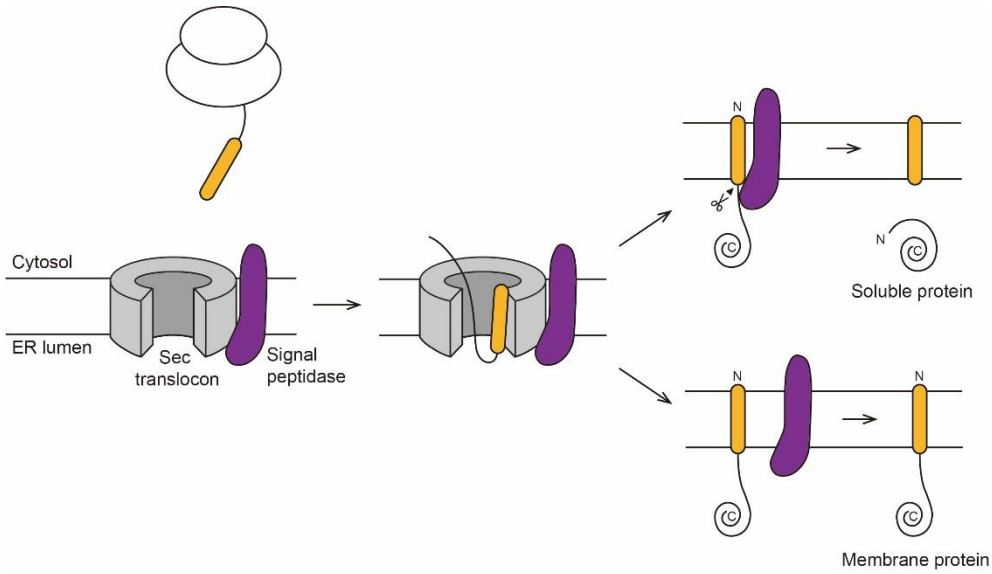

Figure 1. Substrate selection by signal peptidase can determine the fate of secretory and membrane proteins.

Statistical analysis of known signal sequences has revealed that they share common features. The signal sequence is typically ~20–30 amino acids long, consisting of three regions: a positively charged n–region, a central hydrophobic h–region, and a more polar c–region [1] (Fig. 2). In the c–region, cleavable signal sequence must contain cleavage sites from which residues in position −3 and −1 should be small, uncharged amino acids for recognition and processing by SPase. Thus, this so–called "(−3, −1) rule" defines the presence of acceptable cleavage sites within the signal sequence. In contrast, large polar residues (Asn, Gln) were never found in position −3, and Pro residue was absent from the position −3 to +1 relative to the known cleavage site of the signal sequence [2]. A deep learning model–based computational method using signal peptide datasets enabled the development of a tool to calculate the most probable sites for cleavage. For example, SignalP–5.0 server (http://www.cbs.dtu.dk/services/SignalP/) predicts the presence of signal peptides and the location of their cleavage sites in proteins [3]. However, the signal sequence containing potential cleavage sites that fulfill (−3, −1) rule is not always processed by SPase. It is not uncommon that uncleaved signal sequence of a membrane protein has potential cleavage sites within the transmembrane segment. Previously, a single–spanning membrane protein, human asialoglycoprotein receptor H1, was shown to become a substrate of signal peptidase when its original N–terminal domain of 40 residues preceding the hydrophobic segment was truncated to 4 residues [4].

This result demonstrates that the length of n-region is a determinant for signal sequence cleavage. Besides, a cleavable signal sequence with a longer h-region composed of poly-Leu residues was converted to an uncleaved signal sequence, suggesting that the length of h-region also affects signal sequence cleavage [5]. In a recent study, bioinformatics analysis of yeast natural proteins containing single signal sequence has revealed that n-region length of less hydrophobic signal sequences ($\Delta G_{app} > 0.5$ kcal/mol) tends to be shorter than that of highly hydrophobic signal sequences ($\Delta G_{app} < -1.5$ kcal/mol) [6]. Taken together, presence of a cleavage site is not the sole factor for signal sequence cleavage. Rather, characteristics of each region in tripartite structure of the signal sequence contribute to substrate selection by SPase. However, how SPase selects substrates on the basis of various features on signal sequences and which component of the SPase, if any, is involved in substrate selection remains to be solved.

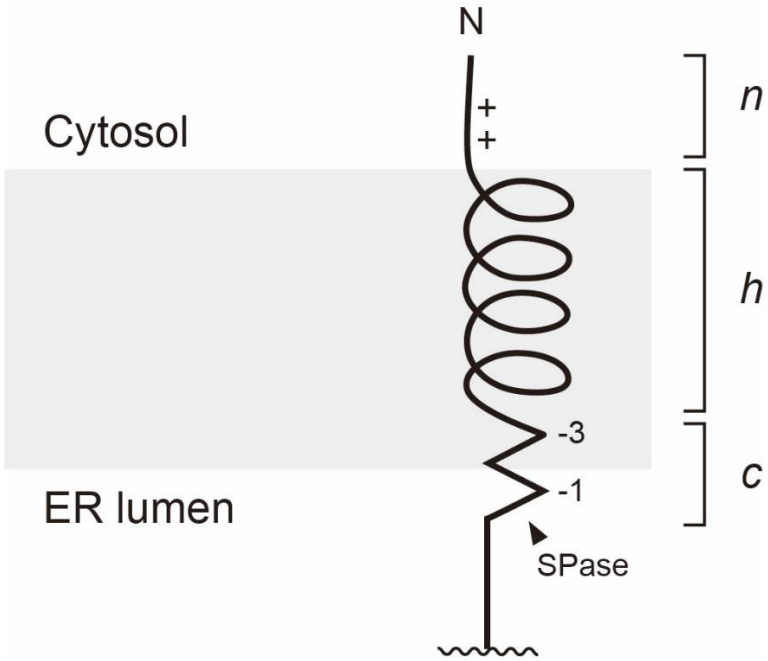

Figure 2. Schematics of the signal sequence. The signal sequence has a tripartite structure, consisting of n-region, h-region, and c-region.

SPase is an evolutionarily conserved ER membrane protease. Comparison of the amino acid sequence of leader peptidase of *E. coli* with that of Sec11 of yeast and Spc18 and Spc21 of mammals has shown that they share similarities [7]. In eukaryotes, signal peptidase resides in the ER membrane in association with other ER membrane proteins [8, 9]. In *Saccharomyces cerevisiae*, the signal peptidase complex (SPC) consists of Spc1, Spc2, Spc3 and Sec11. Spc3 and Sec11 are single-spanning membrane proteins with their large domains placed on the luminal side of the ER membrane, which are thought to be engaged in catalyzing cleavage reaction. Spc1 and Spc2 are double-spanning membrane proteins with their N- and C-terminal domains oriented toward the cytosolic side, whereas little is exposed to the ER luminal side where the catalysis occurs. Topologies of Spc1 and Spc2 thus hint that they might not be directly involved in catalytic activity of signal peptidase. However, earlier studies on Spc1 and Spc2 have revealed that they are important for cell viability and signal peptidase activity under unfavorable conditions. For example, overexpression of Spc1 facilitated signal peptide cleavage in the sec11 mutant, and rescued growth defect of the sec11 mutant [10]. Spc2 was important for cell viability and signal peptidase activity at a high temperature [11]. These findings have shown that Spc1 and Spc2 function in alleviating the signal peptidase defect in abnormal cells, yet their roles remain elusive.

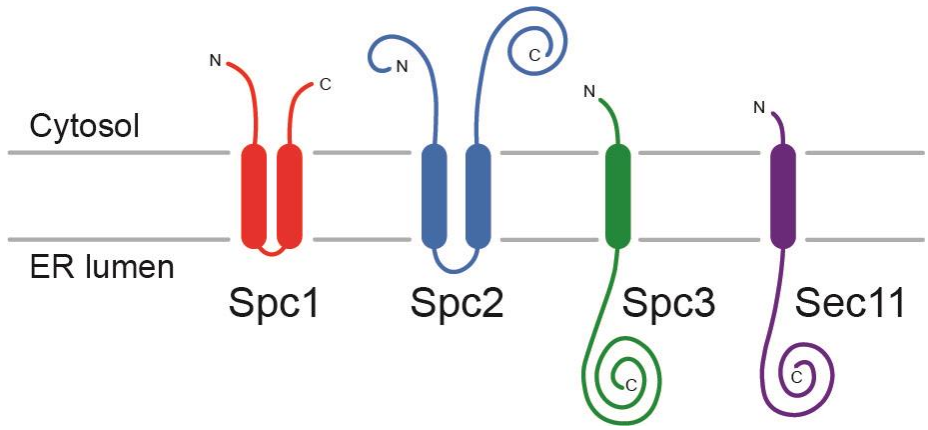

Figure 3. Topologies of the yeast signal peptidase complex (SPC) subunits. Essential subunits Spc3 and Sec11 single-span the membrane with large ER luminal domain. Spc1 and Spc2 span the membrane twice with majority of their domains facing toward the cytosolic side.

Materials and Methods

Yeast strains

The Saccharomyces cerevisiae haploid W303-1α (*MATα, ade2, can1, his3, leu2, trp1, ura3*) was used as a WT strain. The genomic ORF of *SPC1* or *SPC2* in W303-1α were replaced with *HIS3* marker amplified from pCgH plasmid [12] by homologous recombination, and *spc1*Δ (*MATα, spc1Δ::HIS3, ade2, can1, his3, leu2, trp1, ura3*) and *spc2*Δ (*MATα, spc2Δ::HIS3, ade2, can1, his3, leu2, trp1, ura3*) strains were generated. *spc3-2 is* a temperature-sensitive mutant, exhibiting a defect in cleavage when grown at 37 °C [13]. For overexpression of Spc2, pRS426 vector carrying *SPC2* under GPD promoter was transformed into the *spc2*Δ strain.

Construction of plasmids

All N-length and hydrophobicity variants of CPY were derived from pRS424GPDN26CPY-3xHA constructed in the previous study [6]. Using this construct as a template, first, residues 323-532 of CPY were truncated to facilitate the band separation on a SDS-gel by site-directed mutagenesis following the manufacture's protocol (Toyobo, Japan). Next, the N-terminus was gradually truncated or the signal sequence of CPY was modified by site-directed mutagenesis. pRS426 vector containing *SPC2* was cloned by homologous recombination or using the Gibson assembly kit following the manufacture's protocol.

In vivo metabolic labeling

Yeast cells were grown until OD_{600} reached between 0.3 and 0.8 in selective medium. 1.5 OD_{600} units of cells were harvested by centrifugation (2,170g, 5 min, 4°C), washed with −Met medium without ammonium sulfate, and incubated at 30°C for 10 min. Cells were centrifuged and resuspended in 150 μl of −Met medium without ammonium sulfate, and radiolabeled with [^{35}S]Met (40 μCi per 1.5 OD_{600} units of cells) for 5 min at 30 °C. After incubation, labeling was stopped by the addition of 750 μl of ice−cold stop solution buffer containing 20 mM Tris−HCl (pH 7.5) and 20 mM sodium azide. Cell pellets were harvested by centrifugation (16,000g, 1 min, 4°C) and stored at −20°C until use.

Immnunoprecipitation and SDS−PAGE

Radiolabeled cell pellets were resuspended in 100 μl of lysis buffer (20 mM Tris−HCl (pH 7.5), 1% SDS, 1 mM DTT, 1 mM PMSF, and 1X Protease Inhibitor Cocktail) and mixed with 100 μl of ice−cold glass beads. Cell suspensions were vortexed for 2 min twice, keeping the samples on ice for 1 min in between. Subsequently, samples were incubated at 60 °C for 15 min and centrifuged (6,000g, 1 min, 4°C). The supernatant fractions were mixed with 500 μl of immunoprecipitation (IP) buffer (15 mM Tris−HCl (pH 7.5), 0.1% SDS, 1% Triton X−100, and 150 mM NaCl), 1 μl of anti−HA antibody, and 50 μl of pre−washed protein G−agarose beads (Thermo Scientific Pierce) and rotated at room temperature for 3 hr. The

agarose beads were washed twice with IP buffer, once with ConA buffer (500 mM NaCl, 20 mM Tris-HCl (pH 7.5), and 1% Triton X-100), and once with buffer C (50 mM NaCl and 10 mM Tris-HCl (pH 7.5)). The beads were incubated with 50 μl of SDS sample buffer (50 mM DTT, 50 mM Tris-HCl (pH 7.6), 5% SDS, 5% glycerol, 50 mM EDTA, 1 mM PMSF, 1X Protease Inhibitor Cocktail (Quartett), and little bromophenol blue) at 60°C for 15 min, followed by Endo H (Promega) treatment at 37°C for 1 hr. Protein samples were then loaded onto SDS gels and resolved by electrophoresis.

Data quantification

A Typhoon FLA 7000 phosphoimager and a Typhoon FLA 9500 phosphoimager were used for the detection of radiolabeled signals on SDS-gel by autoradiography. Data were processed and quantified using the MultiGauge version 3.0 software. Cleavage efficiency was calculated from the band intensities of glycosylated bands using the formula:

Cleavage (%) = Cleaved band × 100 / cleaved + full-length bands

Carbonate extraction

Yeast cells were grown in selective medium at 30°C overnight. 5 OD_{600} units of cells were harvested by centrifugation (2,170g, 5 min, 4 °C) and washed with dH_2O. Cells were resuspended in 200 μl of lysis buffer (20 mM Tris-HCl (pH 8.0), 10 mM EDTA, 100 mM NaCl, 300 mM Sorbitol, 1mM PMSF, and 1X Protease Inhibitor Cocktail)

and vortexed with glass beads for 10 min at 4℃. After a quick spin down, cell lysate was transferred to a new pre-chilled tube. The remaining glass beads were washed with additional 200 μl of lysis buffer, added to the final lysate after washing. The final lysate was subjected to centrifugation (20,000g, 30 s, 4℃) to remove cell debris, and transferred to a new pre-chilled tube. 100 μl of the lysate was saved for 'Total' fraction, and the remaining lysate was incubated with 300 μl of 0.1M Na_2CO_3 (pH 11.5) on ice for 30 min, followed by centrifugation (20,000g, 20 min, 4℃). The resulting supernatant was saved for 'Supernatant' fraction, and the pellet was washed with additional 200 μl of 0.1M Na_2CO_3. Separate 'Total', 'Supernatant', and 'Pellet' fractions were precipitated with Trichloroacetic acid (TCA) at 12.5% (v/v) final concentration on ice for 30 min, centrifuged (20,000g, 15 min, 4℃), and washed with acetone. Samples were resuspended in SDS sample buffer and analyzed by SDS-PAGE and Western blotting.

Results

Extended N-terminal length preceding the signal sequence decreases cleavage of the signal sequence

For construction of model proteins, N-terminal 26 residues of cytosolic part of a single-span membrane protein Dap2 was fused to the N-terminus of residues 1-323 of secretory protein CPY. The signal sequence of CPY was modified to be more hydrophobic, followed by gradual truncation of the N-terminus of N26CPYt(*h*) (Fig. 4 and Table 1). As CPY variants have three N-glycosylation sites in their C-terminal domain, when they are translocated into the ER, become glycosylated, thus used as an indication of C-terminal translocation. All model proteins were C-terminally HA-tagged, so that they can be immunoprecipitated or Western blotted using anti-HA antibody. When N-length variants were pulse labeled for 5 min at 30° C, immunoprecipitated, resolved by SDS-PAGE, and visualized by autoradiography, one or two protein bands were observed, and the amounts of slowly-migrated band increased as the N-length became longer. After treatment of Endo H, all protein bands were shifted down, indicating that all N-length variants were efficiently translocated into the ER (Fig. 5A). As all N-length variants were sensitive to Endo H, N-length was not a determinant for translocation efficiency. When selected N-length variants (#=1, 9, 16, and 26) were tested in *spc3-2*, a SPase defective strain when grown at high temperature, only slowly-migrated bands were detected on a SDS-gel, indicating slowly-migrated bands in Fig. 5A

are signal sequence-uncleaved species, while fast-migrated bands are signal sequence-cleaved species (Fig. 5B). In addition, this also demonstrates that all cleavages are mediated by SPase. Next, carbonate extraction was carried out to confirm localization of cleaved and uncleaved species. Signal sequence-cleaved species are found in soluble fraction (supernatant fraction), while signal sequence-uncleaved species in membrane fraction (pellet fraction), confirming fast- and slowly-migrated bands on SDS-gels are signal sequence-cleaved and uncleaved species, respectively (Fig. 5C). Quantification data of cleavage efficiency, measured by relative band intensities of a fast-migrated band and a slowly-migrated band in each lane, clearly showed the N-length dependent cleavage pattern, in which cleavage efficiency decreases as the N-length becomes longer (Fig. 5D).

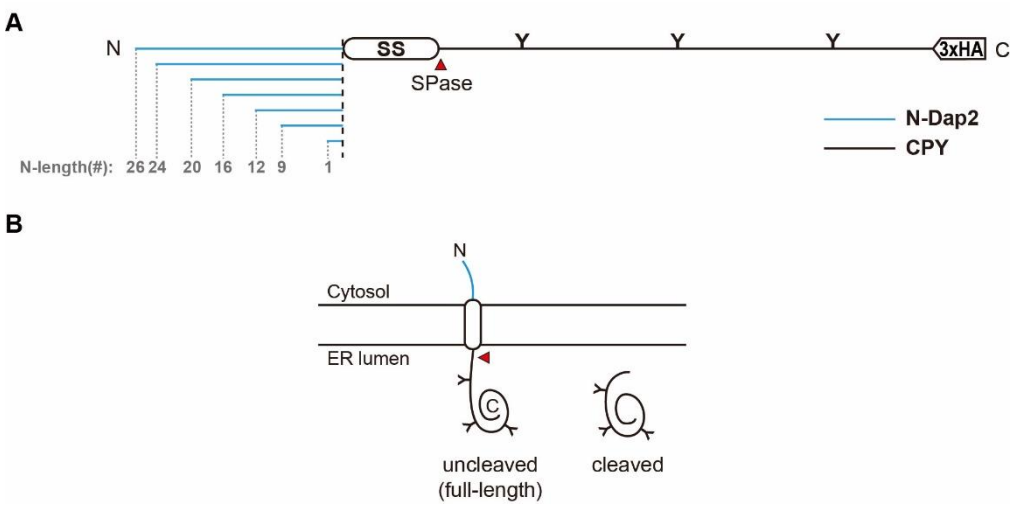

Figure 4. Construction of model proteins. (A) Schematic representation of model proteins. The N-terminal 26 residues of cytosolic part of single-span membrane protein Dap2 (blue) was fused to the N-terminus of residues 1-323 of secretory protein CPY (black), and the N-terminal residues were gradually truncated. SS; Signal sequence (B) Schematic of predicted topology of N#CPYt model proteins.

Name	N-terminal sequence	ΔG_{app} (kcal/mol)	N-length (aa)
N26CPYt(*h*)	MEGGEEEVERIPDELFDTKKKHLLDKLLLTLLLCLLLLSTTLAKAISL	-2.615	26
N24CPYt(*h*)	MGEEEVERIPDELFDTKKKHLLDKLLLTLLLCLLLLSTTLAKAISL	-2.615	24
N20CPYt(*h*)	MVERIPDELFDTKKKHLLDKLLLTLLLCLLLLSTTLAKAISL	-2.615	20
N16CPYt(*h*)	MPDELFDTKKKHLLDKLLLTLLLCLLLLSTTLAKAISL	-2.615	16
N12CPYt(*h*)	MFDTKKKHLLDKLLLTLLLCLLLLSTTLAKAISL	-2.615	12
N9CPYt(*h*)	MKKKHLLDKLLLTLLLCLLLLSTTLAKAISL	-2.615	9
CPYt(*h*)	MKLLTLLLCLLLLSTTLAKAISL	-1.850	1
N26CPYt(*i*)	MEGGEEEVERIPDELFDTKKKHLLDKLAFTLLLCLLLLSTTLAKAISL	-2.011	26
N24CPYt(*i*)	MGEEEVERIPDELFDTKKKHLLDKLAFTLLLCLLLLSTTLAKAISL	-2.011	24
N20CPYt(*i*)	MVERIPDELFDTKKKHLLDKLAFTLLLCLLLLSTTLAKAISL	-2.011	20
N16CPYt(*i*)	MPDELFDTKKKHLLDKLAFTLLLCLLLLSTTLAKAISL	-2.011	16
N12CPYt(*i*)	MFDTKKKHLLDKLAFTLLLCLLLLSTTLAKAISL	-2.011	12
N9CPYt(*i*)	MKKKHLLDKLAFTLLLCLLLLSTTLAKAISL	-2.011	9
CPYt(*i*)	MLAFTLLLCLLLLSTTLAKAISL	-2.676	1
N26CPYt(*l*)	MEGGEEEVERIPDELFDTKKKHLLDKLAFSSLLCLLLLSTTLAKAISL	-0.950	26
N24CPYt(*l*)	MGEEEVERIPDELFDTKKKHLLDKLAFSSLLCLLLLSTTLAKAISL	-0.950	24
N20CPYt(*l*)	MVERIPDELFDTKKKHLLDKLAFSSLLCLLLLSTTLAKAISL	-0.950	20
N16CPYt(*l*)	MPDELFDTKKKHLLDKLAFSSLLCLLLLSTTLAKAISL	-0.950	16
N12CPYt(*l*)	MFDTKKKHLLDKLAFSSLLCLLLLSTTLAKAISL	-0.950	12
N9CPYt(*l*)	MKKKHLLDKLAFSSLLCLLLLSTTLAKAISL	-0.950	9
CPYt(*l*)	MLAFSSLLCLLLLSTTLAKAISL	-0.948	1

Table 1. List of CPY variants used in this study. The N-terminal sequences preceding and hydrophobicity of the signal sequence in N#CPYt variants are shown. Hydrophobicity is predicted by the delta G predictor (http://dgpred.cbr.su.se/). The N-length indicates the number of amino acids preceding the SS of CPY. Predicted SS is underlined.

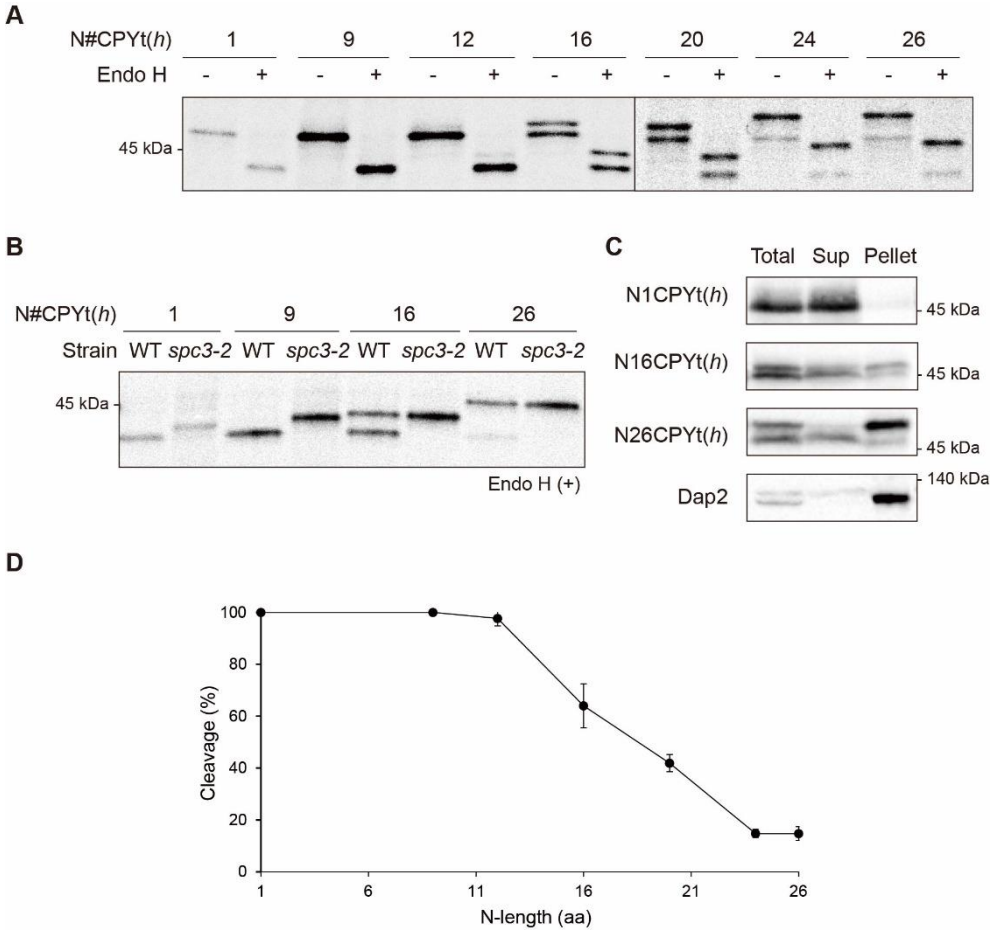

Figure 5. Relative positions of the signal sequence from the N-terminus is a key determinant for signal peptidase processing. (A) Wild-type (WT) cells expressing N-length variants of N#CPYt(*h*) (#=1, 9, 12, 16, 20, 24, and 26) were radiolabeled with [^{35}S]Met for 5 min at 30°C. To facilitate the band separation, protein samples were treated with Endo H (Endo H −, +; endoglycosidase H). (B) Indicated N#CPYt(*h*) variants in WT or *spc3-2* strain were analyzed as in (A), except that N#CPYt(*h*) variants in *spc3-2* strain was incubated at 37°C for 30 min prior to radiolabeling. All samples were treated with Endo H. (C) Western Blot analysis of Total (T), Supernatant (S), and Pellet (P) fractions from carbonate extraction of cells expressing

indicated N#CPYt(*h*) variants and Dap2 are shown. (D) Signal sequence−cleaved bands were quantified and plotted against N−length. Graphs depict means±SD of triplicates.

SPase prefers less hydrophobic signal sequences for processing

To further study the effect of hydrophobicity of the signal sequence on cleavage, CPY variants carrying varying hydrophobicity were constructed (Fig. 6A). Cleavage of CPY variants with intermediate hydrophobicity (*i*) was almost similar compared to the most hydrophobic sets of CPY variants (*h*), but a slight increase in overall cleavage was observed (Fig. 6B). In case of the lowest hydrophobicity (*l*), non-glycosylated bands were detected for N-lengths 24 and 26, which indicates some amounts were not targeted to the ER. N-length dependent cleavage pattern was still observed for the lowest hydrophobicity variants, but signal sequences were more efficiently cleaved compared to the other two hydrophobicity variants. Thus, these results indicate that less hydrophobic signal sequences are preferred by SPase.

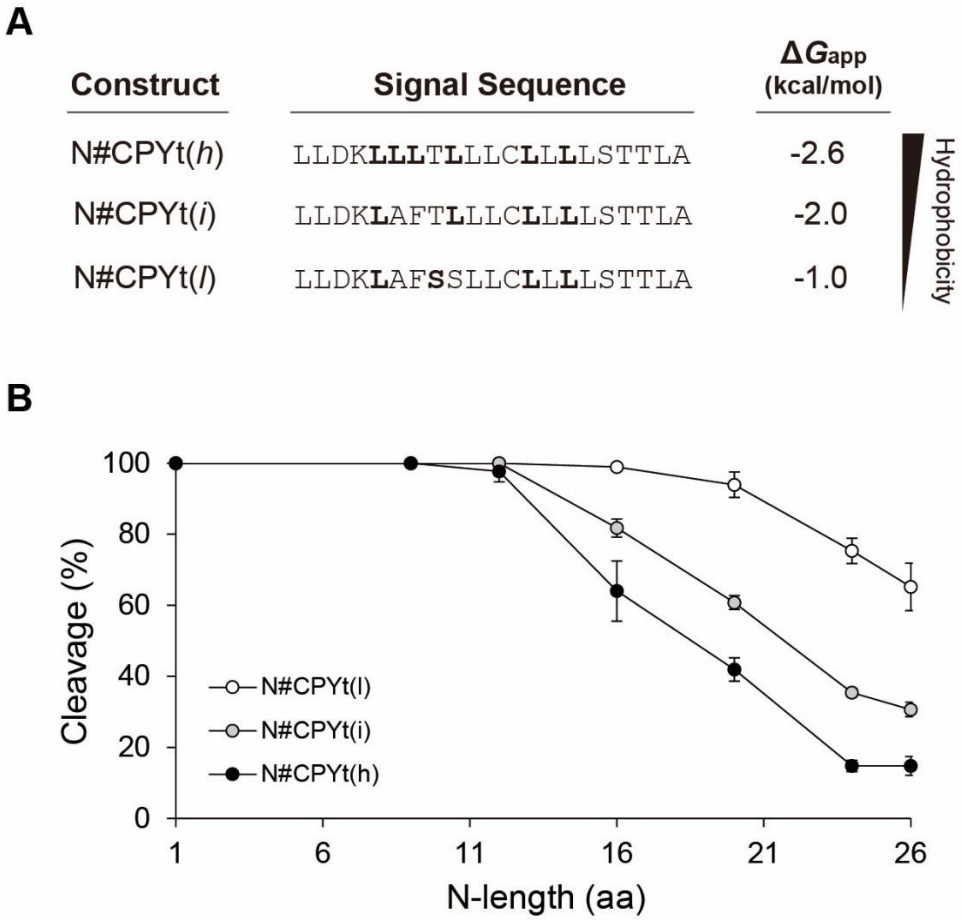

Figure 6. (A) SSs of N#CPYt variants and their hydrophobicities predicted by the delta G predictor (http://dgpred.cbr.su.se/) are shown. (B) The relative amounts of a SPase-processed species over glycosylated products for each CPY variant were measured and plotted against the N-lengths.

Cleavage of the internal signal sequence is enhanced in the absence of Spc1

The signal sequence cleavage was increased in *spc1*Δ cells for N#CPYt(*h*) variants with N−length longer than 16 compared to those in WT cells (Fig. 7A). The difference of the SS cleavage efficiency in *spc1*Δ and WT cells became larger as the N−length became longer (Fig. 7B). Less hydrophobic N#CPYt(*i*) (Figs. 7C and D) and N#CPYt(*l*) (Figs. 7E and F) sets showed similar cleavage patterns that cleavage efficiency increased for the longer N−length variants in *spc1*Δ cells. These data show the altered substrate spectrum for SPase lacking Spc1 that membrane anchored internal signal sequences are recognized as substrates and cleaved.

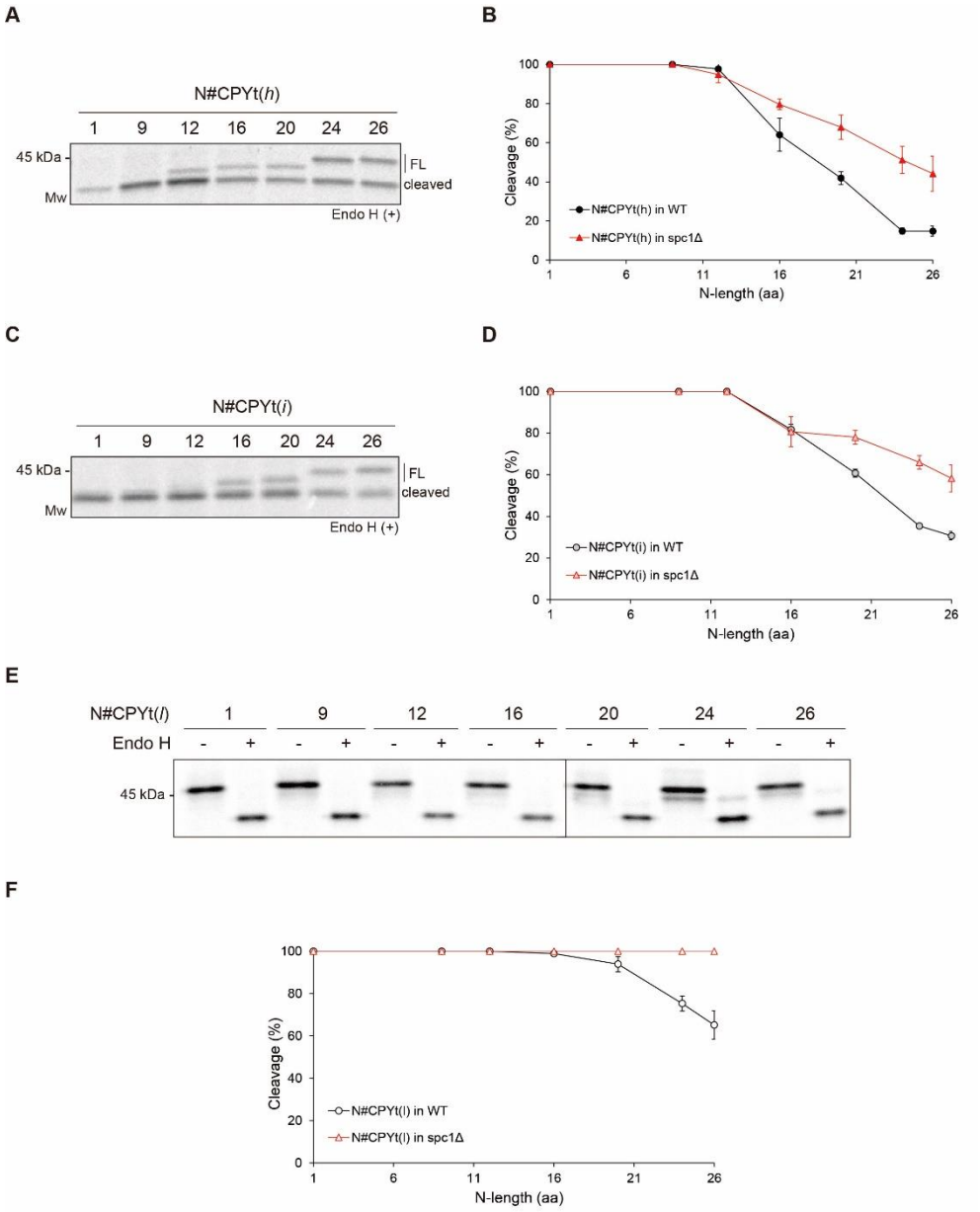

Figure 7. Cleavage of internal signal sequences is enhanced in the absence of Spc1. (A, C, and E) *spc1*Δ cells expressing indicated CPY variants, N#CPYt(*h*) (A), N#CPYt(*i*) (C) and, N#CPYt(*l*) (E) were radiolabeled with [^{35}S]Met for 5 min at 30°C. To facilitate the band separation, protein samples were treated with Endo H. (B, D, and F) *spc1*Δ cells expressing indicated CPY variants, N#CPYt(*h*)

(B), N#CPYt(*i*) (D) and, N#CPYt(*l*) (E) were analyzed as in Fig. 5A. At least three independent experiments were carried out and the average is shown with standard deviations.

Large cytosolic domain of Spc2 is involved in regulating substrate selection for SPase

When cleavages of N#CPYt(*h*) variants were assessed in *spc2*Δ strain, shorter N-length signal sequences became less efficiently cleaved, while longer N-length signal sequences became more efficiently cleaved by SPase (Figs. 8A and B). Despite the altered substrate spectrum in *spc2*Δ cells, it is difficult to tell exactly what kind of substrates are targets for regulation of Spc2 only by looking at this cleavage pattern. Spc2 has a large C-terminal domain on the cytosolic side compared to Spc1. To determine whether the cytosolic C-terminal domain of Spc2 plays role in substrate selection, C-terminal 58 residues were truncated, and the truncated Spc2 (Spc2 tr) was expressed in the *spc2*Δ background (Fig. 8C). For comparison, full-length Spc2 (Spc2 FL) or empty vector was expressed in *spc2*Δ cells in parallel. The cleavage pattern shown in *spc2*Δ cells was also observed when C-terminal domain of Spc2 (residues 121-178) was truncated (Figs. 8C and D), but overall cleavage efficiency was slightly increased compared to empty vector-expressed *spc2*Δ cells. Thus, this result suggests that C-terminal domain of Spc2 contributes to regulating substrate selection for SPase.

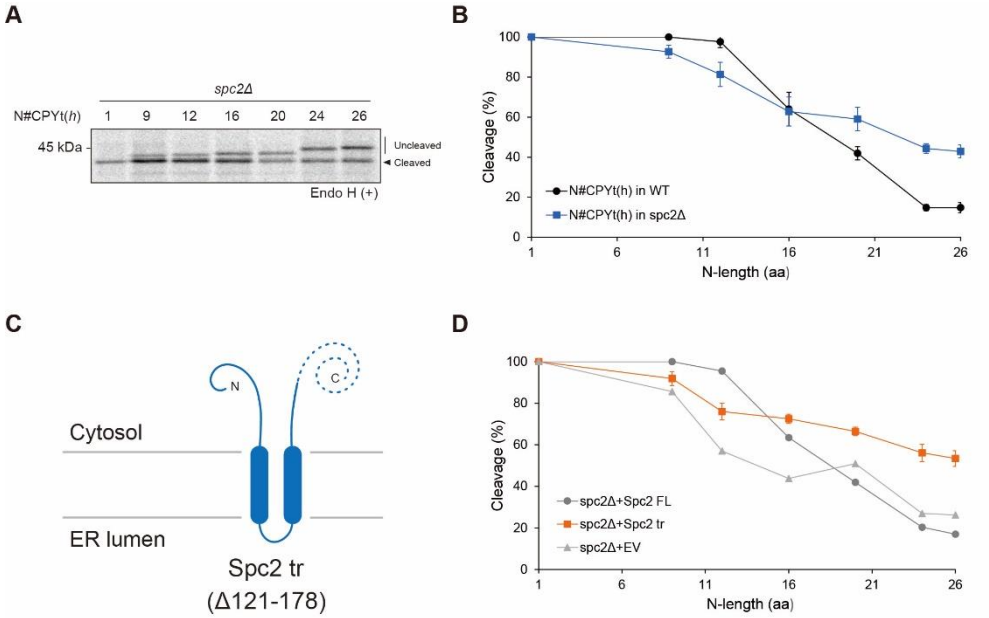

Figure 8. Large C-terminal domain of Spc2 is involved in regulating substrate selection. (A) *spc2*Δ cells expressing N#CPYt(*h*) variants were radiolabeled with [^{35}S]Met for 5 min at 30°C. To facilitate the band separation, protein samples were treated with Endo H. (B) *spc2*Δ cells expressing N#CPYt(*h*) were analyzed as in Fig. 5B. (C) Topology of Spc2 and its truncated form (Spc2 tr) is shown. (D) *spc2*Δ cells expressing one of full-length Spc2, truncated Spc2, and empty vector and N#CPYt(*h*) were analyzed as in Fig. 5B.

Two cleavage sites of N#CPYt are efficiently used by SPase

To check whether the changes in the substrate spectrum of SPase in *spc1*Δ and *spc2*Δ cells are due to altered recognition and/or usage of cleavage sites by SPase lacking Spc1 or Spc2, experiments were designed to investigate the cleavage sites of CPY variants in WT, *spc1*Δ and *spc2*Δ cells. When the signal sequence cleavage sites of CPY variants were searched with SignalP−5.0, two sites were predicted for all CPYt variants (including wild−type CPY) with equal probabilities (for CPYt(*h*), 0.492 for the upstream cleavage site, and 0.482 for the downstream (Fig. 9A). The upstream and the downstream cleavage sites were referred as Cleavage site 1 and 2, respectively, and residues around the Cleavage site 2 were denoted with '(e.g., −3', −1', Fig. 9B). To identify which cleavage site is used, Cleavage site 1 and 2 were selectively eliminated by single amino acid substitution at the Cleavage site 1 or 2 in N#CPYt(*h*) variants. Given that the canonical signal sequence cleavage sites follow the −3, −1 rule and proline (P) at +1 position with respect to the cleavage site inhibits signal sequence processing, a residue at −3 position in Cleavage site 1 was replaced with polar and bulky Glutamine (Q) and +1' position in Cleavage site 2 with P residue (Fig. 9B). Prediction by SignalP−5.0 showed a single cleavage site for each mutant, indicating that Q−3 and P+1' substitutions may selectively destroy Cleavage site 1 and 2, respectively (Figs. 10A and B). To confirm that Cleavage site 1 or 2 is selectively eliminated by Q−3 or P+1' substitution, N#CPYt(*h*) variants carrying the double mutation Q−

3/P+1' were prepared, and confirmed that cleavage no longer occurs. Cleavage of N#CPYt(*h*) variants possessing the cleavage site mutations was assessed by the 5 min pulse experiments and compared the data with the cleavage profile of N#CPYt(*h*) variants. Albeit slightly decreased cleavage of N16, N20 and N26CPYt(*h*) upon inhibition at the Cleavage site 1 (Q−3) or site 2 (P+1'), overall pattern of cleavage profile remained the same whether they contain both or Cleavage site 1 or 2 only, suggesting that SPase recognizes and uses both sites efficiently (Figs. 10C and D). On the other hand, double mutation Q−3/P+1' completely abolished signal sequence processing of all variants, indicating that only predicted two cleavage sites of N#CPYt(*h*) were recognized and used by SPase (Fig. 10C).

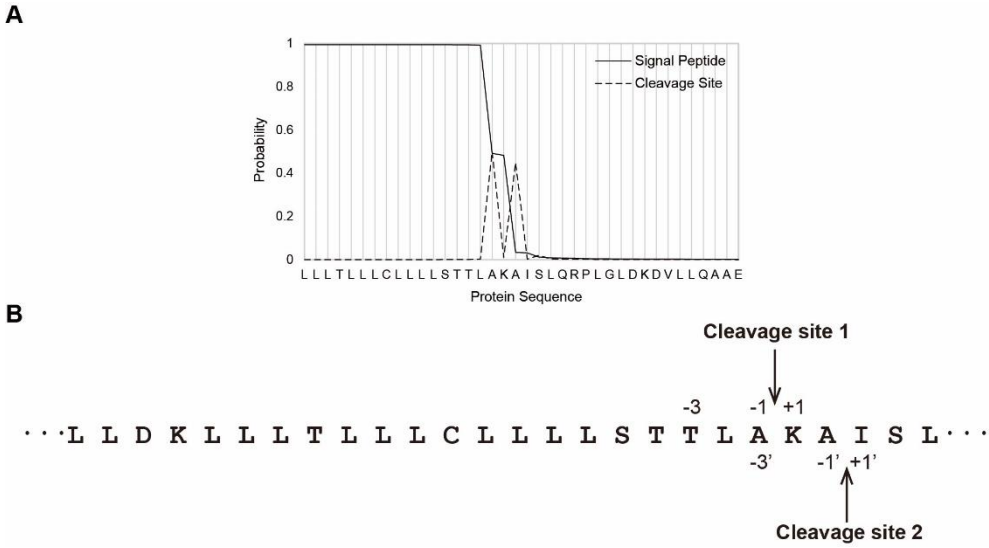

Figure 9. CPY variants have two cleavage sites. (A) Cleavage sites of CPYt(*h*) are predicted using SignalP-5.0 software (http://www.cbs.dtu.dk/services/SignalP/) [3] (B) Cleavage site 1 (upstream) and cleavage site 2 (downstream) are indicated with a downward arrow and an upward arrow, respectively.

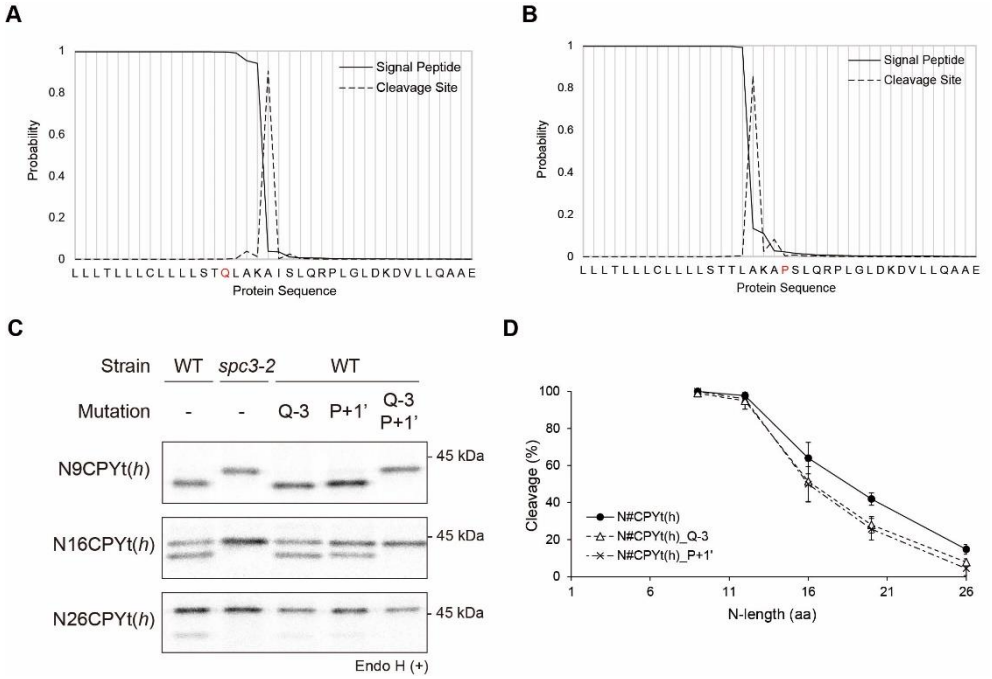

Figure 10. Two cleavage sites are efficiently used for processing. (A) and (B) Cleavage sites of CPYt(*h*) Q−3 and CPYt(*h*) P+1' are predicted as in Fig. 9A, respectively. Peak(s) of dashed line indicates predicted cleavage site(s). (C) Indicated cleavage site mutants of N#CPYt(*h*) variants in WT or *spc3−2* were radiolabeled for 5 min at 30°C (37°C for *spc3−2*), immunoprecipitated by anti−HA antibodies, subjected to SDS−PAGE, Endo H treatment and analyzed by autoradiography. (D) Cleavage (%) of the cleavage site mutants is analyzed.

Cleavage sites used by SPase are not changed with or without Spc1 and Spc2

To determine if SPase lacking Spc1 or Spc2 uses different cleavage sites for processing, signal sequence processing of the cleavage site variants was analyzed in *spc1*Δ or *spc2*Δ cells. Cleavage of N20 and N26CPYt(*h*) variants with Q−3 or P+1' mutation in *spc1*Δ cells increased compared to those in WT cells (Fig. 11A). These results indicate that recognition and usage of cleavage sites are unchanged for SPase with or without Spc1. However, signal sequence processing of N#CPYt(*h*) variants with P+1' mutation was largely impaired in the absence of Spc2, implying Cleavage site 2 is more preferred by SPase lacking Spc2, rather than using both cleavage sites equally (Fig. 11B). These results suggest that the signal sequence cleavage sites are less flexibly used in the absence of Spc2, which may be due to rigid positioning of the signal sequence within the ER membrane or poor recognition of possible adjacent cleavage sites by SPase without Spc2. Next, to determine whether SPase lacking Spc1 or Spc2 uses a non−canonical signal sequence cleavage site, the site without (−3, −1) rule for processing, N#CPYt(*h*) variants with Q−3/P+1' mutation which eliminated both canonical signal sequence cleavage sites were expressed in *spc1*Δ or *spc2*Δ cells and their cleavage was assessed. As in WT cells, no cleavage was detected for these sets of variants, indicating that the canonical signal sequence cleavage sites are still used by SPase in the absence of Spc1 or Spc2 (Fig. 11C). These results thus suggest that the

altered substrate range for SPase lacking Spc1 or Spc2 is not due to its processing in non-canonical cleavage sites.

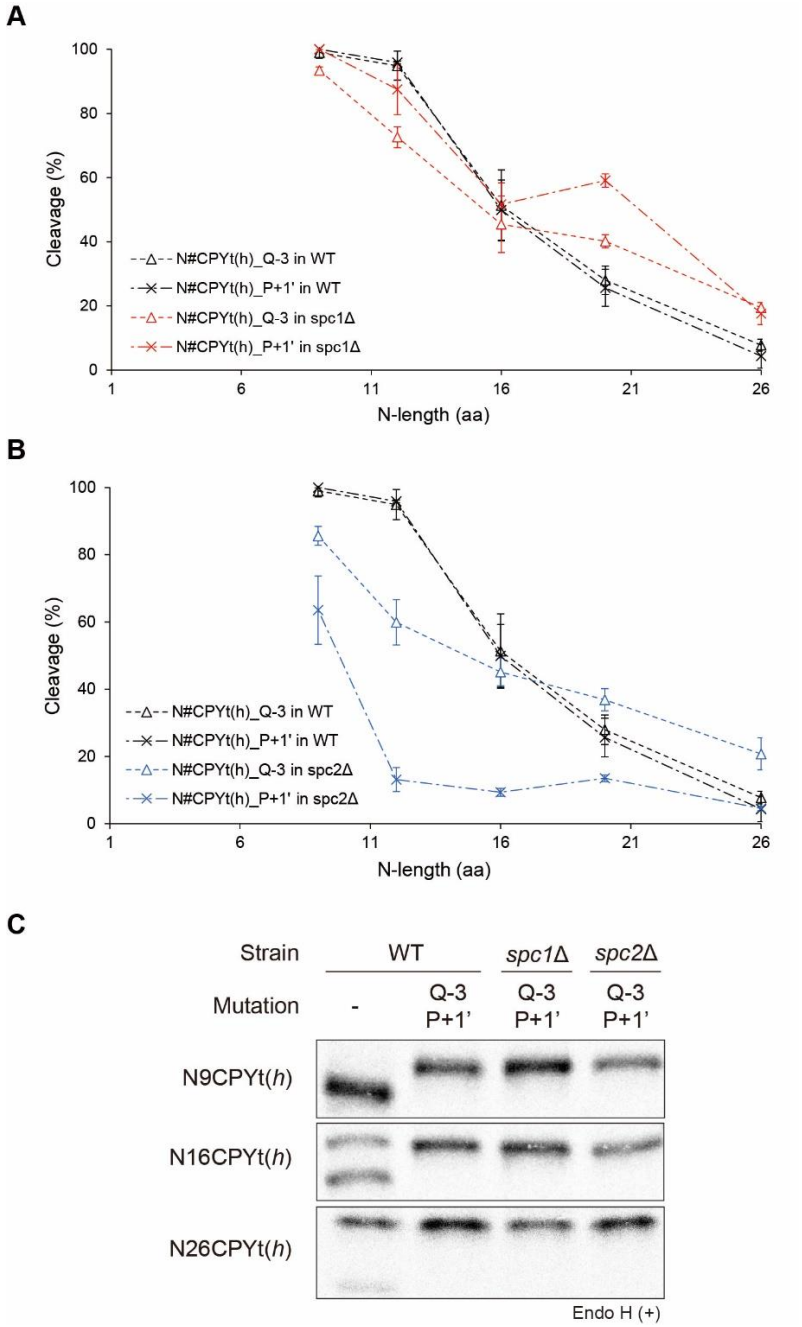

Figure 11. Flexible usage of the cleavage sites is largely impaired in cells lacking Spc2. Cleavage (%) of N#CPYt(*h*) variants with Q−3 or P+1' mutation in WT and *spc1*Δ cells (A) or *spc2*Δ cells (B) is compared. At least three independent experiments are carried out

and the average is shown with standard deviations. (C) Indicated N#CPYt(*h*) cleavage site mutants in WT, *spc1*Δ, and *spc2*Δ cells analyzed as in Fig. 10C.

Discussion

In this thesis, I undertook to study the mechanism by which SPase selects substrates and the roles of Spc1 and Spc2 in this process. For that, test proteins carrying systematically varied signal sequences that differ in N-length and hydrophobicity were prepared, and the SPase substrate spectrum was defined in WT cells. My data show that preferable substrates for SPase contain short N-length and less hydrophobic signal sequence. Next, the substrate spectrum in *spc1*Δ and *spc2*Δ cells was assessed to determine if the SPase substrate spectrum is changed in the absence of Spc1 or Spc2.

In the absence of Spc1, long N-length signal sequences, which were membrane anchored in the WT cells, became more susceptible to be cleaved by SPase. Hence, these results demonstrate that Spc1 functions in discriminating transmembrane segments from cleavable signal sequences. Previously, a genome-wide CRISPR/Cas9-based screen has identified SPCS1, a human homolog of Spc1, as a key host factor in flavivirus protein processing and infection [14]. Since viral proteins of flavivirus family are made as a polyprotein containing internal signal sequences and processed by host SPase and viral protease in the ER membrane, findings of the regulatory function of Spc1 in this study may provide new insight into the regulation of viral protein processing.

In *spc2*Δ cells, shorter N-length signal sequences became less efficiently cleaved while longer N-length signal sequences became more efficiently cleaved by SPase. This perplexing cleavage pattern in *spc2*Δ cells was also observed when the C-terminal domain of Spc2 (residues 121-178) was truncated, suggesting that a large cytosolic domain of Spc2 is critical in selecting substrates for SPase.

However, further experiments are required to confirm whether truncation of the C-terminal domain of Spc2 affects protein stability of Spc2 or complex formation.

Even though little is known about the role of Spc2 in regulating signal sequence cleavage, previous studies have shown that Spc2 is involved in protein translocation into the ER. Co-immunoprecipitation experiments with anti-Spc2 antibodies revealed that Spc2 associates with Sbh1 and Sbh2, the β subunits of the Sec61 complex and Ssh1 (a Sec61 homolog) complex, respectively [15]. More recently, Hosomi *et al*. have shown that the translocation of signal sequence-deficient proteins is facilitated in the absence of Spc2, suggesting that Spc2 affects the suppression of the signal sequence-independent protein translocation [16]. Therefore, the enigmatic signal sequence processing pattern of N#CPYt(*h*) shown in *spc2*Δ cells compared to that in WT cells may be related to the roles of Spc2 in protein translocation into the ER.

Finally, to check whether changes in substrate spectrum in *spc1*Δ and *spc2*Δ cells is due to altered SPase cleavage site, mutagenesis experiments were conducted. Test proteins originally possessing two cleavage sites were engineered to have only one cleavage site by mutation, and their cleavage pattern was assessed. When one of the cleavage sites was inhibited, the other cleavage site was alternatively used, suggesting flexible use of cleavage sites by SPase. Inhibition at one of cleavage sites resulted similarly in WT vs. *spc1*Δ cells.

However, when downstream cleavage site was inhibited, cleavage efficiency drastically decreased in *spc2*Δ cells especially for long N-length variants, indicating that flexible use of cleavage sites may be

impaired in the absence of Spc2. There are two potential explanations for these findings. One possibility is that deletion of Spc2 resulted in rigid positioning of the signal sequence within the ER membrane for processing by SPase. Accordingly, the downstream cleavage site is more exposed to the ER lumen rather than the upstream one. Another possibility is that SPase without Spc2 could not recognize possible adjacent cleavage sites properly, resulting in inefficient cleavage of the signal sequence.

Taken together, Spc1 is engaged in de-selecting transmembranwe segments for SPase, while Spc2 in regulating positioning of the signal sequence within the membrane. As such, findings in this thesis first demonstrate direct engagement of Spc1 and Spc2 in SPase substrate processing.

References

1. von Heijne, G., *Patterns of amino acids near signal-sequence cleavage sites.* Eur J Biochem, 1983. **133**(1): p. 17-21.
2. von Heijne, G., *How signal sequences maintain cleavage specificity.* J Mol Biol, 1984. **173**(2): p. 243-51.
3. Almagro Armenteros, J.J., et al., *SignalP 5.0 improves signal peptide predictions using deep neural networks.* Nat Biotechnol, 2019. **37**(4): p. 420-423.
4. Schmid, S.R. and M. Spiess, *Deletion of the amino-terminal domain of asialoglycoprotein receptor H1 allows cleavage of the internal signal sequence.* J Biol Chem, 1988. **263**(32): p. 16886-91.
5. Nilsson, I., P. Whitley, and G. von Heijne, *The COOH-terminal ends of internal signal and signal-anchor sequences are positioned differently in the ER translocase.* J Cell Biol, 1994. **126**(5): p. 1127-32.
6. Yim, C., et al., *Profiling of signal sequence characteristics and requirement of different translocation components.* Biochim Biophys Acta Mol Cell Res, 2018. **1865**(11 Pt A): p. 1640-1648.
7. van Dijl, J.M., et al., *Signal peptidase I of Bacillus subtilis: patterns of conserved amino acids in prokaryotic and eukaryotic type I signal peptidases.* EMBO J, 1992. **11**(8): p. 2819-28.
8. Evans, E.A., R. Gilmore, and G. Blobel, *Purification of microsomal signal peptidase as a complex.* Proc Natl Acad Sci U S A, 1986. **83**(3): p. 581-5.
9. Meyer, H.A. and E. Hartmann, *The yeast SPC22/23 homolog*

Spc3p is essential for signal peptidase activity. J Biol Chem, 1997. **272**(20): p. 13159-64.

10. Fang, H., et al., *The homologue of mammalian SPC12 is important for efficient signal peptidase activity in Saccharomyces cerevisiae.* J Biol Chem, 1996. **271**(28): p. 16460-5.
11. Mullins, C., et al., *Structurally related Spc1p and Spc2p of yeast signal peptidase complex are functionally distinct.* J Biol Chem, 1996. **271**(46): p. 29094-9.
12. Kitada, K., E. Yamaguchi, and M. Arisawa, *Cloning of the Candida glabrata TRP1 and HIS3 genes, and construction of their disruptant strains by sequential integrative transformation.* Gene, 1995. **165**(2): p. 203-6.
13. Fang, H., C. Mullins, and N. Green, *In addition to SEC11, a newly identified gene, SPC3, is essential for signal peptidase activity in the yeast endoplasmic reticulum.* J Biol Chem, 1997. **272**(20): p. 13152-8.
14. Zhang, R., et al., *A CRISPR screen defines a signal peptide processing pathway required by flaviviruses.* Nature, 2016. **535**(7610): p. 164-8.
15. Antonin, W., H.A. Meyer, and E. Hartmann, *Interactions between Spc2p and other components of the endoplasmic reticulum translocation sites of the yeast Saccharomyces cerevisiae.* J Biol Chem, 2000. **275**(44): p. 34068-72.
16. Hosomi, A., et al., *The ER-associated protease Ste24 prevents N-terminal signal peptide-independent translocation into the endoplasmic reticulum in Saccharomyces cerevisiae.* J Biol Chem, 2020. **295**(30): p. 10406-10419.

국문 초록

진핵생물의 소포체막 또는 원핵생물의 원형질막으로 들어가는 단백질은 대부분 N 말단에 위치하는 신호서열을 가진다. 신호서열의 C 말단에는 번역 시 혹은 번역 후 전구단백질의 막을 통한 전좌가 일어날 때 진화적으로 보존된 신호서열절단효소가 자르는 절단 위치가 존재한다. 잘리지 않은 신호서열은 막 안으로 들어가 막단백질의 막 투과 부분이 된다. 하지만 신호 서열의 다양성으로 인해, 신호서열절단효소의 기질 선택 기전은 잘 알려지지 않았다. 진핵세포에서 신호서열절단효소는 소포체막에 존재하며 다른 막단백질들과 복합체를 이루고 있다. 효모에서 신호서열절단효소 복합체는 Sec11, Spc1, Spc2, Spc3로 이루어져 있고, 그 중 Sec11은 촉매작용을 하며, 나머지 요소들의 역할은 잘 알려져 있지 않다.

이 논문에서는 Spc1과 Spc2의 역할을 연구하기 위해 먼저 다양한 소수성과 N 말단 길이의 신호서열을 가지는 모델 단백질 세트를 만들어 세포 내에서의 신호서열 절단 분석 방법을 마련하였다. 이 분석 방법을 통해 신호서열절단효소 기질의 범위를 정의할 수 있었다. 그 결과, 신호서열절단효소는 짧은 N 말단과 덜 소수성인 신호서열을 가지는 단백질을 기질로 선호한다는 것을 알 수 있었다. 그 다음, *spc1Δ*과 *spc2Δ* 세포에서의 기질 스펙트럼을 WT 세포와 비교해 보았다. WT 세포에서는 막에 고정되어 잘리지 않는 긴 N 말단 길이의 신호서열이 Spc1이 없는 세포에서는 더 잘리기 쉽게 되었다. 반면, Spc2가 없는 세포에서는 짧은 N 말단 신호서열은 덜 효율적으로 잘리고, 긴 N 말단 신호서열은 더 효율적으로 잘리게 되었는데, 이는 Spc2가 없을 때, 잘려야 할 신호서열과 잘리지 않아야 할 신호서열의 사이의 구분이 둔화되는 것을 의미한다.

이러한 결과는 Spc1 또는 Spc2가 없을 때, 막에 고정되는 신호서열이 더 잘 잘리게 됨을 보여준다. 또한 돌연변이 생성 실험을 통해 CPY 변

종들의 원래의 절단 위치는 Spc1 또는 Spc2가 없을 때 변하지 않는 것을 확인하였다. 이는 Spc1과 Spc2가 신호서열절단효소의 기질 선택을 조절하는데 관여함을 보여주며, 막단백질 생성에서의 역할을 암시한다.

주요어: 신호서열, 신호서열절단효소, Spc1, Spc2, 막단백질
학 번: 2018-22166

감사의 글

먼저, 지난 석사 학위 과정동안 부족한 저를 가르쳐 주신 김현아 교수님께 감사의 말씀을 드립니다. 저에게 피가 되고 살이 되는 교수님의 진심어린 조언과 보살핌 덕분에 많이 배우고 성장할 수 있었습니다. 또한, 이번에 제 석사 학위 논문 심사를 맡아 주시고 연구와 실험에 대한 귀한 조언을 아낌없이 주신 조형택 교수님과 안광석 교수님께도 정말 감사드립니다.

지난 2년간 저에게 실험을 가르쳐준 채원 언니, 정말 고맙습니다! 지금 생각해보면 실험실에 들어와서 언니에게 붙어 다니며 실험을 처음 배울 때가 대학원 생활 중 가장 즐겁고 설렜던 것 같아요. 더불어, 실험실 생활 동안 저에게 많은 도움을 주었던 성준 선배, 성민 선배, 서은 언니, 윤제 오빠, 서윤 언니, 미강이를 비롯한 실험실의 다른 분들도 그동안 감사했습니다.

마지막으로 저의 소중한 가족들에게도 고맙다는 말을 꼭 전하고 싶습니다. 정직하고 좋은 연구를 하는 사람이 되겠습니다. 감사합니다.