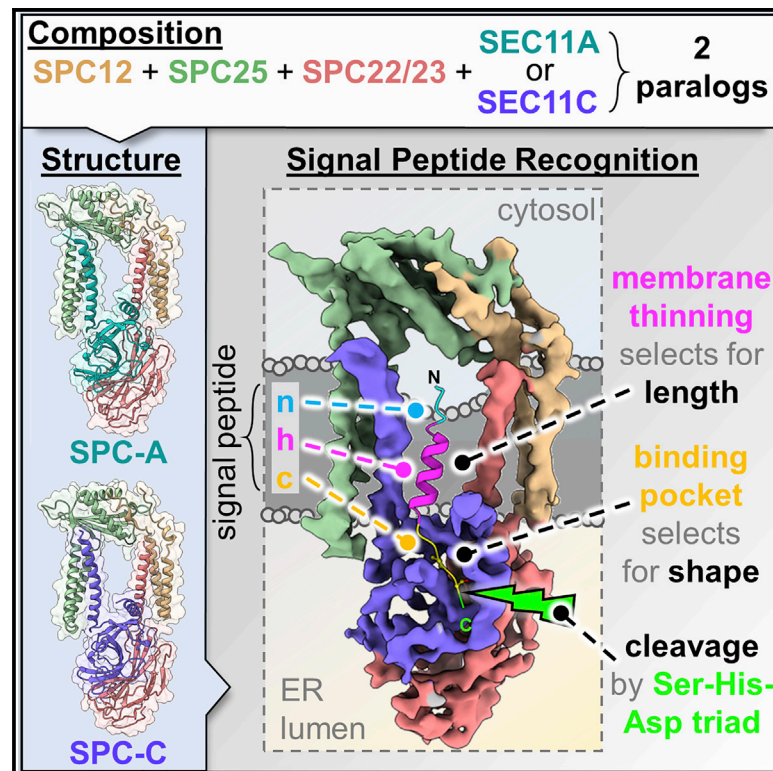# Structure of the human signal peptidase complex reveals the determinants for signal peptide cleavage

Graphical abstract

Authors

A. Manuel Liaci,
Barbara Steigenberger,
Paulo Cesar Telles de Souza, ...,
Siewert J. Marrink,
Richard A. Scheltema,
Friedrich Förster

Correspondence

f.g.forster@uu.nl

In brief

The ER-resident signal peptidase complex (SPC) cleaves signal peptides of nascent secretory proteins and leaves transmembrane helices intact. Combining electron cryo-microscopy, mass spectrometry, and molecular dynamics simulations, Liaci et al. discover that the human SPC uses local membrane thinning and shape complementarity near the active site to generate substrate selectivity.

Highlights

- The human SPC has two paralogs, both of which are structurally characterized here

- The SPC is a serine protease with a catalytic Ser-His-Asp triad

- The c-region binding pocket is conserved to bacterial signal peptidases

- Membrane thinning by the SPC measures signal peptides for length before cleavage

Article

Structure of the human signal peptidase complex reveals the determinants for signal peptide cleavage

A. Manuel Liaci,[1] Barbara Steigenberger,[2,3,7] Paulo Cesar Telles de Souza,[4,5,7] Sem Tamara,[2,3] Mariska Gröllers-Mulderij,[1] Patrick Ogrissek,[1,6] Siewert J. Marrink,[4] Richard A. Scheltema,[2,3] and Friedrich Förster[1,8,*]

[1]Structural Biochemistry, Bijvoet Centre for Biomolecular Research, Utrecht University, Universiteitsweg 99, 3584 CG, Utrecht, the Netherlands
[2]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Centre for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH, Utrecht, the Netherlands
[3]Netherlands Proteomics Centre, Padualaan 8, 3584 CH, Utrecht, the Netherlands
[4]Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Material, University of Groningen, Nijenborgh 7, 9747 AG, Groningen, the Netherlands
[5]Molecular Microbiology and Structural Biochemistry, UMR 5086, CNRS and University of Lyon, Lyon, France
[6]Institute of Chemistry and Metabolomics, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany
[7]These authors contributed equally
[8]Lead contact
*Correspondence: f.g.forster@uu.nl
https://doi.org/10.1016/j.molcel.2021.07.031

SUMMARY

The signal peptidase complex (SPC) is an essential membrane complex in the endoplasmic reticulum (ER), where it removes signal peptides (SPs) from a large variety of secretory pre-proteins with exquisite specificity. Although the determinants of this process have been established empirically, the molecular details of SP recognition and removal remain elusive. Here, we show that the human SPC exists in two functional paralogs with distinct proteolytic subunits. We determined the atomic structures of both paralogs using electron cryo-microscopy and structural proteomics. The active site is formed by a catalytic triad and abuts the ER membrane, where a transmembrane window collectively formed by all subunits locally thins the bilayer. Molecular dynamics simulations indicate that this unique architecture generates specificity for SPs based on the length of their hydrophobic segments.

INTRODUCTION

Approximately one-quarter of the human proteome uses the secretory pathway (Palade et al., 1975; Uhlén et al., 2015), including key proteins for cell signaling, oxygen and nutrient transport, and the immune system. Many secretory pathway proteins are targeted to the endoplasmic reticulum (ER) via a short N-terminal hydrophobic helix called a signal peptide (SP) (Blobel and Dobberstein, 1975). Nascent SPs emerge from the ribosome and target the ribosome-nascent-chain complex to the ER membrane, where it is inserted into the protein-conducting channel Sec61 (Gemmer and Förster, 2020). For many proteins (approximated to exceed 3,000 different physiological protein substrates in humans (Uhlén et al., 2015)), the signal peptidase complex (SPC) cleaves off the SPs from their non-functional pre-forms. The SPC also facilitates the maturation of many viral proteins, including pre-proteins from most flaviviruses (e.g., Zika, Dengue, and Hepatitis C virus), HIV, and SARS coronavirus (Estoppey et al., 2017; Oostra et al., 2007; Snapp et al., 2017; Suzuki et al., 2013; Zhang et al., 2016).

The human SPC comprises the accessory proteins SPC12 (SPCS1), SPC22/23 (SPCS3), and SPC25 (SPCS2) and the two proteolytic subunits SEC11A (SPC18) and SEC11C (SPC21) (Figure 1A; Evans et al., 1986). It is currently unclear whether both proteolytic subunits occur in the same complex or form distinct SPC paralogs (Shelness and Blobel, 1990). Both SEC11A and SEC11C have low, but significant, sequence similarity to bacterial signal peptidases (SPases) (van Dijl et al., 1992), which are monomeric and characterized by a Lys-Ser catalytic dyad (Paetzel et al., 1998; Tschantz et al., 1993). In contrast, eukaryotic SPCs have the active-site lysine replaced with a histidine and might function through either a catalytic His-Ser dyad or an Asp-His-Ser triad (VanValkenburgh et al., 1999), leading to the functional distinction of prokaryotic P-type SPases and ER-type SPases (Paetzel et al., 2002).

The SPC is highly selective for SPs, but the molecular mechanism of SP recognition is largely unexplored. Consequently, SPs are typically predicted using empirical features (Nielsen et al., 2019). SPs are characterized by three distinct regions: (1) an often positively charged, unfolded n-region; (2) a hydrophobic,
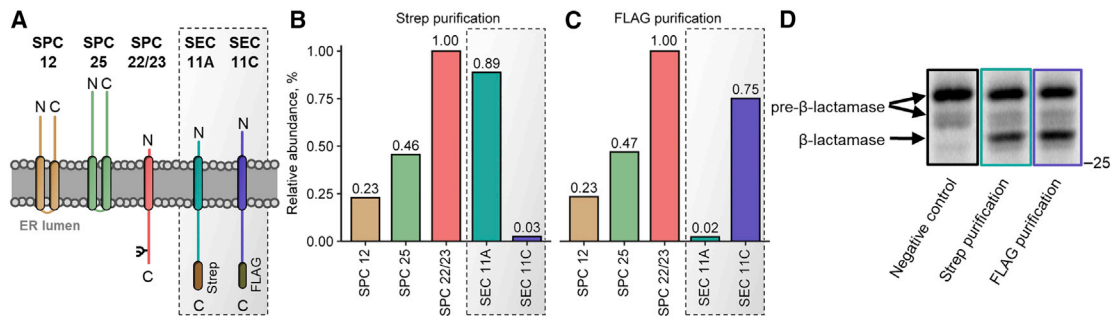
Figure 1. Human SPC exists in two paralogs

(A) Overview of the SPC subunits. The non-proteolytic subunits SPC12 (yellow), SPC25 (green), and SPC22/23 (red) were co-expressed with SEC11A-Strep (teal) and SEC11C-FLAG (purple).

(B and C) Top-down MS quantification of the subunits after Strep (B) or FLAG (C) affinity purification. Abundance normalized to SPC22/23.

(D) Pre-β-lactamase *in vitro* cleavage assay. Negative control, no SPC added. Irrelevant areas of the gel were omitted for clarity; the position of the 25 kDa marker is indicated.

See also Figure S1.

α-helical h-region; and (3) a polar c-region, which contains the scissile bond (von Heijne, 1990). The n-region determines the orientation of the SP in the protein-conducting channel Sec61 and, hence, the membrane protein topology (von Heijne, 2006). The h-region of SPs is invariably hydrophobic and notably shorter than regular transmembrane (TM) helix segments (7–15 amino acids) (Nilsson et al., 1994). The c-region is typically 3–7 amino acids long and contains two crucial positions relative to the scissile bond (−1 and −3), which need to be occupied by small, non-charged residues.

We reconstituted the human SPC and analyzed it by cryo-electron microscopy (cryo-EM) single-particle analysis, coarse-grained molecular dynamics (CGMD) simulations, and structural proteomics-driven mass spectrometry (MS) to elucidate its precise stoichiometry, structure, and mechanism of SP recognition and cleavage.

RESULTS

Human SPC exists in two paralogs

SPC12, SPC22/23, and SPC25 can be found in essentially all eukaryotes, suggesting they have evolved at the advent of eukaryotic life (Figure S1). In most eukaryotic organisms, the SPC consists of only these three subunits and one copy of SEC11. In animals, a duplication event of SEC11 occurred approximately 400 million years ago. SEC11A and SEC11C remained closely related throughout evolution, with ~80% sequence identity in humans. Both genes can individually substitute for yeast SEC11 and even some bacterial SPases functionally (Liang et al., 2003).

To determine whether SEC11A and SEC11C are part of one complex or two distinct paralogous SPCs, we co-expressed the three accessory subunits SPC12, SPC25, and SPC22/23 with Strep-tagged SEC11A and FLAG-tagged SEC11C in HEK293 cells and purified the complexes by either Strep or FLAG affinity chromatography from the same batch of cells (Figure 1). In both cases, we recovered near-stoichiometric amounts of the accessory subunits and the respective tagged SEC11

variant, whereas the other variant was 30–40 times less abundant as determined by top-down MS (Figures 1B and 1C). Both isolates were able to process pre-β-lactamase *in vitro* with similar efficiencies (Figure 1D). We conclude that, in humans and likely in other eukaryotes with two SEC11 paralogs, two functional hetero-tetrameric SPC paralogs exist formed by SPC12, SPC22/23, SPC25, and either SEC11A or SEC11C. In the following sections, we refer to the two paralogous complexes as SPC-A and SPC-C, respectively.

SPC architecture and topology

We determined the structures of both human paralogs, solubilized in amphipol PMAL-C8, using single-particle cryo-EM to an overall resolution of approximately 4.9 Å (Figures 2, S2, and S3; Table 1; Video S1). The low protein mass of the hetero-tetrameric complex (84 kDa, 17 of which are unordered) and the structural variations of the micelle likely limited particle-alignment accuracy and attainable resolution (Herzik et al., 2019). Initial atomic subunit models generated by trRosetta (Yang et al., 2020) yielded excellent fits to the two cryo-EM densities (Figure S3). Using these initial models and the EM maps, we could build atomic models of both SPCs that explain all of the observed density (Figures 2C, 2D, S3, and S4). The models agree with the previously determined TM topologies of the subunits (Kalies and Hartmann, 1996; Shelness et al., 1993) as well as predictions of TM helices, secondary structure and disordered segments, and atomistic molecular dynamics (MD) simulations (Figure S4). When mapping the distance restraints obtained by cross-linking (XL)-MS onto the atomic models, we found that ~80% of the cross-links range within the maximum allowed distance of the PhoX crosslinker (Steigenberger et al., 2019; Figure S5).

In the structures, SEC11A or SEC11C, respectively, interacts with SPC22/23 to form a globular luminal body consisting solely of beta sheets (Figures 2C and 2D). The TM domains of SPC25 and SEC11A/C on one side and SPC12 and SPC22/23 on the opposite side form two distinct three-helix bundles, which frame a characteristic ~15-Å-wide, lipid-filled "TM window" in the
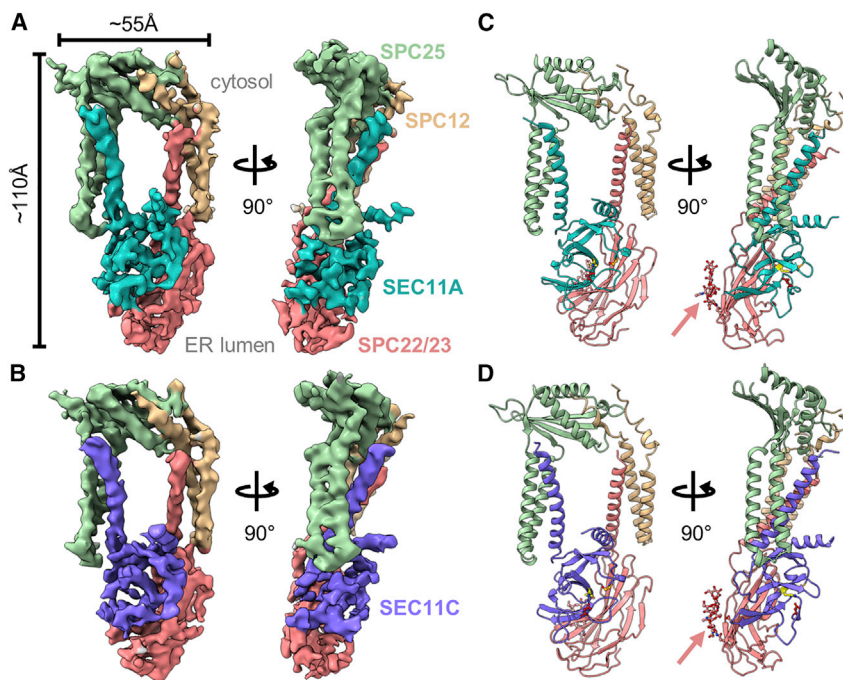
partial N-terminal truncation (all subunits, except for SPC22/23 and SEC11A).

Characterization of the accessory subunits

The luminal domain of SPC22/23 forms an extended beta-sandwich with a fold similar to that of the histone chaperone ASF1 (Daganzo et al., 2003), which embraces the catalytic core of SEC11 (Figures 2C and 2D). This arrangement suggests that SPC22/23 helps to stabilize and position the active center close to the luminal membrane surface, explaining why it is required for catalytic function (Fang et al., 1997; Kamitani et al., 1989; Meyer and Hartmann, 1997). In our sample derived from HEK293 cells, we found 98% of SPC22/23 molecules to be N-glycosylated at Asp141, which is considerably higher than reported for dog pancreatic microsomes (Evans et al., 1986). Top-down MS identified this glycan as a homogeneous biantennary mannose-type structure (Figures 3F and 3J). It is also partially resolved in the EM map and projects toward the membrane.

SPC12 and SPC25 are not essential for catalytic activity (Mullins et al., 1996), but deletion of SPC25 in yeast results in a 2-fold reduction of in vitro SPase activity (Antonin et al., 2000). In our structures, SPC25 accounts for most of the ordered density in the cytosolic portion of the SPC. The protein adopts a novel alpha-beta-sandwich fold, which is interspersed by the two TM helices that interact with SEC11A/C. The N-terminal 50 amino acids of SPC25 are missing from the density but are detected by various MS approaches. In accordance with previous reports (Van Damme et al., 2012), the removed starting methionine at the N terminus of SPC25 is replaced by an N-acetylation. In addition, a subset of SPC25 molecules is N-terminally processed (Figures 3E and 3I).

SPC12 is the only subunit that does not directly interact with SEC11 (Figures 2C and 2D), explaining why it was found to be the least important for catalytic activity (Antonin et al., 2000; Fang et al., 1996). Its cytosolic termini are largely flexible (residues 1–65 and 152–169), and only its membrane-proximal parts constantly interact with SPC25, as supported by XL-MS data (Figure S5). SPC12 exhibits minor N-terminal processing as revealed by top-down MS, along with a low-stoichiometric phosphorylation, which is likely located on the cytosolic portion of the complex (Figures 3G and 3K).

membrane. The cytosolic portion of the complex is formed mainly by SPC12 and SPC25. Together, these two subunits form a clamp-like structure that orients the TM segments of SEC11A/C and SPC22/23.

Consistent with this architecture, native MS shows detectable SEC11A/C-SPC22/23 sub-complexes, which dissociate at comparable activation energies (Figure S6; Table S1). Upon gas-phase activation, SPC dimers overall exhibit comparable dissociation stabilities, whereby SPC-A exhibits a slightly more stable dimer with ~45% of total ion intensity, corresponding to the ejected subunits compared with ~65% for SPC-C at the same dissociation conditions. Additionally, SEC11A has a more stable conformation within the dimeric complex compared with SEC11C, which is estimated based on the intensity-weighted retained charge of an ejected subunit. In the case of SPC-A, SEC11A and SPC22/23 retain ~32% and ~53% of their original precursor charges, respectively, upon dissociation; in the case of SPC-C, SEC11C and SPC22/23 retain ~44% and ~36% of their original precursor charges, respectively. The higher average charge of dissociated SEC11C is indicative of a more-extended conformation, i.e., a higher degree of unfolding (Jurchen and Williams, 2003; Popa et al., 2016), compared with SEC11A. SPC25 and SPC12 were detected only in a free form, likely because removal of the amphipol affected their binding interfaces. The cryo-EM density explains ~80% of the SPC residues, whereby most of the unresolved residues are mapping to the N termini of SPC12 and SPC25 (Figure S4). We detected the bulk of the unmapped N- and C-terminal regions of SEC11A/C, SPC25, and SPC12 by shotgun and top-down MS, which confirms that they are structurally flexible, rather than proteolytically removed (Figure 3). The terminal stretches of the SPC harbor different post-translational modifications (PTMs), such as phosphorylation (SPC12), N-terminal acetylation (SPC25) and

Table 1. Data collection and refinement statistics, related to Figure 2

	SPC-A	SPC-C
Data collection and processing		
Microscope	Talos Arctica	Talos Arctica
Camera	K2 summit	K2 summit
Magnification	165,000	165,000
Voltage (kV)	200	200
Electron exposure (e$^-$/Å^2)	60	60
Defocus range	0.5–4.0	0.5–4.0
Pixel spacing (Å)	0.81	0.81
Symmetry imposed	C1	C1
Final Nr. particle images	29,508	60,598
Map resolution	4.9	4.9
FSC threshold	0.143	0.143
Map sharpening B factor (Å^2)	−180	−180
Model validation		
MolProbity score	1.7	1.6
Clashscore	4.2	4.4
Rotamer outliers (%)	0.2	0.0
Ramachandran plot		
Favored	91.9	95.0
Allowed	100.0	100.0
Outliers	0.0	0.0
Real-space correlation	0.68	0.73
Mean model B factor (Å^2)	152	232

Characterization of SEC11 and the SP c-region binding pocket

The luminal SEC11A/C portion adopts a type-I SPase fold that aligns well with the catalytic core domain of *Escherichia coli* SPase I (Paetzel et al., 1998; Figures 4A and S7). P- and ER-type SPases exhibit low, but significant, sequence similarity in interspersed, conserved sequence stretches, which are commonly referred to as boxes A–E (Paetzel et al., 2002; Figures S7A and S7B). The catalytic residues Ser56/68 (in box B, numbered as in SEC11A/C, respectively) and His96/108 (in box D) are located at highly similar positions as the SPase I Ser-Lys dyad. In contrast to the P-type SPases, it has previously been suggested that ER-type SPases might function through a catalytic Ser-His-Asp triad because SEC11 has three conserved aspartic acid residues, Asp116/128, Asp121/133, and Asp122/134 (all box E), which might complete the active center (VanValkenburgh et al., 1999; Figures 4A and S7C). Our structures show that all three aspartic acid residues are located proximal to the binding pocket, with Asp122/134 best positioned to potentially complete the triad. The models suggest that Asp116/128 points toward the protein core and engages in a salt bridge with the equally conserved Arg97/109, analogous to structures of P-type SPases (Paetzel et al., 1998; Ting et al., 2016). Because the map resolution is insufficient to model side chains reliably, we mutated all three candidate aspartic acids and tested how they affect catalytic activity and protein stability *in vitro* (Figures 4B, S7D, and S7E). Mutating Asp122/134 had only a moderate

effect on protein stability, whereas it completely abolished catalytic activity. As expected, mutating Asp116/128 had a stronger effect on protein stability, but the enzyme complex retained catalytic activity to a reasonable extent, whereas mutating Asp121/133 had little effect on either SPC stability or activity. We, thus, conclude that human SEC11A/C indeed function via a catalytic triad consisting of Ser56/68, His96/108, and Asp122/134.

To model the c-region of SPs in the SPC, we superposed SEC11A/C and *E. coli* SPase I in complex with the lipopeptide inhibitor arylomycin (Paetzel et al., 2004) as a template for the c-region (Figures 4C and S7F–S7H). As in its bacterial counterpart, the catalytic residues of SEC11A/C reside at the end of a shallow, hydrophobic groove that is lined by a β strand formed by box-D residues. In the bacterial enzyme (and most other proteases), the c-region of the SP is forced into a β-strand conformation (Paetzel, 2014; Tyndall et al., 2005). The substrate side chains at the −1 and −3 positions point toward shallow hydrophobic pockets that can only accommodate small hydrophobic residues (Paetzel, 2014). The same principle likely applies to SEC11A/C and provides an explanation for the empirically established c-region consensus motif and the interchangeability of bacterial and eukaryotic SP c-regions (Figures S7F–S7H).

The TM region was not resolved in structures of bacterial SEC11 homologs (Paetzel et al., 1998). In addition to a single, N-terminal TM helix, both SEC11A and SEC11C possess a striking amphiphilic helical segment formed by residues near the C terminus. We termed this amphiphilic helix at the interface between the membrane and the ER lumen the "bowsprit helix" because it prominently projects from the binding pocket (Figures 4E and 4F). The N- and C-terminal stretches of SEC11A/C, which harbor most of the single amino acid variations, are flexible in our structures.

The SPC induces membrane thinning

Inside the lipid-filled TM window, the diameter of the amphipol micelle is reduced to approximately 23 Å compared with the 35–40 Å in the exterior that are comparable to an average mammalian ER membrane (Mitra et al., 2004) (Figures 4E and 4F). The SPC structure suggests that several factors synergistically induce membrane thinning: (1) on the cytosolic face of the SPC, the sides of the three-helix bundles that frame the TM window have notably shorter hydrophobic cores than do those facing the surrounding membrane, and they are notably positively charged at their cytosolic ends (Figures 4F and 4G); (2) on the luminal face of the window, a range of membrane-proximal residues contribute to a negative charge (Figure 4G); whereas (3) SEC11A/C forms a hydrophobic ridge that presses tightly against the membrane and partially inserts itself into the hydrophobic environment (Figures 4E and 4F). The SEC11 bowsprit helix is prominently positioned on the micelle surface, suggesting that it contributes to shaping the membrane surrounding the binding pocket. The thinning of the micelle also occurs when the SPC is solubilized in digitonin (Figure 4H).

Coarse-grained and atomistic MD simulations confirm that the thinning is also present in simple and complex lipid membranes (Figure 5; Video S3; Table S2). For instance, in a complex ER-like membrane, we observed an average thinning of 26%, with fluctuations between 15% and 46%. As a consequence, the TM
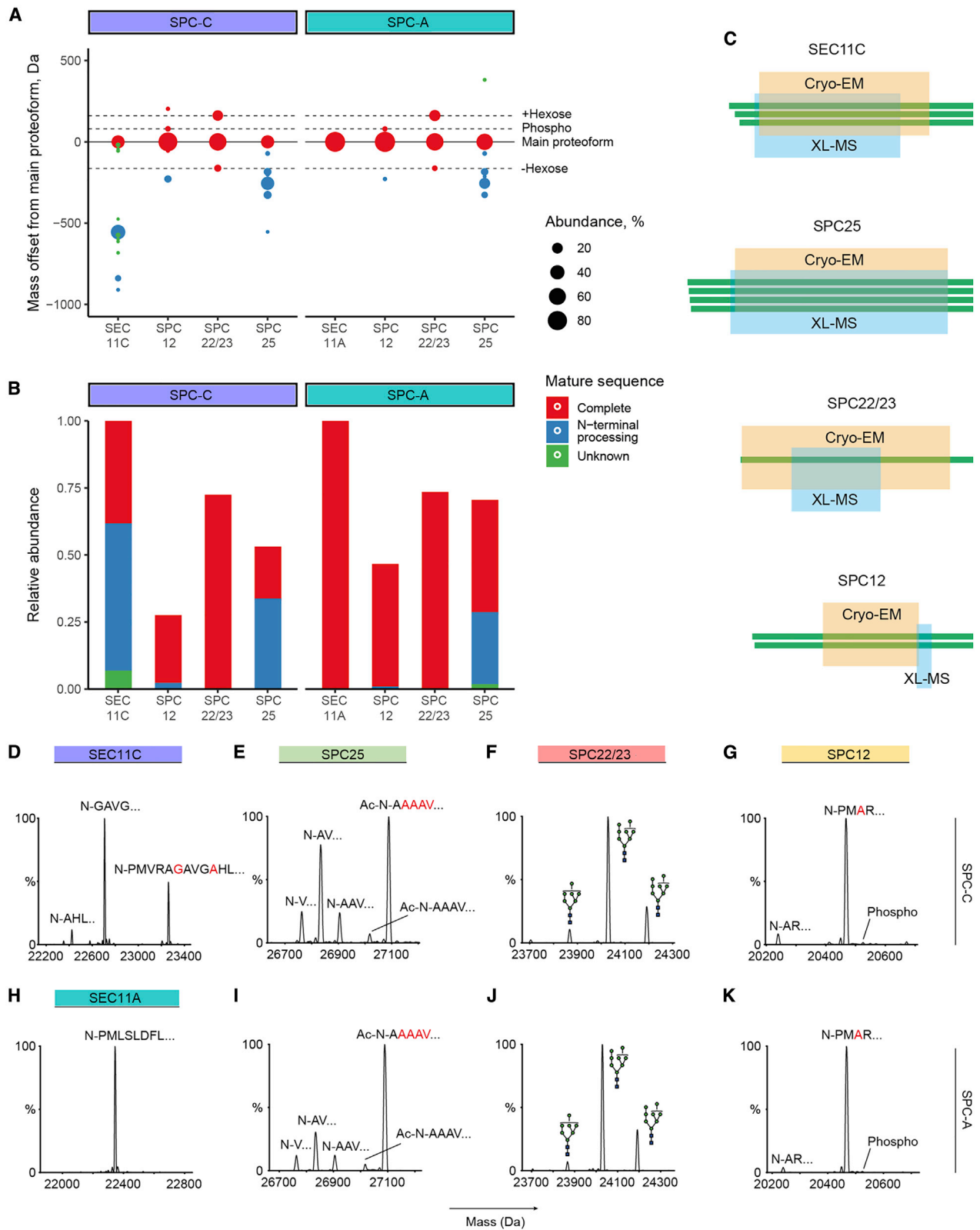
(legend on next page)

window seems enriched with lipids that usually form thin membranes (van Hilten et al., 2020)—especially unsaturated phosphatidylcholine lipids—which spread their acyl chains to squeeze into the window.

The entire c-region of SPs measures five to seven amino acids on average (Figures 6B and 6C), which fits the distance from the active site to the thinnest point of the amphipol micelle located right above the SP binding groove (Figure 4C). At this position, the micelle diameter (~23 Å) coincides with the length of a typical h-region (~11 amino acids, Figures 6A–6C). Thus, the thinned lipid nano-compartment formed by the SPC appears well suited to host the short SP h-region.

Membrane thinning ensures selectivity of the SPC

Early statistical analyses of known SPs suggested that h-regions rarely exceed 16 residues, making them notably shorter than TM helices are (von Heijne, 1985, 1990; Nilsson et al., 1994). Based on the much larger set of SPs deposited on the UniProt database (UniProt Consortium, 2021) to date, we used SignalP 3.0 (Bendtsen et al., 2004) to assign n-, h-, and c-regions of all experimentally verified eukaryotic SPs longer than 22 amino acids (Figures 6B and 6C; Liaci and Förster, 2021). High-confidence assignments were obtained for 921 SPs (62% of the analyzed data). The assignment reveals that the bulk of length variability in this set of relatively long SPs originates from the n-region, whereas the length of the h- and c-regions is relatively constant. In fact, an average and median h-region length of 11 residues and a c-region length of six residues (with standard deviations of 1.1 and 1.4 amino acids, respectively) are conserved across the eukaryotic kingdom. For SPs with high-confidence assignments, the h-region length varied between 8 and 14 residues. The same average values were obtained for a subset with low-prediction probabilities (containing six SPs with h-regions of 17–18 residues). Even a manual curation of all human cases to include the maximum possible h-region length (see Method details) resulted in an average and median h-region length of 13 residues, with only nine cases of more than 18 residues. When comparing the predicted mean and median length of h-regions to predictions for non-cleaved, single-pass TM helices (mean > 20 amino acids) (Sharpe et al., 2010), it becomes clear that h-regions are substantially shorter.

Cell-free assays monitoring the cleavage of model SPs in the ER membrane as a function of the length of poly-leucine h-regions revealed that the eukaryotic SPC does not cleave beyond an h-region exceeding 18–20 amino acids (Nilsson et al., 1994, 2002), consistent with our statistical analysis of eukaryotic SPs. We hypothesized that the membrane thinning induced directly above the c-region binding pocket by the TM window may be the cause for this effect.

Following this reasoning, the shorter h-regions of SPs should be able to diffuse into the window, whereas TM helices of 18–20 residues or more should be excluded by the reduced bilayer thickness. To test that hypothesis, we effectively recapitulated the assays (Nilsson et al., 1994, 2002) in silico: we placed model SPs with h-regions composed of 11 (L11) or 20 (L20) leucine residues near the binding pocket and performed coarse-grained molecular dynamics (CGMD) simulations to evaluate whether the respective peptides can diffuse into the window or remain excluded over time (Figures 6D–6H; Video S4). We simulated both peptides five times for 20 μs. Indeed, although L11 readily diffused into the TM window and, in some instances, even placed its c-region into the c-region-binding pocket, L20 usually diffused away and could not access the pocket. The combined samplings of the CGMD simulations clearly indicate the stability of the short L11 SP in the TM window in contrast to the long L20 peptide.

Taken together, the structure, existing biochemical evidence (Nilsson et al., 1994, 2002), and CGMD simulations strongly suggest that the SPC measures the length of its substrates' h-regions by altering the lipid nano-environment around its SP binding site, along with limited shape complementarity for the −1 and −3 positions in the c-region-binding pocket (Figure 6I). This "molecular ruler" excludes hydrophobic helices greater than a threshold of 18–20 amino acids.

Comparison of the two paralogs

In an attempt to obtain further clues on substrate specificity of SPC-A and SPC-C, we determined the relative abundance of SEC11A and SEC11C in a number of common cancer cell lines (Figure 7A). In all these cell lines, SEC11A is highly expressed, whereas the level of SEC11C is below the detection limit. Nevertheless, SEC11C is reported to be ubiquitously expressed in many tissues at similar levels as SEC11A (Bastian et al., 2008).

The maps for SPC-A and SPC-C are virtually indistinguishable at the current resolution (Figures 7B–7E; Video S1). The residues mapping to the N-terminal SEC11 TM helix and the SP binding groove are completely conserved, whereas the sequence variations (located mostly on surface-exposed loops on the periphery of the SPase domain) do not result in significant structural differences (Figures 7F–7H). However, there are substantial sequence differences in the flexible N- and particularly the C-terminal stretches of SEC11A and SEC11C (Figures 7G and 7H). The cytosolic N-termini of both SEC11A and SEC11C are predicted

---

Figure 3. SPC mature sequences and proteoforms

(A) Proteoform mass offsets determined for SPC subunits. Each offset is defined by the mass difference between a mass of a proteoform and the primary proteoform. The most prominent PTM-related mass shifts are highlighted with dashed lines; main proteoforms are shown with a solid line. Main proteoforms: SEC11A and SPC12 are unmodified, SPC25 is N-acetylated, SEC11C is N-terminally processed, and SPC22/23 is N-glycosylated. Circle size represents the fractional abundance of a proteoform per subunit.

(B) Abundances of SPC subunits demonstrating the contribution of complete, N-terminally processed, and undetermined sequences.

(C) Sequences of distinct SPC proteoforms in the context of sequence coverages achieved with cryo-EM (orange) and XL-MS (blue).

(D–K) Mass profiles displaying the primary proteoforms of SPC subunits for SPC-C (D–G) and SPC-A (H–K) complexes. For subunits with N-terminal sequence processing, a few N-terminal amino acids are displayed above the corresponding peaks. SPC25 was identified in N-acetylated and in N-terminally truncated, non-acetylated forms. For SPC22/23, the identified N-glycans are indicated above the corresponding proteoform peaks. A low-stoichiometric phosphorylated form of SPC12 is annotated in (G) and (K).
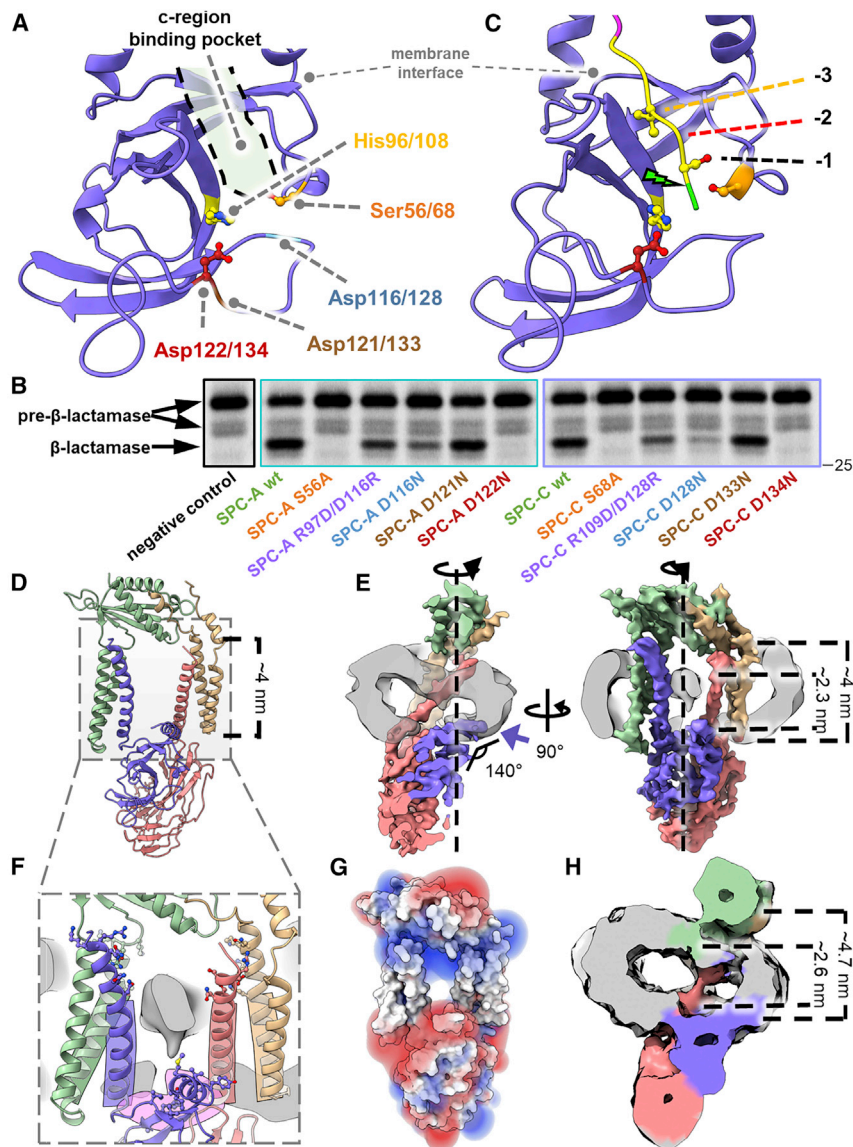
See also Figure S6 and Table S1.

Figure 4. Catalysis and substrate recognition by the SPC

(A) Conserved SPase I fold of SEC11C. The c-region binding pocket and (candidate) catalytic residues are highlighted.

(B) *In vitro* cleavage assay showing the processing of radiolabeled pre-β-lactamase in digitonin by SPC-A and SPC-C mutants. Negative control, no SPC added. The position of the 25-kDa marker is indicated.

(C) C-region of bovine pre-prolactin modeled into the SEC11C binding pocket based on *E. coli* SPase I. The −1 and −3 positions point toward the shallow, hydrophobic pocket. The scissile bond (green arrow) is located close to the catalytic residues.

(D–H) The SPC locally thins the membrane. (D) The TM section of SPC-C is highlighted. (E) Slices through a micelle-containing SPC-C map demonstrate local membrane thinning. Dimensions of the membrane inside and outside the TM window are given. Bowsprit helix is indicated with a purple arrow. SEC11C presses against the membrane from the lumen. Micelle depicted in gray. (F) Polar residues lining the inside of the TM window. The hydrophobic ridge of SEC11C is shown on the luminal side and highlighted in pink; the hydrophobic segments of the TM window helices are boxed for clarity. (G) Electrostatic fields on both sides of the TM window (blue, positive; red, negative). (H) Clipped view of SPC-C in digitonin at 12-Å resolution. The observed membrane thinning effect is akin to that in (E).

See also Figure S7.

to form short, amphiphilic helical segments (Figures 7I–7L). The N-terminal segment of SEC11C is 12 residues longer than that of SEC11A. These residues are conserved among mammals, despite being predicted to be unstructured (Figure S4). Top-down MS reveals significant sequence processing for SEC11C in this area with removal of up to nine N-terminal residues, whereas SEC11A appears as a single unprocessed and unmodified proteoform (Figure 3).

At their C-termini, both SEC11 paralogs are predicted to possess a helix that connects with the bowsprit helix through a conserved proline residue. This "helix-breaker" residue splits the two segments at the interface between the lumen and the ER membrane (Figures 7K and 7L). Deleting the C-terminal segment destabilizes the complex and almost completely abolishes catalytic activity *in vitro* (Figures 7M and 7N). Interestingly, the primary sequence of this C-terminal helix has a hydrophobic stretch of 13 amino acids and a positively charged

C terminus, resembling an inverted SP. Similar to SPs, the length and hydrophobicity—but not the primary sequence—of this segment are conserved among eukaryotes. We named this segment the "C-terminal short (CTS) helix." Given that GFP fused to the C terminus of SEC11C is found on the cytosolic side of digitonin-solubilized particles (Figure 7O) and cross-links only to the cytosolic portions of the SPC (Figure S5B), we employed atomistic MD simulations to test whether the CTS helix could span a lipid membrane (Figure S4C). The data show that the CTS helix, which is not resolved in the EM map, can indeed be stably accommodated as a TM helix in the thinned membrane environment of the SPC, similar to actual SPs. An alternative scenario, which we cannot rule out based on our data, would be that the helix is only partially buried in the membrane and that the C terminus is facing the lumen.

DISCUSSION

SPC organization and comparison to P-type SPases

We show that the SPC exists in two paralogs, each of which is formed by the four subunits SEC11A or SEC11C, SPC22/23, SPC 25, and SPC12. The reason for the existence of two paralogs remains unclear, and there seems to be some functional overlap
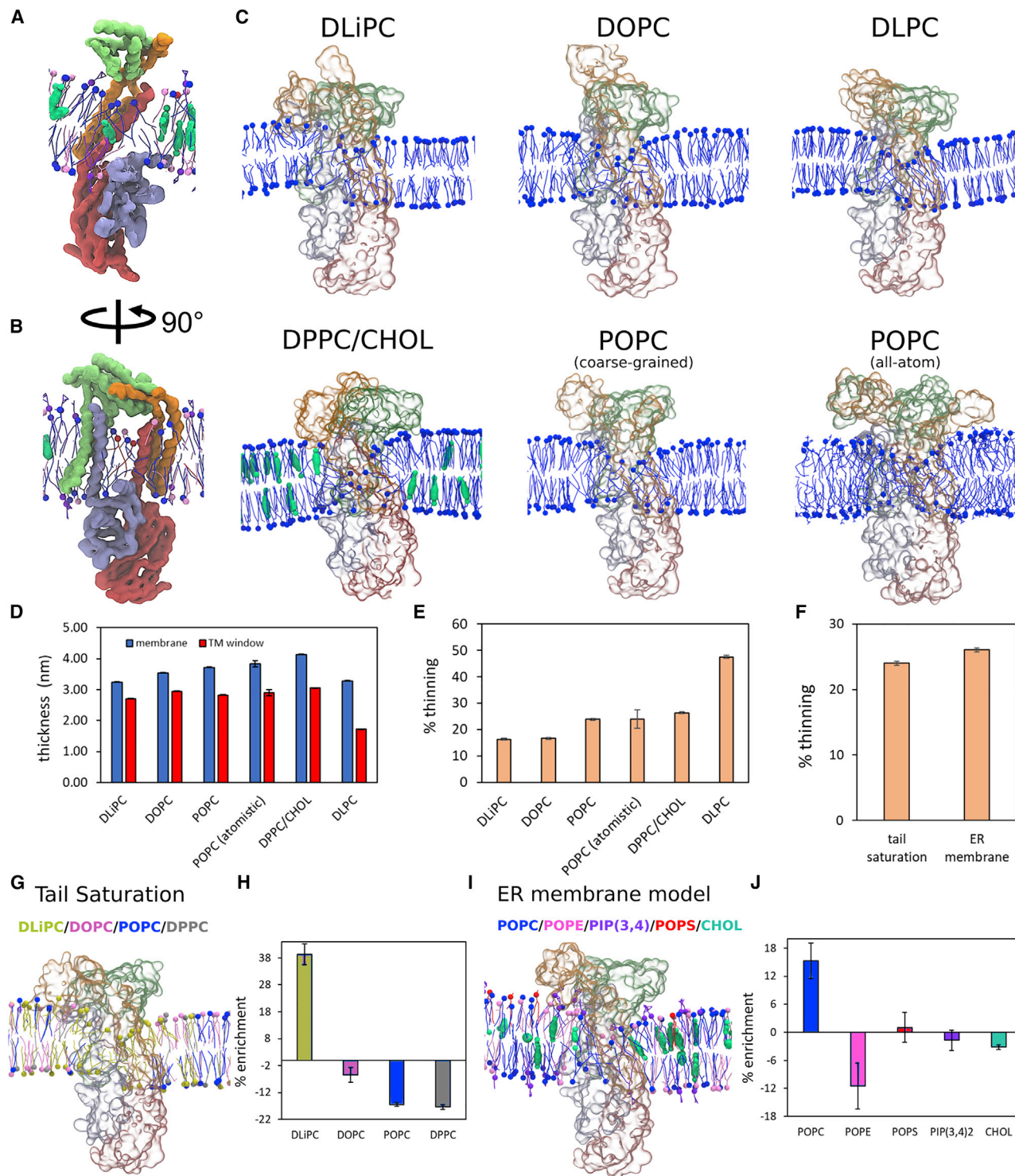
Figure 5. Lipid enrichment and membrane thinning induced by the SPC TM window

(A and B) Coarse-grained MD simulation in a complex ER bilayer showing the lipid distribution within the TM window (blue, POPC; red, POPS; pink, POPE; purple, PI(3,4)P2; green, cholesterol).

(C–E) Thinning induced in simple lipid bilayers. (C) Representative snapshots of coarse-grained MD simulations with SPC (transparent surface, colored as in Figure 4) embedded in DLiPC, DOPC, DLPC, POPC, and DPPC/CHOL. For POPC, the results obtained with atomistic simulations are also shown. Coordinates were smoothed by averaging the coordinates of four neighbor frames of the trajectory. Only a slice of the bilayer around the TM window of SPC is shown. Blue,
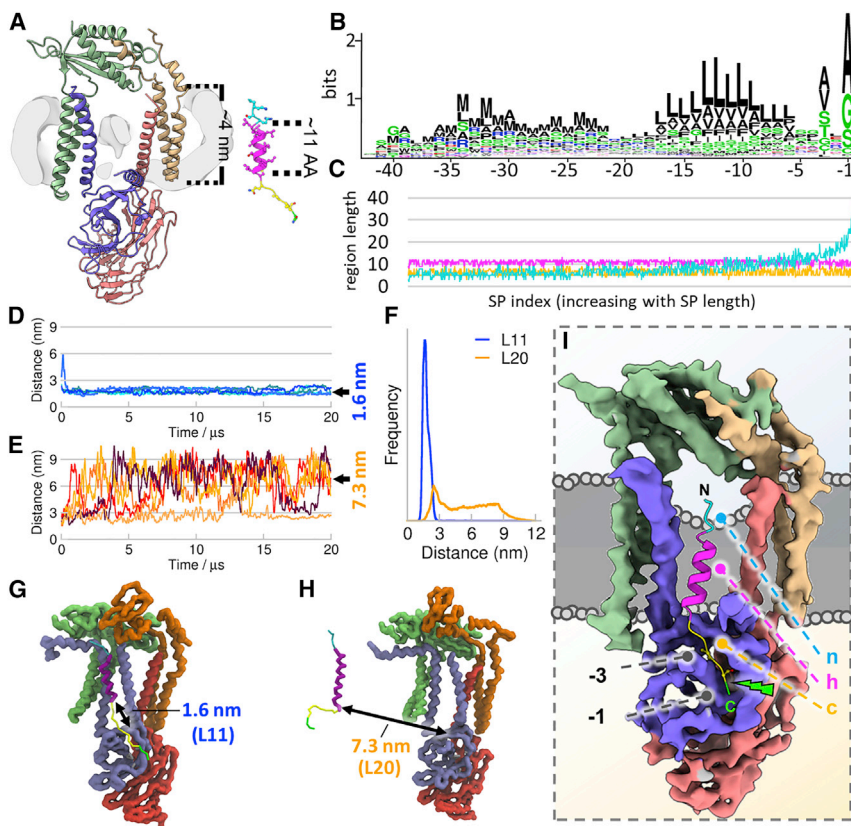
(legend continued on next page)

Figure 6. Membrane thinning provides selectivity for SP engagement

(A) SP h-regions are short compared with the SPC TM helices. The SP of bovine pre-prolactin is shown (cyan, n-region; magenta, h-region; yellow, c-region).

(B) Weblogo (Crooks et al., 2004) showing the frequency of amino acids in a set of 921 experimentally verified SPs, plotted relative to the cleavage site.

(C) Predicted lengths of n-, h-, and c-regions of the same set and colored as in (A). SPs are ordered with increasing overall length from left to right.

(D and E) Time evolution of the SP-TM window distance during five independent 20-µs CGMD simulations (each in a different color shade) of the SPC-C in complex with L11 (D) or L20 (E). Arrows indicate the distances exemplified in panels (G) and (H) for L11 and L20.

(F) Distribution of the SP-TM window distances for CGMD simulations of SPC-C and L11 (blue) or L20 (orange) from 100 µs of sampling.

(G and H) Exemplary snapshots from the CGMD simulations of L11 (G) and L20 (H). Measured distances of the SP to the TM window are indicated.

(I) Proposed model of SP engagement. The SP of bovine pre-prolactin is modeled into the SPC-C binding pocket. The SP, colored as in (A), is recognized based on h-region length and shape complementarity in the c-region. The scissile bond and start of mature sequence are highlighted in green.

(Figure 1D). The high sequence conservation of the subunits suggests that the architecture of the SPC is conserved in eukaryotes.

We reveal that the SPC-A/C active sites are formed by a catalytic triad of Ser56/68, His 96/108, and Asp 122/134. All three amino acids are located on separate structural motifs, which are held together by interactions of three equally conserved neighboring residues of the catalytic triad (Met57/69, Arg97/109, and Asp116/128). Nonetheless, stabilization of this motif by the chaperone-like luminal domain of SPC22/23 is required to uphold catalytic activity to an extent that ensures cell viability (Fang et al., 1997; Kamitani et al., 1989; Meyer and Hartmann, 1997).

The architecture of the human SEC11A/C c-region binding pocket is strikingly similar to P-type SPases, although the latter function through a catalytic Ser-Lys dyad (Figure S7H; Paetzel et al., 1998). Consequently, the principal recognition pattern for SP c-regions is conserved between eukaryotes and prokaryotes. SPs are, in some cases, interchangeable between P- and ER-type SPases, e.g., in the case of *E. coli* β-lactamase (Figure 1D). Both P- and ER-type SPases are resistant to standard serine protease inhibitors (Jackson and Blobel, 1980), which implies that the binding

site and/or catalytic mechanism differ(s) fundamentally from that of other serine proteases. The architectural similarities to the c-region binding pocket of *E. coli* SPase I make it tempting to speculate that the SPC, much like its bacterial homolog, mounts its nucleophilic attack from on the *si*-face of the substrate, as opposed to the *re*-side attack common to other serine proteases (Paetzel et al., 1998).

The most surprising feature of the SPC is the TM window, which is collectively formed by all subunits. The TM helices of the essential subunits SEC11A/C and SPC22/23 form the inner lining of this window, which is framed on either side by TM helix pairs of the non-essential subunits SPC25 and SPC12, respectively. The TM window locally thins the membrane above the SEC11 c-region binding pocket. In contrast, P-type SPases are monomeric and, thus, cannot possess a similarly elaborate TM window. Despite the modest sequence identity, the similarities in tertiary structure of P-type SPases and eukaryotic SEC11 are striking (Figures S7F–S7H). It is, therefore, conceivable that, e.g., *E. coli* SPase I forms a rudimentary version of the TM window through its short, second TM helix, whereas the non-essential SPC12/SPC25 "clamp" is missing.

phospholipids; cyan, cholesterol. (D) Bilayer thickness of the bulk membrane (without SPC, blue) and in the TM window (red) obtained in the MD simulations. (E) Percentage of thinning induced by SPC in different membrane environments.

(F–J) Thinning in complex membranes. (F) Percentage of thinning induced by SPC in the complex lipid mixtures tested here. (G) Representative snapshots of coarse-grained MD simulations of SPC embedded in a mixture of PC lipids with different level of tail saturations. (H) Percentage of PC lipid enrichment in the TM window region in relation to the level of tail saturations. Negative values indicate depletion of lipids. (I and J) Analogous to (G) and (H), but showing different lipid head groups in an endoplasmic reticulum (ER) membrane model.

Errors calculated using Equations 1 and 2; see STAR Methods. For abbreviations and more details about lipid ratios, see Table S2.
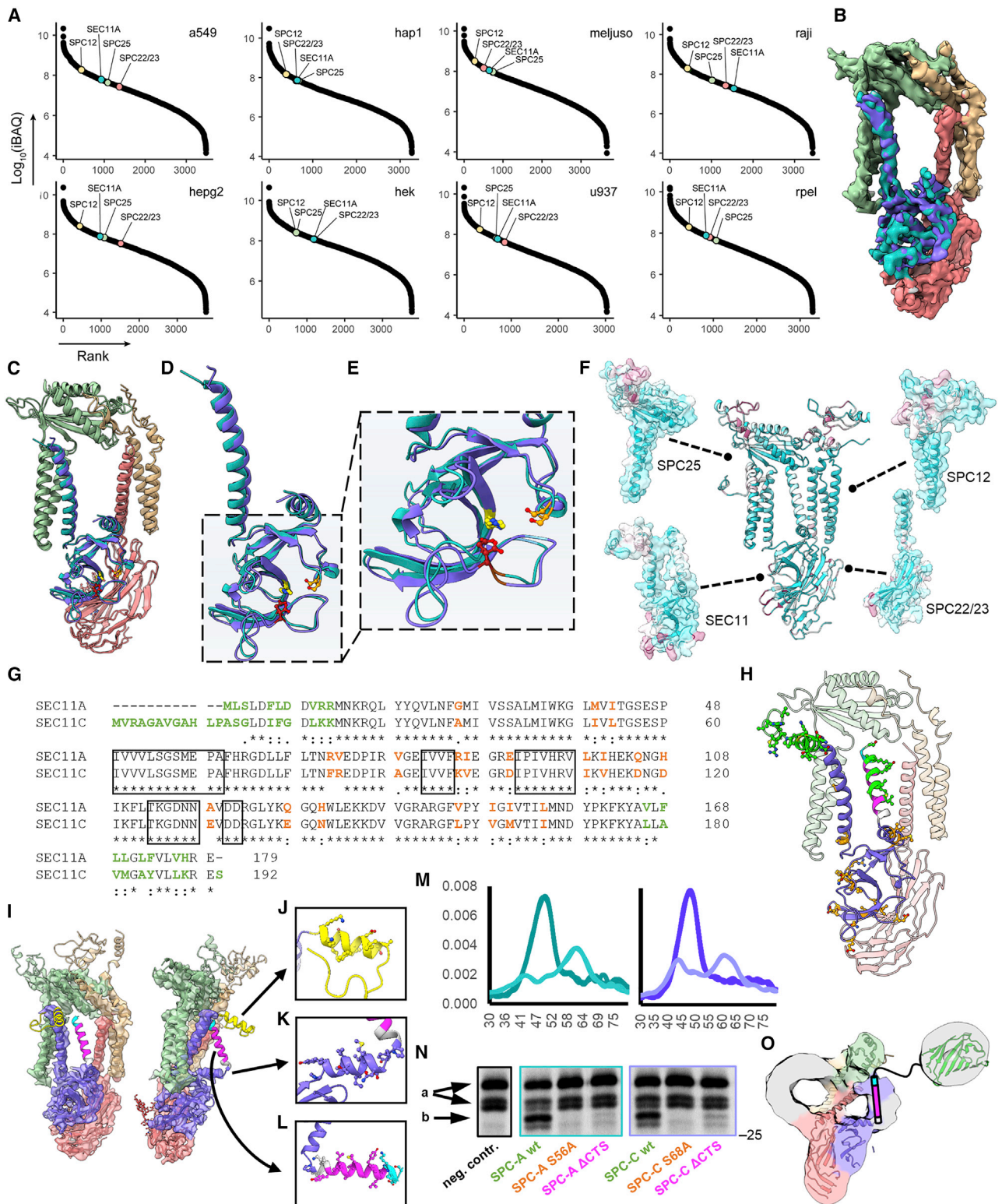
Figure 7. Differences between SEC11A and SEC11C

(A) Intensity-based absolute quantification (iBAQ) values of SPC proteins for eight different cell lines are plotted by rank. SPC12, SPC25, SPC22/23, and SEC11A are marked; SEC11C was not detected in any of the tested cell lines.

(B and C) Superposition of density maps (B) and atomic models (C) for SPC-A and SPC-C (colored according to Figure 1).

(legend continued on next page)

Signal peptides possess short h-regions

The primary sequence and length of SPs can vary substantially, from little more than 10 to more than 100 amino acids. Nonetheless, within the limits of helix prediction accuracy, our bioinformatic analysis of SPs paints a clear picture of the length of the h-region when compared with similar automated analyses of TM helices (Sharpe et al., 2010). The mean length of SP h-regions (11 residues) is shorter than that of TM helices (>20 residues). Even the TM helices in ER and Golgi, which are generally shorter than those of plasma membrane proteins, are predicted to be substantially longer than SP h-regions on average. Only when we manually adjusted for maximum h-region length did we detected a few SPs with h-regions of potentially 18 or more residues, such as human Fc receptor-like A and CD79a, which will have to be studied on a case-by-case basis. Our set also includes some exotic cases, such as the long SPs of the *Drosophila melanogaster* Crumbs (Kilic et al., 2010), UL40 from human cytomegalovirus (Prod'homme et al., 2012), and fusion glycoprotein F0 from canine distemper virus (von Messling and Cattaneo, 2002). In addition, for these extreme examples, the predicted h-regions are not unusually long, and the length variation originates from the n-region. Although the SPC structure does not suggest direct interaction with n-regions, long, folded n-regions might cause steric obstructions of the SP and delay cleavage, as observed, e.g., for flaviviral prME (Alzahrani et al., 2020). In about 19% of the data, the predicted and experimental cleavage sites differed, possibly because of incorrect annotations, e.g., due to *in vivo* post-processing by other peptidases (Bendtsen et al., 2004), which in some cases resulted in long c-regions of more than 30 residues. These cases will have to be verified experimentally.

The SPC uses the membrane as a "molecular ruler" for SP h-regions

The SPCs remarkable specificity for a large and sequentially diverse set of substrates—approximated to exceed 3,000 SPs in humans (Uhlén et al., 2015)—has been a long-standing enigma in cell biology. The SPC-A/C structures suggest that selectivity is achieved through the combination of two key binding determinants (Figure 6I). The structures of bacterial P-type SPases (Paetzel, 2014) have already shown that the shallow, hydrophobic c-region-binding pocket dictates that substrates follow the (−3, −1) rule. The structure of the SPC reveals that this principle is conserved in eukaryotes.

Our data suggest that membrane shaping by the SPC is a second determinant of SP specificity. Systematic assays demonstrate that the eukaryotic SPC cannot cleave SPs with h-regions longer than 18–20 amino acids (Nilsson et al., 1994, 2002). Here, we rationalize this finding using the SPC structure and MD simulations. The data show that the SPC allows SPs with short h-regions to diffuse into the TM window and to access the c-region-binding pocket, whereas regular TM helices of 20 or more amino acids remain excluded (Figure 6). Along these lines, SPs with short h-regions have been experimentally shown to possess affinity for membrane regions with reduced bilayer thickness, whereas longer TM segments do not (Tahara et al., 1992). The low sequence conservation of SP h-regions suggests that they interact, rather, with the SPC's lipid environment than with its TM helices. In that context, the enrichment of phosphatidylcholine within the TM window indicated by our MD simulations also explains why relipidation of the SPC with phosphatidylcholine is required to restore the catalytic activity of the SPC in some detergent systems (Evans et al., 1986; Lively and Walsh, 1983; Uchida et al., 1986). Based on the available data, we conclude that the membrane thinning in the SPC TM window is likely a key determinant for SPC specificity (Figure 6; Videos S2 and S4).

In combination, the selectivity of the TM window for short h-regions and of the active site for shallow, hydrophobic residues at positions −1 and −3 relative to the cleavage site ensure specificity and allow a high degree of flexibility that can retain the diverse up- or downstream functions of SPs (Hegde and Bernstein, 2006; Kapp et al., 2009). For example, in the supposedly rare cases of TM helices that are sufficiently short to diffuse into the TM window (e.g., some Golgi-resident TM proteins [Sharpe et al., 2010]), the lack of shape complementarity in the c-region may provide protection from cleavage. Based on the conservation of the SPC subunits and the overall similarity of SPs, it appears conceivable that this mechanism is conserved among all eukaryotes.

Membrane shaping has recently emerged as a common theme among insertases, such as YidC (Kumazaki et al., 2014), the TMCO1 translocon (McGilvray et al., 2020), the guided entry of tail-anchored proteins (GET) insertase complex (McDowell et al., 2020), the ER membrane complex (Pleiner et al., 2020), and translocases, such as the Hrd1 - ER-associated protein degradation (ERAD) complex (Schoebel et al., 2017; Wu et al., 2020). Although insertases and translocases use membrane thinning to decrease the energetic cost of protein insertion and

---

(D and E) Superposition of atomic models for SEC11A (teal) and SEC11C (purple).

(F) Conservation mapped onto the SPC structure (teal, conserved; red, variable). Contact points between subunits as well as the SP binding groove are highly conserved.

(G) Pairwise alignment of SEC11A and SEC11C. SP binding site residues highlighted in black boxes. Variable residues resolved in the maps are highlighted in orange, variable residues in flexible regions are highlighted in green.

(H) differences between SEC11A and SEC11C mapped onto the SPC-C structure, colored as in (G).

(I–L) Unresolved regions of SEC11A/C are shown as predicted by trRosetta. (J) N-terminal amphipathic helix (yellow); (K) bowsprit helix; (L) CTS helix (gray/magenta/cyan).

(M) Nano-differential scanning fluorimetry showing the effect of CTS helix deletion on the stability of SPC-A (teal) and SPC C (purple). ΔCTS constructs colored in lighter shade. The first derivative of the fluorescence ratio (350/330 nm; y axis) is plotted against the temperature in °C (x axis). Melting profiles were determined as duplicates.

(N) *In vitro* cleavage assay showing the processing of radiolabeled pre-β-lactamase in digitonin by SPC-A and SPC-C ΔCTS. a, non-cleaved; b, cleaved; negative control, no SPC added. The position of the 25-kDa marker is indicated.

(O) GFP (green cartoon) fused to the C terminus of SEC11C is located at the cytosolic face of the particle, indicating that the C-terminal CTS helix is a TM segment.

translocation, the SPC uses the lipid compartment in the TM window as a molecular ruler to measure the length of its substrates, which is distinct from that of canonical TMs.

Conclusions

In summary, we reveal the existence of two SPC paralogs, SPC-A and SPC-C and report their structures and post-translational modifications. The structures suggest that the determinants for SP binding are (1) a c-region binding pocket similar to P-type SPases, and (2) a TM window formed by all subunits, which locally thins the membrane and acts as a molecular ruler that separates SPs from TM segments based on helix length. This study will allow for more accurate predictions of SP sequences in eukaryotic organisms. The molecular details of SP recognition and cleavage by the SPC will benefit the development of potent SPs for protein mass production of, e.g., antibodies, hormones, or protein-based vaccines from human, yeast, and plant sources.

Limitations of the study

Our work reveals the architecture of the SPC. Because of the small size (67 kDa of ordered density), absence of symmetry, the presence of a large micelle that occupies about 50% of the particle volume, and the EM equipment used here, the resolution of our reconstruction is limited to about 4.9 Å. At that resolution, it is not possible to reliably model amino acid side chains. It will be interesting to improve the resolution of the EM map and atomic model using cutting-edge equipment (Nakane et al., 2020; Yip et al., 2020). A cryo-EM reconstruction of the SPC in complex with a SP will further validate the model for SP engagement, the catalytic mechanism, and the role of membrane thinning put forward in this work.

Ultimately, it will be exciting to obtain mechanistic insights into the transfer of SPs from Sec61 to SPC and the different roles of the SPC paralogs for a mechanistic understanding of the SPC in the cell. To that end, further experiments addressing the nature of the CTS helix are warranted. Possible functions of the CTS helix might include increased robustness of membrane thinning in absence of the accessory subunits, direct interaction with the h-region of SPs, a steric "guide" for the diffusion of SPs into the SPC binding pocket, or possibly an interaction with the ER translocon, e.g., via the lateral gate of Sec61.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Design and cloning of expression constructs
  - Protein expression and purification
  - Activity assay
  - Differential scanning fluorimetry

  - Shotgun mass spectrometry
  - Native mass spectrometry
  - Intact mass and top-down mass spectrometry
  - Cross-linking mass spectrometry
  - Single particle analysis
  - Surface potential calculation
  - Secondary structure and disorder prediction
  - Phylogenic analysis and residue conservation
  - Molecular dynamics simulations
  - Empirical SP analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

AUTHOR CONTRIBUTIONS

A.M.L. and F.F. conceived the project. A.M.L. designed the expression constructs. A.M.L. and M.G.-M. cloned constructs and point mutants. A.M.L. expressed cells and established a protein purification protocol. A.M.L. and P.O. performed protein purification and sample purification for cryo-EM, and MS. B.S., S.T., R.A.S., A.M.L., and F.F. conceived the MS experiments. B.S., R.A.S., and S.T. performed the MS experiments and data analysis. A.M.L. and P.O. solved the structures of SPC-A and SPC-C, including sample preparation, data collection, data processing, and refinement. A.M.L. and F.F. generated, refined, and analyzed the atomic models. A.M.L. conceived and performed the stability and activity assays. P.C.T.d.S., A.M.L., S.J.M., and F.F. conceived the MD simulation experiments, which P.C.T.d.S. performed and analyzed. M.G.-M. grew cells for the shotgun analysis. A.M.L., B.S., S.T., P.C.T.d.S., S.J.M., R.A.S., and F.F. wrote the manuscript and prepared the figures with help from all co-authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and Lindah, E. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX *1–2*, 19–25.

Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A., and Adams, P.D. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. Acta Crystallogr. D Struct. Biol. *74*, 531–544.

Albanese, P., Tamara, S., Saracco, G., Scheltema, R.A., and Pagliano, C. (2020). How paired PSII-LHCII supercomplexes mediate the stacking of plant thylakoid membranes unveiled by structural mass-spectrometry. Nat. Commun. *11*, 1361.

Alzahrani, N., Wu, M.J., Shanmugam, S., and Yi, M. (2020). Delayed by design: Role of suboptimal signal peptidase processing of viral structural protein precursors in flaviviridae virus assembly. Viruses *12*, 1090.

Antonin, W., Meyer, H.A., and Hartmann, E. (2000). Interactions between Spc2p and other components of the endoplasmic reticulum translocation sites of the yeast Saccharomyces cerevisiae. J. Biol. Chem. *275*, 34068–34072.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. *44* (W1), W344–W350.

Bastian, F.B., Parmentier, G., Roux, J., Moretti, S., Laudet, V., and Robinson-Rechavi, M. (2008). Bgee: integrating and comparing heterogeneous transcriptome data among species. In Proceedings of Data Integration in Life Sciences: Lecture Notes in Computer Science, Fifth International Workshop, DILS 2008, S. Istrail, P. Pevzner, and M. Waterman, eds. (Springer), pp. 124–131.

Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. *340*, 783–795.

Blobel, G., and Dobberstein, B. (1975). Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. J. Cell Biol. *67*, 835–851.

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. J. Chem. Physiol. *126*, 014101.

Carpenter, T.S., López, C.A., Neale, C., Montour, C., Ingólfsson, H.I., Di Natale, F., Lightstone, F.C., and Gnanakaran, S. (2018). Capturing phase behavior of ternary lipid mixtures with a refined martini coarse-grained force field. J. Chem. Theory Comput. *14*, 6050–6062.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: A Sequence Logo Generator. Genome Research *14*, 1188–1190.

Daganzo, S.M., Erzberger, J.P., Lam, W.M., Skordalakes, E., Zhang, R., Franco, A.A., Brill, S.J., Adams, P.D., Berger, J.M., and Kaufman, P.D. (2003). Structure and function of the conserved core of histone deposition protein Asf1. Curr. Biol. *13*, 2148–2158.

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. J. Chem. Phys. *98*, 10089–10092.

de Jong, D.H., Singh, G., Bennett, W.F.D., Arnarez, C., Wassenaar, T.A., Schäfer, L.V., Periole, X., Tieleman, D.P., and Marrink, S.J. (2013). Improved parameters for the martini coarse-grained protein force field. J. Chem. Theory Comput. *9*, 687–697.

de Jong, D.H., Baoukina, S., Ingólfsson, H.I., and Marrink, S.J. (2016). Martini straight: boosting performance using a shorter cutoff and GPUs. Comput. Phys. Commun. *199*, 1–7.

Drozdetskiy, A., Cole, C., Procter, J., and Barton, G.J. (2015). JPred4: a protein secondary structure prediction server. Nucleic Acids Res. *43* (W1), W389–W394.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Essletzbichler, P., Konopka, T., Santoro, F., Chen, D., Gapp, B.V., Kralovics, R., Brummelkamp, T.R., Nijman, S.M.B., and Bürckstümmer, T. (2014). Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. Genome Res. *24*, 2059–2065.

Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. (1995). A smooth particle mesh Ewald method. J. Chem. Phys. *103*, 8577–8593.

Estoppey, D., Lee, C.M., Janoschke, M., Lee, B.H., Wan, K.F., Dong, H., Mathys, P., Filipuzzi, I., Schuhmann, T., Riedl, R., et al. (2017). The natural product cavinafungin selectively interferes with zika and dengue virus replication by inhibition of the host signal peptidase. Cell Rep. *19*, 451–460.

Evans, E.A., Gilmore, R., and Blobel, G. (1986). Purification of microsomal signal peptidase as a complex. Proc. Natl. Acad. Sci. USA *83*, 581–585.

Fang, H., Panzner, S., Mullins, C., Hartmann, E., and Green, N. (1996). The homologue of mammalian SPC12 is important for efficient signal peptidase activity in *Saccharomyces cerevisiae*. J. Biol. Chem. *271*, 16460–16465.

Fang, H., Mullins, C., and Green, N. (1997). In addition to SEC11, a newly identified gene, SPC3, is essential for signal peptidase activity in the yeast endoplasmic reticulum. J. Biol. Chem. *272*, 13152–13158.

Gemmer, M., and Förster, F. (2020). A clearer picture of the ER translocon complex. J. Cell Sci. *133*, jcs231340.

Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: meeting modern challenges in visualization and analysis. Protein Sci. *27*, 14–25.

Hegde, R.S., and Bernstein, H.D. (2006). The surprising complexity of signal sequences. Trends Biochem. Sci. *31*, 563–571.

Herzik, M.A., Jr., Wu, M., and Lander, G.C. (2019). High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM. Nat. Commun. *10*, 1032.

Herzog, F.A., Braun, L., Schoen, I., and Vogel, V. (2016). Improved side chain dynamics in MARTINI simulations of protein-lipid interfaces. J. Chem. Theory Comput. *12*, 2446–2458.

Hess, B., Bekker, H., Berendsen, H.J.C., and Fraaije, J.G.E.M. (1997). LINCS: a linear constraint solver for molecular simulations. J. Comput. Chem. *18*, 1463–1472.

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B.L., Grubmüller, H., and MacKerell, A.D., Jr. (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat. Methods *14*, 71–73.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J. Mol. Graph. *14*, 33–38, 27–28.

Jackson, R.C. (1983). Quantitative assay for signal peptidase. Methods Enzymol. *96*, 784–794.

Jackson, R.C., and Blobel, G. (1980). Post-translational processing of full-length presecretory proteins with canine pancreatic signal peptidase. Ann. N Y Acad. Sci. *343*, 391–404.

Jorgensen, W.L. (1981). Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. J. Am. Chem. Soc. *103*, 335–340.

Jurchen, J.C., and Williams, E.R. (2003). Origin of asymmetric charge partitioning in the dissociation of gas-phase protein homodimers. J. Am. Chem. Soc. *125*, 2817–2826.

Kalies, K.U., and Hartmann, E. (1996). Membrane topology of the 12- and the 25-kDa subunits of the mammalian signal peptidase complex. J. Biol. Chem. *271*, 3925–3929.

Kamitani, K., Narita, K., and Sakiyama, F. (1989). Purification and characterization of hen oviduct N α-acetyltransferase. J. Biol. Chem. *264*, 13188–13193.

Kapp, K., Schrempf, S., Lemberg, M.K., and Dobberstein, B. (2009). Post-targeting functions of signal peptides. In Madame Curie Bioscience Database: Protein Transport into the Endoplasmic Reticulum (Landes Bioscience), pp. 1–16.

Kelstrup, C.D., Bekker-Jensen, D.B., Arrey, T.N., Hogrebe, A., Harder, A., and Olsen, J.V. (2018). Performance evaluation of the Q exactive HF-X for shotgun proteomics. J. Proteome Res. *17*, 727–738.

Kilic, A., Klose, S., Dobberstein, B., Knust, E., and Kapp, K. (2010). The *Drosophila* Crumbs signal peptide is unusually long and is a substrate for signal peptide peptidase. Eur. J. Cell Biol. *89*, 449–461.

Klauda, J.B., Venable, R.M., Freites, J.A., O'Connor, J.W., Tobias, D.J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell, A.D., Jr., and Pastor, R.W. (2010). Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. J. Phys. Chem. B *114*, 7830–7843.

Klausen, M.S., Jespersen, M.C., Nielsen, H., Jensen, K.K., Jurtz, V.I., Sønderby, C.K., Sommer, M.O.A., Winther, O., Nielsen, M., Petersen, B., and Marcatili, P. (2019). NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. Proteins *87*, 520–527.

Klykov, O., Steigenberger, B., Pektaş, S., Fasci, D., Heck, A.J.R., and Scheltema, R.A. (2018). Efficient and robust proteome-wide approaches for cross-linking mass spectrometry. Nat. Protoc. *13*, 2964–2990.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol. Biol. Evol. *35*, 1547–1549.

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

Lee, J., Cheng, X., Swails, J.M., Yeom, M.S., Eastman, P.K., Lemkul, J.A., Wei, S., Buckner, J., Jeong, J.C., Qi, Y., et al. (2016). CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. J. Chem. Theory Comput. *12*, 405–413.

Liaci, A.M., and Förster, F. (2021). Empirical analysis of eukaryotic ER signal peptides (Mendeley Data). version 1. https://doi.org/10.17632/P65tkrr89v.1.

Liang, H., VanValkenburgh, C., Chen, X., Mullins, C., Van Kaer, L., Green, N., and Fang, H. (2003). Genetic complementation in yeast reveals functional similarities between the catalytic subunits of mammalian signal peptidase complex. J. Biol. Chem. *278*, 50932–50939.

Lively, M.O., and Walsh, K.A. (1983). Hen oviduct signal peptidase is an integral membrane protein. J. Biol. Chem. *258*, 9488–9495.

MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B *102*, 3586–3616.

Marrink Laboratory (2020). Martinize and Vermouth: The Ultimate Resolution Transformation Tools (GitHub). https://github.com/marrink-lab/vermouth-martinize.

Marrink, S.J., Risselada, H.J., Yefimov, S., Tieleman, D.P., and de Vries, A.H. (2007). The MARTINI force field: coarse grained model for biomolecular simulations. J. Phys. Chem. B *111*, 7812–7824.

Martínez, L. (2015). Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. PLoS ONE *10*, e0119264.

Marty, M.T., Baldwin, A.J., Marklund, E.G., Hochberg, G.K.A., Benesch, J.L.P., and Robinson, C.V. (2015). Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. Anal. Chem. *87*, 4370–4376.

McDowell, M.A., Heimes, M., Fiorentino, F., Mehmood, S., Farkas, Á., Coy-Vergara, J., Wu, D., Bolla, J.R., Schmid, V., Heinze, R., et al. (2020). Structural basis of tail-anchored membrane protein biogenesis by the GET insertase complex. Mol. Cell *80*, 72–86.e7.

McGilvray, P.T., Anghel, S.A., Sundaram, A., Zhong, F., Trnka, M.J., Fuller, J.R., Hu, H., Burlingame, A.L., and Keenan, R.J. (2020). An ER translocon for multi-pass membrane protein biogenesis. eLife *9*, 1–43.

Meyer, H.A., and Hartmann, E. (1997). The yeast SPC22/23 homolog Spc3p is essential for signal peptidase activity. J. Biol. Chem. *272*, 13159–13164.

Mitra, K., Ubarretxena-Belandia, I., Taguchi, T., Warren, G., and Engelman, D.M. (2004). Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol. PNAS *101* (12), 4083–4088.

Mullins, C., Meyer, H.A., Hartmann, E., Green, N., and Fang, H. (1996). Structurally related Spc1p and Spc2p of yeast signal peptidase complex are functionally distinct. J. Biol. Chem. *271*, 29094–29099.

Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P.M.G.E., Grigoras, I.T., Malinauskaite, L., Malinauskas, T., Miehling, J., et al. (2020). Single-particle cryo-EM at atomic resolution. Nature *587*, 152–156.

Nielsen, H., Tsirigos, K.D., Brunak, S., and von Heijne, G. (2019). A brief history of protein sorting prediction. Protein J. *38*, 200–216.

Nilsson, I., Whitley, P., and von Heijne, G. (1994). The COOH-terminal ends of internal signal and signal-anchor sequences are positioned differently in the ER translocase. J. Cell Biol. *126*, 1127–1132.

Nilsson, I., Johnson, A.E., and von Heijne, G. (2002). Cleavage of a tail-anchored protein by signal peptidase. FEBS Lett. *516*, 106–108.

Oostra, M., te Lintelo, E.G., Deijs, M., Verheije, M.H., Rottier, P.J.M., and de Haan, C.A.M. (2007). Localization and membrane topology of coronavirus nonstructural protein 4: involvement of the early secretory pathway in replication. J. Virol. *81*, 12323–12336.

Paetzel, M. (2014). Structure and mechanism of Escherichia coli type I signal peptidase. Biochim. Biophys. Acta *1843*, 1497–1508.

Paetzel, M., Dalbey, R.E., and Strynadka, N.C.J. (1998). Crystal structure of a bacterial signal peptidase in complex with a β-lactam inhibitor. Nature *396*, 186–190.

Paetzel, M., Karla, A., Strynadka, N.C.J., and Dalbey, R.E. (2002). Signal peptidases. Chem. Rev. *102*, 4549–4580.

Paetzel, M., Goodall, J.J., Kania, M., Dalbey, R.E., and Page, M.G.P. (2004). Crystallographic and biophysical analysis of a bacterial signal peptidase in complex with a lipopeptide-based inhibitor. J. Biol. Chem. *279*, 30781–30790.

Palade, G., Arch, A., Locke, M., Locke, E., and Palade, G. (1975). Intracellular aspects of the process of protein synthesis. Science *189*, 347–358.

Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. J. Appl. Physiol. *52*, 7182–7190.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Pleiner, T., Tomaleri, G.P., Januszyk, K., Inglis, A.J., Hazu, M., and Voorhees, R.M. (2020). Structural basis for membrane insertion by the human ER membrane protein complex. Science *369*, 433–436.

Popa, V., Trecroce, D.A., McAllister, R.G., and Konermann, L. (2016). Collision-induced dissociation of electrosprayed protein complexes: An all-atom molecular dynamics model with mobile protons. J. Phys. Chem. B *120*, 5114–5124.

Prod'homme, V., Tomasec, P., Cunningham, C., Lemberg, M.K., Stanton, R.J., McSharry, B.P., Wang, E.C.Y., Cuff, S., Martoglio, B., Davison, A.J., et al. (2012). Human cytomegalovirus UL40 signal peptide regulates cell surface expression of the NK cell ligands HLA-E and gpUL18. J. Immunol. *188*, 2794–2804.

Rohou, A., and Grigorieff, N. (2015). CTFFIND4: fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. *192*, 216–221.

Rosenthal, P.B., and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J. Mol. Biol. *333*, 721–745.

Scheltema, R.A., Hauschild, J.P., Lange, O., Hornburg, D., Denisov, E., Damoc, E., Kuehn, A., Makarov, A., and Mann, M. (2014). The Q Exactive HF, a benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field orbitrap analyzer. Mol. Cell. Proteomics *13*, 3698–3708.

Scheres, S.H.W. (2012). A Bayesian view on cryo-EM structure determination. J. Mol. Biol. *415*, 406–418.

Schoebel, S., Mi, W., Stein, A., Ovchinnikov, S., Pavlovicz, R., DiMaio, F., Baker, D., Chambers, M.G., Su, H., Li, D., et al. (2017). Cryo-EM structure of the protein-conducting ERAD channel Hrd1 in complex with Hrd3. Nature *548*, 352–355.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Senko, M.W., Remes, P.M., Canterbury, J.D., Mathur, R., Song, Q., Eliuk, S.M., Mullen, C., Earley, L., Hardman, M., Blethrow, J.D., et al. (2013). Novel parallelized quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. Anal. Chem. *85*, 11710–11714.

Sharpe, H.J., Stevens, T.J., and Munro, S. (2010). A comprehensive comparison of transmembrane domains reveals organelle-specific properties. Cell *142*, 158–169.

Shelness, G.S., and Blobel, G. (1990). Two subunits of the canine signal peptidase complex are homologous to yeast SEC11 protein. J. Biol. Chem. *265*, 9512–9519.

Shelness, G.S., Lin, L., and Nicchitta, C.V. (1993). Membrane topology and biogenesis of eukaryotic signal peptidase. J. Biol. Chem. *268*, 5201–5208.

Snapp, E.L., McCaul, N., Quandte, M., Cabartova, Z., Bontjer, I., Källgren, C., Nilsson, I., Land, A., von Heijne, G., Sanders, R.W., and Braakman, I. (2017). Structure and topology around the cleavage site regulate post-translational cleavage of the HIV-1 gp160 signal peptide. eLife *6*, 1–25.

Souza, P.C.T., Alessandri, R., Barnoud, J., Thallmair, S., Faustino, I., Grünewald, F., Patmanidis, I., Abdizadeh, H., Bruininks, B.M.H., Wassenaar, T.A., et al. (2021). Martini 3: a general purpose force field for coarse-grained molecular dynamics. Nat. Methods *18*, 382–388.

Steigenberger, B., Pieters, R.J., Heck, A.J.R., and Scheltema, R.A. (2019). PhoX: An IMAC-enrichable cross-linking reagent. ACS Cent. Sci. *5*, 1514–1522.

Suzuki, R., Matsuda, M., Watashi, K., Aizaki, H., Matsuura, Y., Wakita, T., and Suzuki, T. (2013). Signal peptidase complex subunit 1 participates in the assembly of hepatitis C virus through an interaction with E2 and NS2. PLoS Pathog. *9*, e1003589.

Tahara, Y., Murata, M., Ohnishi, S., Fujiyoshi, Y., Kikuchi, M., and Yamamoto, Y. (1992). Functional signal peptide reduces bilayer thickness of phosphatidylcholine liposomes. Biochemistry *31*, 8747–8754.

Tegunov, D., and Cramer, P. (2019). Real-time cryo-electron microscopy data preprocessing with Warp. Nat. Methods *16*, 1146–1152.

Ting, Y.T., Harris, P.W.R., Batot, G., Brimble, M.A., Baker, E.N., and Young, P.G. (2016). Peptide binding to a bacterial signal peptidase visualized by peptide tethering and carrier-driven crystallization. IUCrJ *3*, 10–19.

Tschantz, W.R., Sung, M., Delgado-Partin, V.M., and Dalbey, R.E. (1993). A serine and a lysine residue implicated in the catalytic mechanism of the Escherichia coli leader peptidase. J. Biol. Chem. *268*, 27349–27354.

Tyndall, J.D.A., Nall, T., and Fairlie, D.P. (2005). Proteases universally recognize beta strands in their active sites. Chem. Rev. *105*, 973–999.

Uchida, M., Watanabe, Y., and Fujimoto, Y. (1986). Is phospholipid a required cofactor for the activity of mammalian signal peptidase? FEBS Lett. *200*, 343–346.

Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics: tissue-based map of the human proteome. Science *347*, 1260419.

UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. *49* (D1), D480–D489.

Van Damme, P., Lasa, M., Polevoda, B., Gazquez, C., Elosegui-Artola, A., Kim, D.S., De Juan-Pardo, E., Demeyer, K., Hole, K., Larrea, E., et al. (2012). N-terminal acetylome analyses and functional insights of the N-terminal acetyltransferase NatB. Proc. Natl. Acad. Sci. USA *109*, 12449–12454.

van Dijl, J.M., de Jong, A., Vehmaanperä, J., Venema, G., and Bron, S. (1992). Signal peptidase I of Bacillus subtilis: patterns of conserved amino acids in prokaryotic and eukaryotic type I signal peptidases. EMBO J. *11*, 2819–2828.

van Hilten, N., Stroh, K.S., and Risselada, H.J. (2020). Membrane thinning induces sorting of lipids and the amphipathic lipid packing sensor (ALPS) protein motif. Front. Physiol. *11*, 250.

VanValkenburgh, C., Chen, X., Mullins, C., Fang, H., and Green, N. (1999). The catalytic mechanism of endoplasmic reticulum signal peptidase appears to be distinct from most eubacterial signal peptidases. J. Biol. Chem. *274*, 11519–11525.

von Heijne, G. (1985). Signal sequences. The limits of variation. J. Mol. Biol. *184*, 99–105.

von Heijne, G. (1990). The signal peptide. J. Membr. Biol. *115*, 195–201.

von Heijne, G. (2006). Membrane-protein topology. Nat. Rev. Mol. Cell Biol. *7*, 909–918.

von Messling, V., and Cattaneo, R. (2002). Amino-terminal precursor sequence modulates canine distemper virus fusion protein function. J. Virol. *76*, 4172–4180.

Wagner, T. (2020). MPI-Dortmund/sphire-janni: JANNI (Version v0.1.2) (Zendodo). https://doi.org/10.5281/zenodo.3378300.

Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., Raisch, T., Prumbaum, D., et al. (2019). SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. Commun. Biol. *2*, 218.

Walker, S.J., and Lively, M.O. (2013). Signal peptidase (eukaryote). In Handbook of Proteolytic Enzymes, Volume 1, A.J. Barrett, J.F. Woessner, N.D. Rawlings, and G. Salvesen, eds. (Academic Press), pp. 3512–3517.

Wassenaar, T.A., Pluhackova, K., Böckmann, R.A., Marrink, S.J., and Tieleman, D.P. (2014). Going backward: A flexible geometric approach to reverse transformation from coarse grained to atomistic models. J. Chem. Theory Comput. *10*, 676–690.

Wassenaar, T.A., Ingólfsson, H.I., Böckmann, R.A., Tieleman, D.P., and Marrink, S.J. (2015). Computational lipidomics with insane: A versatile tool for generating custom membranes for molecular simulations. J. Chem. Theory Comput. *11*, 2144–2155.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. (Springer). https://doi.org/10.1007/978-0-387-98141-3.

Wu, X., Siggel, M., Ovchinnikov, S., Mi, W., Svetlov, V., Nudler, E., Liao, M., Hummer, G., and Rapoport, T.A. (2020). Structural basis of ER-associated protein degradation mediated by the Hrd1 ubiquitin ligase complex. Science *368*, eaaz2449.

Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., and Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. Proc. Natl. Acad. Sci. USA *117*, 1496–1503.

Yip, K.M., Fischer, N., Paknia, E., Chari, A., and Stark, H. (2020). Atomic-resolution protein structure determination by cryo-EM. Nature *587*, 157–161.

Zamdborg, L., LeDuc, R.D., Glowacz, K.J., Kim, Y.B., Viswanathan, V., Spaulding, I.T., Early, B.P., Bluhm, E.J., Babai, S., and Kelleher, N.L. (2007). ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. Nucleic Acids Res. *35*, W701–W706.

Zhang, R., Miner, J.J., Gorman, M.J., Rausch, K., Ramage, H., White, J.P., Zuiani, A., Zhang, P., Fernandez, E., Zhang, Q., et al. (2016). A CRISPR screen defines a signal peptide processing pathway required by flaviviruses. Nature *535*, 164–168.

Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods *14*, 331–332.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
cT$_2$ rabbit reticulocyte lysate	Green Hectares	N/A
Chemicals, peptides, and recombinant proteins		
Fluorinated fos-choline-8	Anatrace	Cat#F300F
PMAL-C8	Anatrace	Cat#P5008
Digitonin	Calbiochem	Cat#300410
n-dodecyl-β-D-maltoside (DDM)	Sigma-Aldrich	Cat#D4641
Cholesteryl hemisuccinate (CHS)	Sigma-Aldrich	Cat#C6512
cOmplete protease inhibitor cocktail	Roche	Cat#11836145001
TEV protease	This study	N/A
EasyTag express ^{35}S Protein Labeling Mix	PerkinElmer	Cat#NEG772002MC
PhoX	Steigenberger et al., 2019	N/A
D-biotin	VWR Life Science	Cat#0340-5G
Critical commercial assays		
2x Gibson assembly master mix	New England Biolabs	Cat#E2611S
Q5 site-directed mutagenesis kit	New England Biolabs	Cat#E0554S
Streptactin XT superflow high capacity resin	IBA	Cat#2-4030-010
BioBeads SM-2 resin	Bio-Rad	Cat#1523920
Superdex 200 Increase 10/300 GL	GE Healthcare	Cat#28990944
Superose 6 Increase 10/300 GL	GE Healthcare	Cat#29091596
MAbPac Reversed Phase HPLC Column	Thermo Fisher Scientific	Cat#303184
Micro Bio-Spin Columns with Bio-Gel P-6	Bio-Rad	Cat#7326221
PepMap100 C$_{18}$ resin	Thermo Fisher Scientific, packed in-house	Cat#160454
Poroshell EC-C$_{18}$ 2.7 μm resin	Agilent Technologies, packed in-house	Cat#899999-777
Sep-Pak C$_{18}$ 1 cc Vac Cartridge	Waters	Cat#WAT054955
Deposited data		
Electron microscopy structure of SPC-A	This study	PDB: 7P2P
Electron microscopy density map of SPC-A in PMAL-C8	This study	EMDB: EMD-13171
Electron microscopy structure of SPC-C	This study	PDB: 7P2Q
Electron microscopy density map of SPC-C in PMAL-C8	This study	EMDB: EMD-13172
Mass spectrometry proteomics data	This study	http://proteomecentral.proteomexchange.org; ProteomeXchange: PXD022002; detailed listing in Table S1
Protein sequences for SPC subunits from various organisms	UniProtKb	https://www.uniprot.org/help/uniprotkb; detailed accession numbers in Figure S1
Signal peptide analysis	This study	Liaci and Förster, 2021. Mendeley Data: 10.17632/p65tkrr89v.1
Raw images of SDS-PAGE, Activity assays, and representative micrographs	This study	Liaci and Förster, 2021. Mendeley Data: 10.17632/whvwv62rn4.1
Experimental models: Cell lines		
HEK293E$^+$ cells	U-Protein Express BV	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
HAP1 cells	Essletzbichler et al., 2014	N/A
Raji cells	Emmanuel Wiertz	N/A
A549 cells	Emmanuel Wiertz	N/A
HEPG2 cells	Emmanuel Wiertz	N/A
U937 cells	Emmanuel Wiertz	N/A
HEK293F cells	Emmanuel Wiertz	N/A
RPE1 cells	Emmanuel Wiertz	N/A
Meljuso cells	Emmanuel Wiertz	N/A
Oligonucleotides		
E.coli pre-β-lactamase mRNA	Promega	Included with Cat#Y4041
Recombinant DNA		
Plasmid: pUPE-2961 SPC-A-(GFP-Strep-HA)	This study	N/A
Plasmid: pUPE-2961 SPC-C-(GFP-Strep-HA)	This study	N/A
Plasmid: pUPE-2961 SPC-A[S56A]-(GFP-Strep-HA)	This study	N/A
Plasmid: pUPE-2961 SPC-A[D116N]-(GFP-Strep-HA)	This study	N/A
Plasmid: pUPE-2961 SPC-A[D116R/R97D]-(GFP-Strep-HA)	This study	N/A
Plasmid: pUPE-2961 SPC-A[D121N]-(GFP-Strep-HA)	This study	N/A
Plasmid: pUPE-2961 SPC-A[D122N]-(GFP-Strep-HA)	This study	N/A
Plasmid: pUPE-2961 SPC-C[S68A]-(GFP-Strep-HA)	This study	N/A
Plasmid: pUPE-2961 SPC-C[D128N]-(GFP-Strep-HA)	This study	N/A
Plasmid: pUPE-2961 SPC-C[D128R/R109D]-(GFP-Strep-HA)	This study	N/A
Plasmid: pUPE-2961 SPC-A[D133N]-(GFP-Strep-HA)	This study	N/A
Plasmid: pUPE-2961 SPC-A[D134N]-(GFP-Strep-HA)	This study	N/A
Plasmid: pUPE-2961 SPC25-12-22/23	This study	N/A
Plasmid: pUPE-2961 SEC11A-(Strep)	This study	N/A
Plasmid: pUPE-2961 SEC11C-(FLAG)	This study	N/A
Plasmid: pUPE-2961 SPC-A[ΔCTS]-(gGFP-Strep-HA)	This study	N/A
Plasmid: pUPE-2961 SPC-C[ΔCTS]-(gGFP-Strep-HA)	This study	N/A
Software and algorithms		
EPU	Thermo Fischer	https://www.thermofisher.com/us/en/home/electron-microscopy/products/software-em-3d-vis-epu-software.html
Warp	Tegunov and Cramer, 2019	http://www.warpem.com/warp/
Motioncor2	Zheng et al., 2017	https://emcore.ucsf.edu/ucsf-software
CTFFIND4	Rohou and Grigorieff, 2015	https://grigoriefflab.umassmed.edu/ctffind4
Relion (v3.1)	Scheres, 2012	https://www3.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SPHIRE crYOLO	Wagner et al., 2019	https://sphire.mpg.de/wiki/doku.php?id=pipeline:window:cryolo
SPHIRE JANNI	Wagner, 2020	https://sphire.mpg.de/wiki/doku.php?id=janni
trRosetta	Yang et al., 2020	https://yanglab.nankai.edu.cn/trRosetta/
Coot (v0.9)	Emsley et al., 2010	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
PHENIX real space refine	Afonine et al., 2018	https://phenix-online.org/
PyMol	Schrödinger	https://pymol.org
UCSF Chimera	Pettersen et al., 2004	https://www.cgl.ucsf.edu/chimera/
UCSF ChimeraX	Goddard et al., 2018	https://www.cgl.ucsf.edu/chimerax/
VMD	Humphrey et al., 1996	https://www.ks.uiuc.edu/Research/vmd/
JPred4	Drozdetskiy et al., 2015	http://www.compbio.dundee.ac.uk/jpred/
NetSurfP-2.0	Klausen et al., 2019	http://www.cbs.dtu.dk/services/NetSurfP/
Mega-X	Kumar et al., 2018	https://www.megasoftware.net/
Consurf	Ashkenazy et al., 2016	https://consurf.tau.ac.il/
Martinize2	de Jong et al., 2013; Marrink Laboratory, 2020	https://github.com/marrink-lab/vermouth-martinize
INSANE	Wassenaar et al., 2015	http://www.cgmartini.nl/index.php/330-it-s-insane
GROMACS (version 2020, incl. thermostat, barostat and P.M. Ewald)	Abraham et al., 2015	https://manual.gromacs.org/documentation/2020/index.html
CHARMM-GUI	Lee et al., 2016	http://charmm-gui.org/
LINCS	Hess et al., 1997	http://www.cs.rug.nl/~bekker/publications/lincs.pdf
backward	Wassenaar et al., 2014	http://www.cgmartini.nl/index.php/downloads/tools/240-backward
MDLovoFit	Martínez, 2015	http://leandro.iqm.unicamp.br/mdlovofit/home.shtml
PR.ThermControl	NanoTemper	https://nanotempertech.com/prometheus-pr-thermcontrol-software/
MaxQuant	Cox and Mann, 2008	https://www.maxquant.org/
iBAQ	Schwanhäusser et al., 2011	N/A
UniDec	Marty et al., 2015	http://unidec.chem.ox.ac.uk/
Xtract/ReSpect	Thermo Fischer	https://www.thermofisher.com/order/catalog/product/IQLAAEGABSFANOMBAQ; Cat#OPTON-30932
Proteome Discoverer (version 2.4.0.305)	Thermo Fischer	https://www.thermofisher.com/us/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/proteome-discoverer-software.html
ProSightPD 3.0 Plugin for Proteome Discoverer	Zamdborg et al., 2007; Proteinaceous	https://proteinaceous.net/product/prosightpd/
XlinkX node for Proteome Discoverer, incl. PD node	Klykov et al., 2018; Thermo Fischer	https://www.thermofisher.com/order/catalog/product/OPTON-30946#/OPTON-30946; https://www.hecklab.com/software/xlinkx/
R	R Core Team	https://www.r-project.org/
ggplot2	Wickham, 2009	N/A
Weblogo	Crooks et al., 2004	https://weblogo.berkeley.edu/logo.cgi
Other		
Martini 3 coarse-grained force field	Souza et al., 2021	http://cgmartini.nl/index.php/force-field-parameters/particle-definitions
CHARMM force field for proteins	Huang et al., 2017	https://www.charmm.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Friedrich Förster (f.g.forster@uu.nl).

Materials availability

This study did not generate new unique reagents. Expression constructs are available from the lead contact upon request.

Data and code availability

- The cryo-EM 3D maps of human SPC-A and SPC-C have been deposited at the EMDB database and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. The corresponding atomic models were deposited at the RCSB PDB and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. The mass spectrometry proteomics data have been deposited at the ProteomeXchange Consortium and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. A detailed listing of the mass spectrometry datasets can be found in Table S1. The empirical signal peptide analysis (Liaci and Förster, 2021) as well as original SDS-PAGE, activity assay and representative micrograph images have been deposited to Mendeley Data and are publicly available as of the date of publication. DOIs are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Plasmids were obtained through expression in E.coli TOP10. Proteins were obtained through recombinant expression in HEK293 E$^+$ cells.

METHOD DETAILS

Design and cloning of expression constructs

All subunits of SPC-A (SPC12,25,22/23, and SEC11A) and SPC-C (SPC12,25,22/23, and SEC11C), respectively, were expressed from a single pUPE-2961 vector (U-Protein Express BV) in one large ORF. The individual subunits were separated by picornaviral 2A modules. Codon-optimized DNA constructs based on UniProtKB entries UniProt: Q9Y6A9-1 (SPC12), UniProt: Q15005 (SPC25), UniProt: P61009 (SPC22/23), UniProt: P67812 (SEC11A), and UniProt: Q9BY50 (SEC11C) were synthesized by Twist Bioscience and cloned into the vector backbone by Gibson assembly (New England Biolabs) in the sequence SPC25-[E2A]-SPC12-[P2A]-SPC22/23-[T2A]-SEC11A (SPC-A) or SPC25-[E2A]-SPC12-[P2A]-SPC22/23-[T2A]-SEC11C (SPC-C). The respective SEC11 subunit was C-terminally tagged with a TEV-cleavable eGFP-TwinStrep-HA tag.

For composition analysis, proteolytic subunits were removed from the expression constructs using blunt end deletion following a standard Q5 mutagenesis workflow (New England Biolabs). Additionally, SEC11A and SEC11C were separately cloned into pUPE-2961 by Gibson assembly. A C-terminal Strep tag was added to SEC11A, and a C-terminal FLAG tag was added to SEC11C by Q5 mutagenesis, and *vice versa*. Similarly, point mutants and ΔCTS constructs were generated from the parental constructs by Q5 mutagenesis. All constructs were evaluated by sequencing.

Protein expression and purification
Expression

All SPC constructs were transiently expressed for ~48h in suspension HEK293-E$^+$ cells by U-Protein Express BV (Utrecht, the Netherlands) using 0.5 mg vector DNA per L cell culture. The final cell densities ranged between 1-2 million cells per mL. All subsequent steps were performed at 4°C unless stated otherwise. Cells were pelleted by centrifugation at 500 g and washed three times with ice-cold PBS to remove biotin from the expression medium. The resulting cell pellets were flash-frozen in 0.5 L aliquots and

stored at −80°C until further use. For composition analysis, cells from 1 L culture were co-transfected with 0.333 mg of the vector containing the accessory subunits SPC25, SPC12, and SPC22/23 as well as 83.3 µg of SEC11A-Strep and SEC11C-FLAG, respectively.

For assays: Purification in digitonin

Cell pellets from 0.2-0.5 L culture medium were thawed in 35 mL lysis buffer per L cell culture (50 mM HEPES pH 7.8 100 mM NaCl 5 mM EDTA, 1 mM DTT, 10% (v/v) glycerol, 0.7 µg/mL DNase I, 1% (w/v) digitonin) and incubated 1.5 h at 4°C in a rotating wheel. Since the SPC is resistant to common protease inhibitors (Jackson and Blobel, 1980), one cOmplete inhibitor tablet (Roche, containing EDTA) was added during cell lysis. Samples were cleared by ultracentrifugation at 100,000 g for 30 min in a fixed-angle rotor. The resulting supernatant was immobilized twice on 5 mL pre-equilibrated Streptactin XT high capacity beads (IBA) in a gravity flow column, and the immobilized sample was washed with 20 column values (CV) wash buffer (20 mM HEPES pH 7.8, 85 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% (v/v) glycerol, 0.1% (w/v) digitonin). Retained SPC was eluted with 5-10 CV elution buffer (100 mM HEPES pH 7.8, 85 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% (v/v) glycerol, 0.09% (w/v) digitonin, 50 mM biotin). These preparations were used for XL-MS and to analyze the SPC point mutants and ΔCTS constructs. For the point mutants and ΔCTS constructs, samples were buffer-exchanged (1:400) into wash buffer and concentrated to 1.0 mg/mL.

A similar protocol was used for composition analysis, except that DTT was omitted from all buffers until affinity chromatography was completed. 300 mL of centrifuged cell lysate were split into two batches and immobilized four times on two gravity flow columns - one containing 0.5 mL Streptactin XT high capacity beads (IBA) and one containing 0.5 mL anti-FLAG M2 affinity resin (Sigma-Aldrich). After each round of immobilization, the flow through from each column was re-applied to the other column. Both columns were washed with 40 CV of wash buffer without DTT before elution at room temperature with 10x 1 CV of elution buffer without DTT. For the FLAG resin, 50 mM biotin were replaced with 200 µg/mL 3x FLAG peptide (Sigma-Aldrich). Samples were buffer-exchanged into wash buffer containing DTT (dilution factor 1:400) and concentrated to 1 mg/mL.

For cryo-EM: Purification in PMAL-C8

Cell pellets were thawed in 35 mL EM lysis buffer (50 mM HEPES pH 7.8, 100 mM NaCl 5 mM EDTA, 1 mM DTT, 10% (v/v) glycerol, 1 large Roche cOmplete inhibitor tablet (containing EDTA), 0.7 µg/mL DNase I, 1% (w/v) DDM, and 0.2% (w/v) CHS) per L cell culture and incubated 1.5 h at 4°C in a rotating wheel. Samples were cleared by ultracentrifugation at 100,000 g for 30 min in a fixed-angle rotor. The resulting supernatant was immobilized twice on 5 mL Streptactin XT high capacity beads in a gravity flow column, and the immobilized sample was washed with 20 CV EM wash buffer (20 mM HEPES pH 7.8, 85 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% (v/v) glycerol, 0.0174% (w/v) DDM, 0.00348% (w/v) CHS). Retained SPC was eluted with 5-10 CV EM elution buffer (100 mM HEPES pH 7.8, 85 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% (v/v) glycerol, 0.0174% (w/v) DDM, 0.00348% (w/v) CHS, 50 mM biotin). The eluate was concentrated to 1 mg/mL, diluted 1:1 (v/v) with a buffer containing 10 mM HEPES pH 7.8, 10% (v/v) glycerol, 1 mM EDTA, 1 mM DTT and incubated 1h at 4°C with PMAL-C8 (Anatrace, mass ratio protein:PMAL-C8 1:2.25). 30 mg BioBeads (BioRad) and TEV protease (laboratory stock, 0.6 mg/mL in 500 mM NaCl, 60% (v/v) glycerol, added at a 1:6 v/v ratio) were added, and the samples were incubated 30 min at rt followed by an overnight incubation at 4°C. BioBeads were removed by centrifugation, and the samples were exchanged into EM amphipol buffer (10 mM HEPES pH 7.8, 85 mM NaCl, 10% glycerol, 1 mM EDTA, 1 mM DTT) using a HiTrap column (GE Healthcare). The resulting eluate was passed over 1 mL Streptactin XT high capacity beads (IBA) on a gravity flow column and washed with 5 CV EM amphipol buffer to remove non-cleaved SPC. The flowthrough was collected, concentrated, and applied to size exclusion chromatography on a Superdex 200 increase 10/300 column (GE Healthcare) equilibrated with size exclusion buffer (10 mM HEPES pH 7.8, 85 mM NaCl, 1 mM EDTA, 1 mM DTT) (Figure S2E). Peak fractions were combined and concentrated again if necessary.

Activity assay

Activity assays are based on a commonly used *in vitro* cleavage protocol (Jackson, 1983; Walker and Lively, 2013). ^{35}S-labeled pre-β-lactamase was generated by *in-vitro* translation. Per 5 µL reaction, 25 ng *E.coli* pre-β-lactamase mRNA (Promega) were incubated with 2.5 µL rabbit reticulocyte lysate (Green Hectares), 0.5 mM DTT, and 1 µL EasyTag Express ^{35}S protein labeling mix (PerkinElmer) in the presence of 0.1% (w/v) digitonin. *In vitro*-translation was allowed to proceed for 60 min at 30°C. 2 µL of *in vitro* translated pre-β-lactamase were then incubated with 2 µL SPC (1.0 mg/mL) at 25°C for 90 min. The samples were denatured for 15 min at 70°C in reducing SDS-sample buffer and resolved on a 15% Tris-glycine SDS-PAGE. Gels were dried and exposed to Kodak MR films (Kodak) over night. [35]S signal collected on the phosphor screens was scanned using a Typhoon FLA-7000 scanner (GE Healthcare). Due to the poor accessibility of the substrate signal peptide *in vitro*, partial cleavage of the precursor protein is expected (Jackson, 1983).

Differential scanning fluorimetry

Melting profiles were acquired using a Prometheus NT.48 (NanoTemper). Experiments were performed using standard capillaries and a sample volume of 12 µL per capillary. SPC samples at 1.0 mg/mL were heated from 20°C to 90°C with 1°C/min. UV absorbance at 350 and 330 nm were recorded at 10% excitation power. To determine the melting onset (T_{on}) and melting point (T_m), the shift in native tryptophan fluorescence was monitored by plotting changes in the emission at 350 and 330 nm. T_{on} and T_m were determined automatically using PR.ThermControl (NanoTemper). All experiments were performed in duplicates.

Shotgun mass spectrometry

Sample preparation

Wild-type cell lines were cultured as follows: U937, Raji and Meljuso in RPMI-1640, 5% FBS, glutamine and penicillin-streptomycin (glu + Pen-Strep); A549 in DMEM, 5% FBS, glu + Pen-Strep; RPE1 in DMEM, 10% FBS, glu + Pen-Strep; HAP1 (Essletzbichler et al., 2014) in IMDM, 10% FBS and Pen-Strep. HepG2 in RPMI-1640, 10% FBS, glu + Pen-Strep; RPMI-8226 in RPMI-1640, 10% FBS, HEPES, glu + Pen-Strep. RPMI-8226 and U937 cells were cultured in suspension in T75 flasks at 37°C and 5% CO_2. All other cell lines were cultured as a monolayer in T75 flasks at 37°C with 5% CO_2. Monolayer cultures were dislodged using 0.05% Trypsin-EDTA. 10^6 cells were centrifuged, washed once with PBS, snap frozen, and stored at −80°C. Following mild centrifugation, the cell pellets were resuspended at a concentration of 1e6 cells per 100 μL of ice-cold lysis buffer containing 8 M urea, 50 mM Tris (pH 8), and EDTA-free mini protease inhibitor cocktail cOmplete (Sigma-Aldrich). The resuspended cell mixture was vortexed gently for 10 min and soni-cated for five cycles of 30 s with a Bioraptor Plus (Diagenode SA) at 4°C. The final protein concentration was measured using a Pierce BCA protein assay kit (Thermo Fisher Scientific). An aliquot of ∼50 μg cell lysate was reduced with freshly dissolved 10 mM DTT for 30 min at 37°C and alkylated with freshly dissolved iodoacetamide (IAA; Sigma-Aldrich) for 30 min at 37°C in the dark. LysC (Wako Chemicals) was added at a 1:50 (w/w) enzyme-to-protein ratio and the mixture was incubated for 3 h at 37°C. Next, the mixture was diluted to 2 M final urea concentration with 50 mM ammonium bicarbonate. Trypsin (Sigma-Aldrich) was added in 1:20 (w/w) enzyme-to-protein ratio. Digestion was performed overnight at 37°C, followed by quenching with 0.1% trifluoroacetic acid (TFA). The digest was desalted using a Sep-Pac C_{18} 1cc vacuum cartridge (Waters). The cartridge was washed twice with 100% acetonitrile (ACN) and twice with 0.1 M acetic acid prior to sample loading. Elution was done with 80% (v/v) acetonitrile (ACN) and 0.1 M acetic acid in Milli-Q water. The desalted peptides were lyophilized by vacuum centrifugation to near-complete dryness. The final peptide mixture was resuspended in 2% (v/v) formic acid prior to LC-MS/MS data acquisition.

LC-MS/MS data acquisition

All data were acquired using a Thermo Scientific Ultimate 3000 RSLCnano system coupled on-line to an Orbitrap HF-X mass spec-trometer (Kelstrup et al., 2018; Scheltema et al., 2014) (Thermo Scientific). Peptides were first trapped on the trapping cartridge (Pep-Map100 C_{18}, 5 μm, 5 mm × 300 μm; Thermo Scientific) prior to separation on an analytical column (Poroshell EC-C_{18}, 2.7 μm, 50 cm × 75 μm; packed in-house), heated to 40°C. Trapping was performed for 1 min in solvent A (0.1% v/v formic acid in water), and the gradient was as follows: 9%–13% solvent B (0.1% v/v formic acid in 80% v/v ACN) over 1 min, 13%–44% solvent B over 95 min, 44%–99% solvent B over 3 min, and finally 99% B for 4 min (flow was set to 300 nL/min). Mass spectrometry data was collected in a data-dependent fashion with survey scans from m/z 300 to 1500 Th (resolution of 60,000 at m/z = 200 Th), and up to 15 of the top precursors selected and fragmented using higher-energy collisional dissociation (HCD) with a normalized collision energy of value of 27%. The MS2 spectra were recorded at a resolution of 15,000 (at m/z = 200 Th). The AGC targets for both MS and MS2 scans were set to standard within a maximum injection time of 50 and 35 ms, respectively.

Data analysis

Raw data were processed using the MaxQuant computational platform (Cox and Mann, 2008) with standard settings applied. In short, the extracted peak lists were searched against the reviewed Human UniProtKB database (date 15-07-2020; 20353 entries), with an allowed precursor mass deviation of 4.5 ppm and an allowed fragment mass deviation of 20 ppm. MaxQuant by default en-ables individual peptide mass tolerances, which was used in the search. Cysteine carbamidomethylation was set as static modifi-cation, and methionine oxidation and N-terminal acetylation as variable modifications. The iBAQ algorithm was used for calculation of approximate abundances for the identified proteins (Schwanhäusser et al., 2011), which normalizes the summed peptide inten-sities by the number of theoretically observable peptides of the protein.

Native mass spectrometry

Sample Preparation

Samples were stored at −80°C in either digitonin or amphipol (PMAL-C8)-containing buffer prior to native MS analysis. Approxi-mately 10-20 μg of the membrane protein complex was concentrated and buffer-exchanged into 150 mM aqueous ammonium ac-etate (pH 7.5) with 0.01% (w/v) DDM, by using gel filtration with P-6 Bio-Spin columns (BioRad). The resulting protein concentration was estimated to be ∼1-2 μM before native MS analysis.

Data acquisition

Samples containing the membrane protein complex were directly infused into a Q Exactive Ultra High Mass Range Orbitrap instru-ment (QE-UHMR) (Thermo Fisher Scientific, Bremen) by using in-house prepared gold-coated borosilicate capillaries. Mass spec-trometer parameters were used as follows: capillary voltage at 1.2 kV, positive ion mode, source temperature at 250°C, S-lens RF level at 60, injection time at 50 ms, noise level parameter at 3.64. To release the membrane proteins from the detergent micelles, in-source trapping was used with a desolvation voltage of −200 V without additional collisional activation. Automatic gain control (AGC) mode was set to fixed. Resolution at 8,750 (at m/z = 200 Th), which corresponds to a 32 ms transient. Ion transfer optics and voltage gradients throughout the instrument were manually tuned to achieve optimal transmission of the membrane protein com-plex. Nitrogen was used in the higher-energy collisional dissociation (HCD) cell with trapping gas pressure setting set to 3, which corresponds to ∼2.2e-10 mBar ultra-high vacuum (UHV). The instrument was calibrated in the m/z range of interest using an aqueous cesium iodide solution. Acquisition of spectra was performed by averaging 1000 μscans in the time domain and subsequently

recording 10 scans (2 μscans each). Peaks corresponding to the protein complex of interest were isolated with a 10 Th isolation window and probed for fragmentation using elevated HCD voltages, HCD direct eV setting of 100-150 V.

Data Analysis

Raw native spectra were deconvoluted with UniDec (Marty et al., 2015) to obtain zero-charged masses. For annotation, masses of the ejected subunits obtained upon activation of the membrane protein complexes were matched to the subunits identified in top-down LC-MS/MS experiment. For the reconstruction of the native MS spectrum from top-down MS data, distinct proteoforms of SPC22/23 and SEC11A or SEC11C proteins were randomly combined to obtain the masses of dimers and the products of corresponding abundances were used as abundances of the dimers. Final reconstructed native spectra were overlaid with respective native spectra obtained for the catalytic dimers of SPC-A and SPC-C complexes.

Intact mass and top-down mass spectrometry

Sample preparation

Samples stored in digitonin or amphipol-containing buffer were diluted to a final concentration of 0.2 μg/μL. Approximately 1 μg of sample was injected for a single top-down LC-MS(/MS) experiment.

LC-MS(/MS) data acquisition

Chromatographic separation was achieved by using a Thermo Scientific Vanquish Flex UHPLC instrument coupled on-line, via a 1 mm x 150 mm MAbPac reversed-phase analytical column, to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher). The column compartment and the column preheater were heated to 80°C during analysis. Membrane proteins were separated over a 22 min LC-MS/MS run at a flow rate of 150 μL/min. Gradient elution was performed using mobile phases A (Milli-Q water/0.1% formic acid) and B (acetonitrile/0.1% formic acid) with 30 to 57% B ramp-up in 14 min. LC-MS(/MS) data were collected with the mass spectrometer set to Intact Protein / Low Pressure. Two acquisition approaches were used with complementary full MS resolutions of either 7,500 or 120,000 (both at m/z = 200 Th). At 7,500 ions with masses above ~30 kDa can be detected and at 120,000 ions with masses below ~30 kDa can be resolved with accurate masses. Full MS scans were acquired for a mass range of m/z 500-3,000 Th with the AGC target set to 250% with a maximum injection time of 50 ms for the resolution of 7,500 and 250 ms for the resolution of 120,000. A total of 2 μscans were averaged and recorded for the 7,500 resolution scans and 5 μscans for the 120,000 resolution scans. All MS/MS scans were acquired with a resolution of 120,000, a maximum injection time of 250 ms, an AGC target of 10,000% and 5 μscans for the most or the first 2 most intense proteoform(s) in each cycle for medium and high resolution, respectively. The ions of interest were mass-selected by quadrupole isolation in a m/z = 4 Th window and collected to an AGC Target of 5e6 ions prior to electron transfer dissociation (ETD). The ETD reaction time was set to 16 ms with a maximum injection time of 200 ms and the AGC target of 1e6 for the ETD reagent. For data-dependent MS/MS acquisition strategy, the intensity threshold was set to 2e5 of minimum precursor intensity. MS/MS scans were recorded in the range of m/z = 350-5000 Th using high mass range quadrupole isolation.

Data Analysis

Full MS spectra were deconvoluted with either Xtract or ReSpect (Thermo Fisher Scientific) for isotopically-resolved or unresolved data, respectively. Automated proteoform searches against a custom sequence database were performed in Thermo Proteome Discoverer (version 2.4.0.305) extended with the ProSightPD 3.0 nodes (Zamdborg et al., 2007). Parameters were set as follows. ReSpect: precursor m/z tolerance – 0.2 Th; relative abundance threshold – 0%; precursor mass range – 3-100 kDa; precursor mass tolerance – 30 ppm; charge range – 3-100. Xtract: signal/noise threshold – 3; m/z range – 500-3,000 Th. Initially, a large precursor tolerance window of 5 kDa was used to allow for detection of unknown PTMs and sequence processing followed by cycles of database filtering and sequence adjustment to determine a final set of isoforms/proteoforms. For the final database search, ProSightPD parameters were: precursor mass tolerance – 500 Da; fragment mass tolerance – 20 ppm. To verify unreported isoforms/proteoforms, in-house R and C# scripts were used to group replicate fragmentation scans for a precursor of interest followed by automated fragment annotation and manual spectrum inspection. A similar approach was used to characterize unidentified abundant precursors (Albanese et al., 2020). Representation of proteoforms per protein was achieved by summing full MS scans per protein elution peak and converting spectra to zero-charged mass profiles in UniDec (Marty et al., 2015). Any additional visualization of the data was performed in R extended with the ggplot2 package (Wickham, 2009).

Cross-linking mass spectrometry

Sample preparation

Proteins were incubated with the cross-linking reagent PhoX (Steigenberger et al., 2019) for 45 min at room temperature (buffer conditions specified below). The cross-linking reaction was quenched by addition of Tris·HCl (100 mM, pH 7.5) to a final concentration of 10 mM. Cross-linked proteins were further purified from aggregation products by size exclusion chromatography on a Superose 6 increase column (GE Healthcare) equilibrated with size exclusion buffer (10 mM HEPES pH 7.8, 85 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.09% (w/v) digitonin). The concentration of peak fractions ranged between 0.2 mg/mL (SPC-A) to 0.4 mg/mL (SPC-C). Crosslinked proteins were denatured by addition of urea (8 M in 100 mM Tris) and reduced by addition of DTT (final concentration of 2 mM) for 30 min at 37°C, followed by alkylation with IAA (final concentration of 4 mM) for 30 min at 37°C. Afterward the sample was digested by incubation with a combination of LysC (1:75 enzyme to protein) and Trypsin (1:50 enzyme to protein) for 10 h at 37°C, after which formic acid (final concentration 1%) was added to quench the digestion. Finally, peptides were desalted by Sep-Pak C$_{18}$ prior to Fe-IMAC enrichment.

Cross-linked peptides were enriched with Fe(III)-NTA cartridges (Agilent Technologies) using the AssayMAP Bravo Platform (Agilent Technologies) in an automated fashion. Cartridges were primed at a flow rate of 100 μL/min with 250 μL of priming buffer (0.1% TFA, 99.9% ACN) and equilibrated at a flow rate of 50 μL/min with 250 μL of loading buffer (0.1% TFA, 80% ACN). The flow-through was collected into a separate plate. Dried samples were dissolved in 200 μL of loading buffer and loaded at a flow rate of 5 μL/min onto the cartridge. Cartridges were washed with 250 μL of loading buffer at a flow rate of 20 μL/min and cross-linked peptides were eluted with 35 μL of 10% ammonia directly into 35 μL of 10% formic acid. Samples were dried down and stored at −20°C prior to further use. Before to LC–MS/MS analysis, the samples were resuspended in 10% formic acid.

LC-MS/MS data acquisition

All data were acquired using an UHPLC 1290 system (Agilent Technologies) coupled on-line to an Orbitrap Fusion (Senko et al., 2013) mass spectrometer (Thermo Scientific). Peptides were first trapped (Dr. Maisch Reprosil C_{18}, 3 μm, 2 cm × 100 μm) prior to separation on an analytical column (Agilent Poroshell EC-C_{18}, 2.7 μm, 50 cm × 75 μm). Trapping was performed for 10 min in solvent A (0.1% v/v formic acid in water), and the gradient for protein complexes was as follows: 0 – 10% solvent B (0.1% v/v formic acid in 80% v/v ACN) over 5 min, 10- 40% solvent B over 70 min, 40%–100% solvent B over 3 min, and finally 100% B for 4 min (flow was passively split to approximately 200 nL/min). The mass spectrometer was operated in a data-dependent mode. Full-scan MS spectra were collected in a mass range of m/z 350 – 1300 Th in the Orbitrap at a resolution of 60,000 after accumulation to an AGC target value of 1e6 with a maximum injection time of 50 ms. In-source fragmentation was activated and set to 15 eV. The cycle time for the acquisition of MS/MS fragmentation scans was set to 3 s. Charge states included for MS/MS fragmentation were set to 3-8, respectively. Dynamic exclusion properties were set to n = 1 and to an exclusion duration of 20 s. HCD fragmentation (MS/MS) was performed in stepped collision energy mode (31.5, 35, 38.5%) in the Ion Trap and the mass spectrum acquired in the Orbitrap at a resolution of 30,000 after accumulation to an AGC target value of 1e5 with an isolation window of m/z = 1.4 Th and maximum injection time of 120 ms.

Data analysis

The acquired raw data were processed using Proteome Discoverer (version 2.4.0.388) with the XlinkX/PD nodes integrated (Klykov et al., 2018; Steigenberger et al., 2019). The linear peptides search was performed using the standard Mascot node as the search engine with the full Human database from UniProtKB (20,230 entries, downloaded from UniProtKB downloaded at 2018_01). Cysteine carbamido-methylation was set as fixed modification. Methionine oxidation and protein N-term acetylation was set as dynamic modification. For the search of mono-links, water-quenched ($C_8H_5O_6P$) and Tris-quenched ($C_{12}H_{14}O_8PN$) were set as dynamic modifications. Trypsin/P was specified as the cleavage enzyme with a minimal peptide length of six and up to two miss cleavages were allowed. Filtering at 1% false discovery rate (FDR) at the peptide level was applied through the Percolator node. For crosslinked peptides, a database search was performed against a FASTA containing the proteins under investigation supplemented with a common contaminants list of 200 proteins using XlinkX/PD nodes for cross-link analysis. Cysteine carbamidomethylation was set as fixed modification and methionine oxidation and protein N-term acetylation were set as dynamic modifications. Trypsin/P was specified as enzyme and up to two missed cleavages were allowed. Furthermore, identifications were only accepted with a minimal score of 40 and a minimal delta score of 4. Otherwise, standard settings were applied. Filtering at 1% false discovery rate (FDR) at peptide level was applied through the XlinkX Validator node with setting simple.

Single particle analysis

Sample preparation

Size exclusion chromatography (SEC) peak fractions were either used directly for vitrification (at 0.5 mg/mL) or concentrated to 4 mg/mL and vitrified in the presence of 1.5 mM fluorinated fos-choline (FFosC, Anatrace). 3 μL samples were applied to freshly glow-discharged Cu 200 Holey Carbon R1.2/1.3 grids (Quantifoil, for samples without FFosC), or Cu 200 Holey Carbon R2/1 grids (Quantifoil, for samples containing 1.5 mM FFosC). In both cases, grids were flash-frozen using a Vitrobot Mark IV (Thermo Fischer Scientific) with 595 blotting paper (Ted Pella) at 4°C, 100% humidity, and either a blot force of 0 for 4 s or a blot force of −2 for 3 s and a liquid ethane/propane mixture.

Data collection

Data were collected on a 200 kV Talos Arctica microscope (Thermo Fischer Scientific) equipped with a post-column energy filter (slit width 20 eV) and a K2 summit direct electron detector (Gatan). EPU (Thermo Fischer Scientific) was used for automated data collection in counted mode. Movies were acquired in 45-50 frames at an effective pixel size of 0.81 Å/px, with a dose rate of ∼4 e⁻/px/s (measured in an empty hole without ice), and a total dose of 60 e⁻/Å². Defocus values ranged between 0.5 and 4 μm. Data quality was monitored in real time using Warp (Tegunov and Cramer, 2019).

Image processing

All four datasets collected for SPC-A and SPC-C in PMAL-C8 were processed analogously (for detailed dataset statistics, refer to Table 1 and Figures S2–S4). Collected movie stacks were manually inspected and imported into Relion 3.1 (Scheres, 2012). Motion correction was performed with MotionCor2 (Zheng et al., 2017), and CTFFIND4 (Rohou and Grigorieff, 2015) was used for CTF estimation using exhaustive search in Relion 3.1. Movies with an estimated resolution worse than 10 Å were discarded. For particle picking, the motion-corrected micrographs were denoised using a trained model in SPHIRE-JANNI (Wagner, 2020). Particle picking was performed with SPHIRE-crYOLO 1.5.5 (Wagner et al., 2019), using models trained separately on subsets with similar defocus (0-1 μm, 1-2 μm, 2-3 μm, and 3-4 μm) to ensure maximal particle recovery across the whole defocus range. Particles

were extracted in Relion 3.1 at 3-fold binning and subjected to 3 rounds of 2D classification, during which information until the first CTF zero was ignored (Figure S2). Once 2D classification converged, the respective datasets with and without FFosC were combined and subjected to 2 additional rounds of 2D classification. Unbinned images of the remaining particles were used to generate an initial 3D model, followed by 3D refinement, and re-extraction with updated coordinates. Since the micelle occupies about 50% of the particle volume, it was expected to have detrimental effects on particle alignment. Therefore. the refined map was used as a reference for a 3D classification without particle alignment using a mask enclosing only the protein portion. The respective best classes were subjected to CTF correction, Bayesian polishing, and 3D refinement in Relion 3.1. Further attempts to perform CTF corrections and subtract the micelle did not improve the reconstructions. A post-processing step in Relion was employed to partially mask the micelle, correct with the detector MTF, and apply a sharpening B-factor of -180 Å^2.

Model building & refinement

All models were generated by subjecting the full-length proteins to structure prediction by trRosetta (Yang et al., 2020; Figure S3). The precise pixel spacing was determined by map correlation of the SEC11 luminal domain in UCSF Chimera (Pettersen et al., 2004). Models for SPC12 and SPC25 were docked as complete rigid bodies, whereas models of SEC11A and SEC11C were separated into two rigid body groups: (i) the N–terminal part containing the cytosolic portion and the single TM helix, and (ii) the luminal domain including the C terminus. Both groups were docked individually in UCSF Chimera. Similarly, SPC22/23 was divided into three rigid body groups: (i) the N-terminal TM helix, (ii) a short strand with poor density fit that connects the TM helix and the rest of the protein, and (iii) the luminal domain. An additional model containing only the luminal portion was calculated and used to mitigate influences of the TM helix on the fold of the luminal portion. All domains except SPC12 could be docked unambiguously based on their soluble parts (Figure S3). SPC12 was fitted based on the different lengths of its two TM segments and the presence of a hydrophilic stretch that is unlikely to be exposed to the membrane interface. Existing information in the literature, top-down, native, and XL-MS data, the presence of known anchor points (glycosylation site at Asp141 of SPC22/23, visible tryptophan side chains), secondary structure predictions, disorder predictions, atomistic molecular dynamics simulations of the protein in POPC using a CHARMM force field (see below), differences in the top-5 trRosetta models, and overall biochemical properties of the resulting model were used to evaluate the docking and domain architecture (Figure S4).

Unresolved parts of the assembled SPC were trimmed manually in *Coot* version 0.9 (Emsley et al., 2010; Figure S3). Alternating rounds of manual adjustment in *Coot* and PHENIX real space refine (Afonine et al., 2018) were applied to yield the final models. In *Coot*, all-atom restraints and tight geometry restraints were used at all times. Only loops with poor density fit and, if necessary, side chains with low-probability rotamers were adjusted. In PHENIX, morphing and simulated annealing were applied to enable fitting of loops. Map-model FSCs were calculated using PHENIX Mtriage. All figures were prepared using UCSF ChimeraX (Goddard et al., 2018) or PyMol (Schrödinger).

Surface potential calculation

Surface potentials were calculated in vacuum using the APBS plugin in PyMol (Schrödinger), based on pKa calculations using the built-in pdb2pqr module.

Secondary structure and disorder prediction

Secondary structure elements were predicted using JPred4 (Drozdetskiy et al., 2015). Disordered regions were predicted using the HHblits method with NetSurfP-2.0 (Klausen et al., 2019).

Phylogenic analysis and residue conservation

Representative, curated protein sequences of SPC subunits from all major branches of life were manually extracted from UniProtKB. Sequence alignments and phylogenetic trees were computed using MUSCLE in Mega-X (Kumar et al., 2018). Residue conservation was calculated using the Consurf server (Ashkenazy et al., 2016) searching against the cleaned UniProtKB database using the HMMER algorithm.

Molecular dynamics simulations

Coarse-grained models

The most recent development version of the Martini 3 Coarse-Grained (CG) force field (Souza et al., 2021) was used to perform all CG molecular dynamics (MD) simulations. The CG protein model was generated with the new version of the program Martinize (de Jong et al., 2013; Marrink Laboratory, 2020). The refined full-length SPC-C model generated by trRosetta and fitted into the observed density was used as a reference to define bonded parameters dependent of the secondary structure. The glycan chain attached to Asp141 of SPC22/23 was not included in the model. Elastic networks were applied to each monomer of the SPC-C complex, with a distance cut-off of 0.85 nm using a force constant of 1300 kJ mol^{-1} nm^{-2}. Unresolved parts of SPC-C were kept free, without any elastic bonds. Additional harmonic bonds were added between the protein monomers, to further increase stability of the SPC-C complex. With a cut-off of 0.7 nm, these extra harmonic potentials mimic hydrogen bonds between the backbone beads of different monomers.

In addition to the SPC-C complex, two SP structures with sequence MIKKKKL(n)VPSAQAAYVSSSDSG based on SPs sampled in Nilsson et al. (1994, 2002) were generated with PyMol (Schrödinger), varying only the number of leucines in the h-region: 11 (called

here L11) and 20 (L20). The h-regions were modeled in a helical conformation, while the other regions were kept in coil/extended configurations. These structures were used as reference to generate the CG models in Martinize (de Jong et al., 2013; Marrink Laboratory, 2020), with the same parameters for the elastic network applied to SPC-C, which was only applied to the h-region of the sequences. All CG protein models also included addition of side chain corrections (Herzog et al., 2016) (except for coil/extended regions), based on reference structures. Lipids models were inspired on the previous Martini 2 force-field (Marrink et al., 2007; Wassenaar et al., 2015), but now following the rules of Martini 3 and with adaptations in the bonded parameters inspired by the "extensible model" of Carpenter et al. (2018).

Setup of coarse-grained MD simulations

SPC-C was embedded in different membrane environments including an endoplasmic reticulum membrane model, as described in Table S2. All systems were solvated using a Martini water model solution with 0.15 M concentration of NaCl, mimicking physiological conditions. All simulation boxes were built using the INSANE program with dimensions of $18 \times 18 \times 15 \ nm^3$ (Wassenaar et al., 2015). The principal axis of the SPC-C complex was set to be parallel to the normal of the lipid bilayers. First, the system was minimized for 2000 steps with the steepest descent method, followed by an equilibration stage performed for 500 ps with 10 fs as time step. After minimization and equilibration, the production run was performed for 20 μs, using a time step of 20 fs. In the case of systems containing the SPs L11 or L20, the MD simulations were performed 5 times, assuring enough sampling of the possible binding modes between SPC-C and the respective SP. Each SP was initially placed at a 2.8 nm distance of the TM window. Settings for the CG simulations followed the "new" Martini set of simulation parameters (de Jong et al., 2016). The temperature of the systems was kept at 310 K with the velocity rescaling thermostat (Bussi et al., 2007). For the pressure, we used semi-isotropic coupling at 1 bar using the Parrinello and Rahman (1981) barostat. Additional MD simulations of pure lipid bilayers were performed to be compared with SPC-C embedded in the same environments. All simulations were performed with GROMACS (version 2020) (Abraham et al., 2015).

Setup of atomistic MD simulations

The CHARMM force field for proteins (Huang et al., 2017) and lipids (Klauda et al., 2010) was used to perform the atomistic MD simulations. Water was modeled explicitly using the modified CHARMM TIP3P model (Jorgensen, 1981; MacKerell et al., 1998). The temperature was coupled to a heat bath at 310 K, using the velocity rescale thermostat (Bussi et al., 2007). The pressure was kept at 1.0 bar, using a Parrinello and Rahman (1981) barostat with a compressibility of $4.5 \times 10^{-5} \ bar^{-1}$ and coupling time of 4.0 ps. Particle Mesh Ewald (Darden et al., 1993; Essmann et al., 1995) was used to compute the electrostatic interactions, with a real-space cut-off of 1.2 nm. Van der Waals interactions were switched to zero between 1.2 and 1.4 nm. Neighbor lists were updated every 10 steps. Bonds involving hydrogens were constrained using the LINCS algorithm (Hess et al., 1997). The integration time step used was 2 fs and the overall center of mass motion was removed every 10 steps. Simulation box and topologies were initially built with CHARMM-GUI (Lee et al., 2016). SPC-C was embedded in a POPC bilayer and solvated with water in a simulation box of $18 \times 18 \times 15 \ nm^3$. NaCl was added to bring the system to neutral charge and an ionic strength of 0.15 M. Minimization and equilibrations was based on the CHARMM-GUI protocol, followed by a 150 ns production run with time step of 2 fs. As the relaxation of the bilayer (which includes the thinning in the TM window) demanded longer simulations, an additional initial configuration of SPC-C in POPC was built from the final configuration of the 20 μs CG MD simulations. The CG configuration was backmapped to the atomistic resolution using the backward program (Wassenaar et al., 2014), which was followed by another 150 ns production run. All atomistic simulations were performed with GROMACS (version 2020) (Abraham et al., 2015).

Analysis of the trajectories

The thickness (d_i) of the membranes were estimate based on the average distance of the phosphate bead (in CG simulations) or phosphorus atom (in atomistic simulations) of the lipids using the *gmx density tool* of GROMACS (Abraham et al., 2015). The percentage of thinning (%T) was estimated based on the difference of thickness in the bulk membrane ($d_{membrane}$) and the lipids inside the TM window ($d_{TM \ window}$) of the SPC-C complex, according to Equation 1:

$$\%T = 100 \times \frac{d_{membrane} - d_{TM \ window}}{d_{membrane}} \qquad \text{(Equation 1)}$$

Lipid enrichment (%E_i) near the TM window was computed according to Equation 2:

$$\%E_i = 100 \times (x_i - m_i) \qquad \text{(Equation 2)}$$

where x_i is molar fraction of lipids near the TM window and (m_i) is the molar fraction of the lipids in relation of the whole membrane. m_i was simply defined by the composition of the membrane. On the other hand, x_i was defined as the number of contacts of the lipid with the TM window ($c_{i-TM \ window}$) in relation of the total number of contacts with all the lipids, according to Equation 3:

$$x_i = \frac{C_{i-TM \ window}}{\sum_j C_{j-TM \ window}} \qquad \text{(Equation 3)}$$

$c_{i-TM \ window}$ was computed using the *gmx mindist tool* of GROMACS (Abraham et al., 2015). A distance cut-off of 1.1 nm was used between the PO4 and ROH beads (from phospholipids and cholesterol, respectively), and the BB beads of the following residues around the TM helices of SPC-C complex: Chain A - Tyr31, Tyr32, Gln36, Tyr185, Val186 and Leu,187; Chain B – Ala8, Asn9 and Ser10; Chain C – Thr129, Ile130 and Tyr131; Chain D – Glu88, Gln89 and Met90.

Root mean square fluctuation (RMSF) calculation was performed using a Fortran program, based on the MDLovoFit code (Martínez, 2015). The alignment protocol of the Cα atoms starts with a standard rigid-body alignment of the structures using all Cα atoms as reference, followed by the calculation of the RMSF per residue for atoms. After the identification of the residues with RMSF lower than 2 Å, a new rigid body alignment uses only this subgroup of residues that represents the most rigid parts of SPC-C. This protocol is repeated until the RMSD and the residues used in the rigid subgroup are converged. As structure reference for the trajectory alignment, we used the SPC-C average structure from the trajectory.

The distance between each SP and the TM window was computed using the *gmx distance tool* of GROMACS (Abraham et al., 2015). This metric was calculated as the distance between the backbone beads of Ile61 of SEC11C and valine at the C-terminal end of the h-region (Val18 in L11, and Val26 in L20). For better visualization of the plots, the results were smoothed out using the moving average method, with a window of 50 steps.

Visual inspection and figure rendering of the trajectory snapshots were performed with VMD (Humphrey et al., 1996).

Empirical SP analysis

Experimentally verified SP sequences were extracted from UniProtKB (UniProt Consortium, 2021) using the following query: annotation:(type:positional signal length:[22 TO 200] evidence:experimental) taxonomy:(2759). Entries that did not contain SP sequences were removed manually, resulting in a set of 1,492 SPs. Additionally, computationally predicted human SPs with of 60 or more amino acids were extracted using the query annotation:(type:positional signal length:[60 TO 200]) AND organism'':*Homo sapiens* (Human) [9606],'' resulting in six additional unique sequences. Lastly, a set of nine well-documented long SPs were extracted manually (Liaci and Förster, 2021). The sequence 'KFEKLKFEKL' was appended to each extracted SP, and all sequences were submitted to SignalP 3.0, which is the latest iteration of SignalP that features SP region predictions using Hidden Markov models (HMM) (Bendtsen et al., 2004). Each residue was assigned to the region with the highest HMM probability (minimum 0.2). The data was split into the following subsets: (i) Sequences with cleavage site probabilities > 0.5 and identical experimental and predicted cleavage sites (n = 921); (ii) sequences with cleavage site probabilities > 0.5 and differing experimental and predicted cleavage sites (n = 277); (iii) sequences with cleavage site probabilities < 0.5 (n = 289), and (iv) sequences for which SP prediction failed completely (n = 5). Set (i) was used to generate Figures 6B and 6C, and is the set discussed in the main text. Sequence plots from the other subsets can be found in (Liaci and Förster, 2021).

In order to get a measure of the maximal possible h-region length, all 412 human entries were also manually adjusted to include stretches that were originally assigned to the neighboring regions and that consist of the hydrophobic amino acids Val, Leu, Ile, Phe, Met, or the less hydrophobic residues Trp, Ala, Tyr, Pro, or Gly. Additionally, one interruption by one polar, non-charged amino acid (Ser, Asn, Gln, His) was allowed on each side if followed by at least two hydrophobic residues. Pro-Pro, Pro-X-Pro, and Pro-Gly segments were considered disruptive for TM helices and thus not allowed, and only the first residue of such motives was added to the h-region. Lastly, a minimal c-region length of 4 resides was assumed (as is suggested by the structure). The raw data can be found in Liaci and Förster (2021).

QUANTIFICATION AND STATISTICAL ANALYSIS

In Figure 5, average percentage of thinning and the average enrichment were computed according to Equations 1 and 2, respectively, with the error calculated with block average approach.

In Figure S3, resolution estimations of cryo-EM density maps are based on the 0.143 Fourier shell correlation (FSC) criterion (Rosenthal and Henderson, 2003).