The biogenesis of multispanning membrane proteins at the Sec61-translocon and their integration into the lipid bilayer of the endoplasmic reticulum

---

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

Von

Marco Janoschke

aus Öhringen (Deutschland)

Basel 2020

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Dr. Martin Spiess

Dr. Simon Bernêche

Basel, den 11.12.2018

Prof. Dr. Martin Spiess
Dekan

Acknowledgements

First of all I would like to thank the whole Spiess Lab, especially my boss Martin Spiess for allowing me to be a part of his lab. I also thank Tina June for introducing me into the topic, Nicole Beuret for taking care of all the stuff that has to be done to keep a lab going and Mirjam Pennauer and Simon Schlienger for fruitful 'Friday talks'. Also the other members of the lab provided an atmosphere in which I liked to work in: Cristina Baschong, Ghaetan Bader, Dominik Buser, Valentina Millarte and Jennifer Reck.
I would also like to thank Simon Bernèche and Jean Pieters for being part of my committee and Dominic Höpfner with whom I had a nice cooperation concerning the cavinafungin topic, which ended up in a publication in which I was able to participate.

Last but not least, I would like to thank my family and friends: My parents, Gerhard und Sieglinde, who supported me through all the years and my wife Natalie. I would also like to thank my friends Albert, Thomas and Felix for all the non-scientific and happy hours we spend together.

Summary

The key entry point of most membrane proteins into the lipid bilayer is the Sec61/SecYEG translocon, that mediates the transfer of hydrophilic sequences across the membrane and integration of mostly apolar α-helical transmembrane domains into the lipid bilayer. Three distinct integrations steps can be distinguished: (1) a first hydrophobic signal sequence targets the protein to the translocon, integrates itself into the membrane, and initiates translocation of the downstream polypeptide. (2) A subsequent hydrophobic segment laterally exits the translocon into the bilayer and thus stops further transfer. (3) The next hydrophobic sequence triggers re-integration into the translocon, re-initiating polypeptide transfer. Successive stop-transfer and re-integration sequences result in complex multispanning proteins. The major determinant of membrane topology appears to be the hydrophobicity of transmembrane domains. This has been best demonstrated for potential stop-transfer segments, suggesting a sequence-autonomous thermodynamic equilibration between the hydrophilic environment of the translocon and the apolar lipid phase.In this thesis, we analyzed in detail the hydrophobicity threshold for a potential re-integration TM domain downstream of different cytoplasmic loop sequences. Surprisingly, we discovered a strong dependence on the length of this cytoplasmic sequence. Short sequences are facilitating re-integration, while long ones seem to impede it. This demonstrates, that re-integration is not independent from the sequence-context. Further investigations revealed that loop sequences containing isolated folding domains, intrinsically disordered sequences, or sequences with a high affinity for chaperones enhance the re-integration efficiency, whereas those with low affinity to chaperones, and fragments of natural protein domains impair re-integration. We propose that the latter sequences, as they collapse to molten globules – i.e. near-native conformation of high compactness with already pronounced secondary structure and increased amount of hydrophobic residues on the surface area – compete with the translocon for interaction with the potential transmembrane segment. Our results thus define the environment of the nascent polypeptide chain when re-integration can occur and may serve as a guide in de novo membrane protein design.

In a second part, we characterized the antiviral natural product cavinafungin as an inhibitor of signal peptidase for Dengue virus as well as host substrates, inhibiting biogenesis of viral proteins from a single precursor membrane polyprotein.

Table of Content

1. Introduction

1.1. Protein biogenesis in eukaryotic cells

Proteins were first described in the 18th century as an own class of biomolecules by Antoine Fourcroy. He first named them albuminoids, because of the similar characteristics of these biomolecules to egg albumin. In 1833 the chemist Gerardus Johannes Mulder discovered that the elemental analysis of egg albumin, fibrin and wheat gluten all resulted in basically the same empirical formula, suggesting that all these biomolecules are made up of one basic substance. However, the exact nature and function of proteins long remained elusive. While some amino acids were discovered already in the 19th century, only in the beginning of the 20th century it was stated by the german chemists Emil Fischer and Franz Hofmeister that proteins are mostly made up of amino acids. In the 1920s it was shown by James Sumner that the enzyme urease is a protein, appointing the first enzymatical function to a protein itself (Tanford and Reynolds, 2003).

Today we know that proteins fulfill a plethora of functions in a cell and are not only catalyzing numerous reactions, but are also involved in the regulation of an organism`s homeostasis, fulfill structural functions in the shaping of a cell, and act as transporters for other proteins, molecules and a great number of external signals.

Not only the function of proteins was of great interest for scientists during the last century, but also how proteins exist in a physiological environment. The first hypothesis that proteins are not just long polypeptide chains in solution, but are folding in three dimensions, was already established in the 1930s by William Astbury and others. In 1951, Linus Pauling proposed two different secondary structure elements: α- and β-type (Eisenberg, 2003). Alpha-helicity is a secondary structure that is energetically favored when compared to an unfolded sequence, due to an energetic sink resulting from the hydrogen-bond formation between the peptide bonds i and i+4. The helicity resulting from this hydrogen-bond formation buries the peptide bonds in the core or backbone of the secondary structure, while the amino acid side chains are oriented to the solvent. The different properties of each amino acid side chain make the α-helix an ideal secondary structure, since the characteristics of the helix can be modulated to match the characteristics of the solvent and other parts of the protein. The second secondary structure motif is called β-sheet and is made up of several β-strands. While

hydrogen-bonds in α-helices are formed within the strand, the hydrogen-bonds in β-sheets are formed between two β-strands, that are oriented towards each other either parallel or anti-parallel. In contrast to α-helices, the β-sheet is a nearly extended conformation, having the polypeptide backbone stretched and the amino acid residues pointing into opposite directions (Eisenberg, 2003).

A leading experiment in the field of protein folding was performed in 1961 by the group of Christian B. Anfinsen. His group revealed that the folding and unfolding process of ribonuclease A is reversible. Upon treatment with reducing and denaturing agents the protein unfolds and loses all enzymatic activity. After removal of these agents the protein refolds and regains enzymatic function (ANFINSEN et al., 1961). These findings lead to the thermodynamic hypothesis: the folding information of small domains is only encoded within the amino acid sequence and is not determined by external features. It states, that the native state of a protein corresponds to its free energy minimum. A totally unfolded and thus unstructured domain is high in energy, due to the exposure of hydrophobic amino acids into a hydrophilic environment. Several models were proposed to describe what happens during the early states of folding (Udgaonkar, 2013):

- The framework model: Protein folding is initiated by secondary structure formation, tertiary structures form due to diffusion and collision of the polypeptide (Ptitsyn, 1973).
- The nucleation model: Formation of native secondary structure elements starts locally at a nucleation point in the protein and is propagated through the polypeptide until the whole protein is folded (Wetlaufer, 1973).
- The hydrophobic collapse model: The unfolded polypeptide first collapses into a molten globule-state, due to the energetic pressure caused by the exposure of hydrophobic residues to the cytosol. Folding into the native structure is achieved by rearrangement of the molten globule-state into the native protein fold (Brylinski et al., 2006).
- The nucleation-condensation model: The model tries to unify the abovementioned models. Secondary and tertiary structures are formed simultaneously (Fersht, 1995).

These models are not mutually exclusive and are often pointing out extreme cases of protein folding that do not rule out other possibilities. In each of these models the folding process points towards low energy, resulting in a state that is called a molten globule (Ohgushi and Wada, 1983).

A molten globule is by definition a near-native conformation of high compactness with already pronounced secondary structure but only little tertiary structure interactions. It still exposes hydrophobic amino acids on the surface. The folding funnel diagram is providing a model for the driving-forces of folding (**Figure 1**).

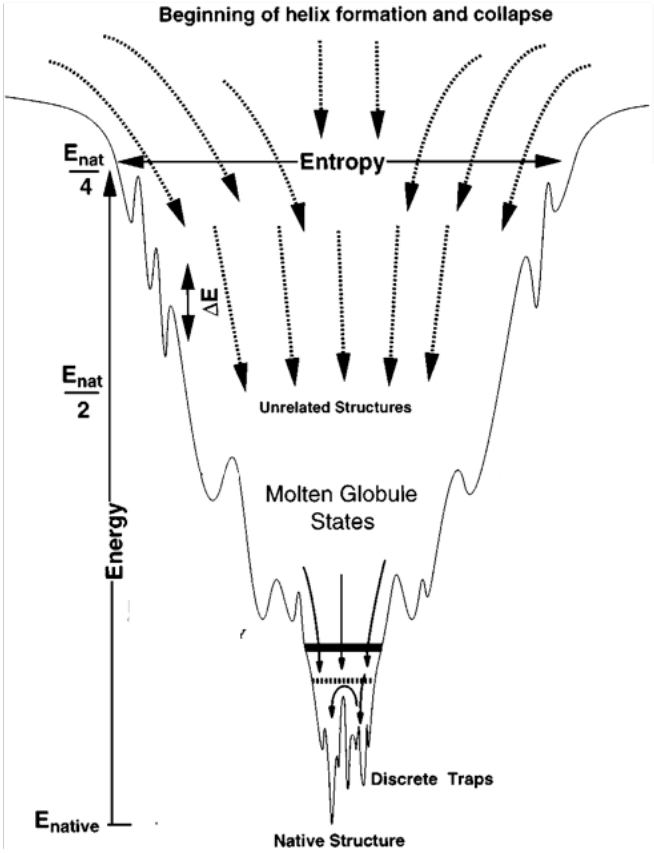


Figure 1: **Folding funnel diagram.** The free energy E for unfolded proteins is high as well as the number of possible conformations. The polypeptide reduces its free energy and the number of possible conformations by collapsing into a molten-globule. Rearrangements lead to the native conformation. Adapted from (Onuchic et al., 1997)

However, in the physiological environment of a living cell, this simple model of protein folding needs modifications. *In vivo,* proteins are normally synthesized in the cytoplasmic environment of a cell on ribosomes by translation of mRNA and do not fold independent of the translational process, since the codon translation rate is much slower than the folding rate (Fedorov and Baldwin, 1997). Furthermore, the co-translational folding of proteins is restricted by a 100 Å long ribosome exit tunnel that

can cover 30 – 40 amino acids, and thus prevents folding (Malkin and Rich, 1967). Although co-translational formation of secondary structures can already occur in the ribosome and even complete small domains, like zincfinger motifs, have been shown to form within the ribosome exit tunnel (Marino et al., 2016; Nilsson et al., 2015b), large domains *in statu nascendi* with only partial exposure of the domain sequence to the cytosol cannot fold into their native structure.

Furthermore, in living cells a problem emerges that is only of minor importance when characterizing folding in the test-tube. The high density (300-400 mg/ml biomolecules) in the cellular environment increases the possibility of a nascent chain polypeptide to interact with other proteins (Ellis and Minton, 2006). Although the collapse of a polypeptide chain in general happens very fast, the folding pathway to the native state can be energetically hindered, resulting in kinetically trapped intermediates, which expose hydrophobic amino acids to the cytosol that are prone to aggregation. To avoid this, a class of proteins have emerged that recognize these hydrophobic stretches, passively aid in folding and prevent aggregation: chaperones. Many chaperones have been identified as heat shock proteins (Hsps). The Hsps are upregulated upon heat shock to protect heat-destabilized proteins from aggregation (Kim et al., 2013a). The different classes of Hsps were named after their molecular weights: Hsp40s, Hsp60s, Hsp70, Hsp90s, and Hsp100s, with the Hsp70 chaperones being the best characterized. They function via an ATP/ADP-cycle (**Figure 2**).

The ATP-bound open conformation of Hsp70 chaperones is screening substrates for hydrophobic stretches. Hsp40 co-chaperones often recruit unfolded proteins to Hsp70. The so-called J-domain of Hsp40 chaperones interact with the ATPase domain of Hsp70 and stimulates hydrolysis of ATP to ADP. The ADP-bound closed conformation of Hsp70 chaperones is binding to the unfolded protein. Upon ADP-ATP exchange, which is catalyzed by nucleotide-exchange factors, the polypeptide is released and a new cycle can start (Kim et al., 2013b).

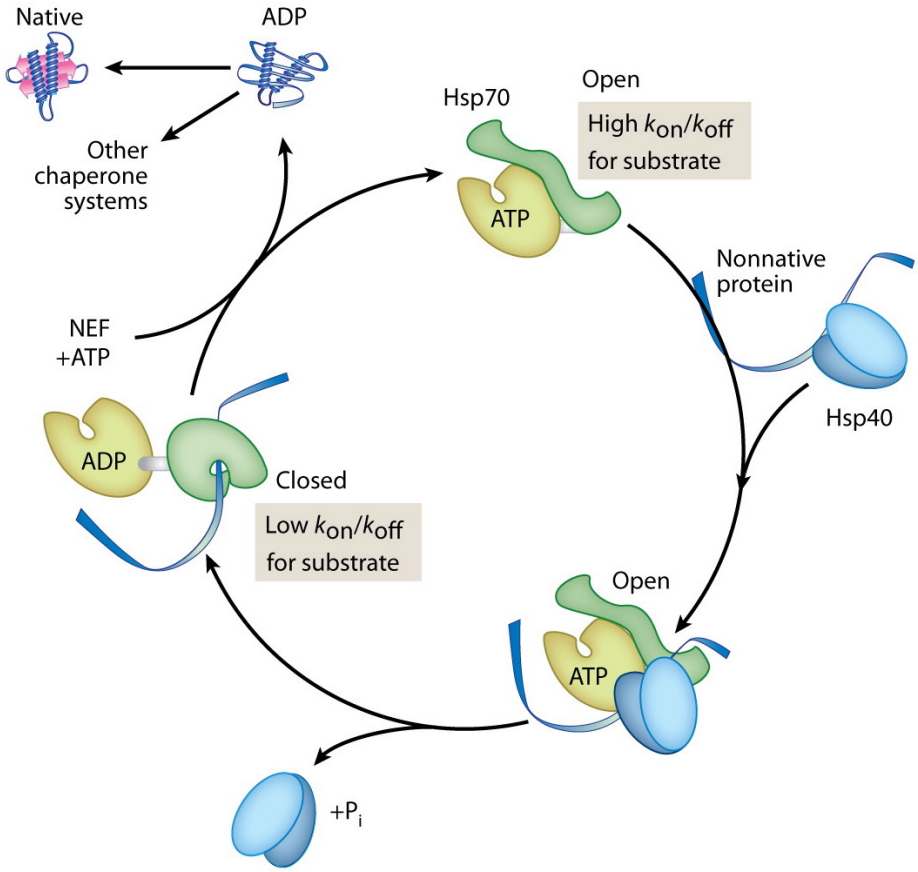


Figure 2: ATP-dependent Hsp70 chaperone cycle. The ATP-bound Hsp70 chaperone is screening polypeptides for unfolded stretches with a high on/off rate. Hsp40 co-chaperones are often recruiting unfolded polypeptides to the Hsp70 chaperone and are mediating the hydrolysis of ATP to ADP, which leads to an closed conformation, where the unfolded protein stretch is hidden inside the chaperone, making it unable to aggregate with other proteins. Nucleotide-exchange factors reopen the chaperone and the substrate is released. (Kim et al., 2013a).

Importantly, binding of proteins to an unfolded polypeptide itself does not facilitate efficient folding, but is limiting the concentration of unfolded intermediates, thus decreasing the possibility of protein aggregation. Chaperones do so by assisting in protein folding in basically every step of protein synthesis, starting already at the ribosome. Ribosome-associated Hsp70-chaperone (RAC in yeast, Hsp70L1 in mammals) and nascent-chain associated chaperone (NAC in archaea and eukaryotes) aid the polypeptide during the earliest stages of protein synthesis, when the nascent chain emerges from the ribosomal exit tunnel and is exposed to the cytosolic environment (Preissler and Deuerling, 2012). The exact mechanism is still elusive and the function of many chaperones might be redundant. It was shown that both, ribosome-associated chaperones (RAC) and soluble chaperones bind to nascent chains, keeping them in a partially unfolded conformation to prevent misfolding and aggregation until larger parts of the protein have been synthesized, ensuring cooperative domain folding

(Kim et al., 2013a). Upon protein synthesis into the cytosol, the nascent chain is further chaperoned by other cytosolic Hsp70-Hsp40 chaperones, like Ssa1-4 (in yeast) (Bukau et al., 2000). Some proteins are transferred to so-called chaperonins post-translationally. Chaperonins are large complexes (800 – 1000 kDa) with a central cavity, that allow the polypeptide to fold in a protected environment (Bukau and Horwich, 1998).
Notably, Hsps do not only prevent misfolding and aggregation of proteins until they are correctly folded in the cytosol. When proteins fail to fold they are degraded via the ubiquitin-proteasome system (UPS), a process that also requires the binding of chaperones to the polypeptide (Chakrabarti et al., 2011).

Proteins in a living organism have to fulfill many inter- and intracellular functions. Several membrane-enclosed organelles (e.g. endoplasmic reticulum, mitochondria, peroxisomes, nucleus) provide specific environments, where specialized enzymes are responsible for specific maturation steps. Due to the compartmentalization of the eukaryotic cells, most proteins are not resident in the cytosol and therefore have to pass at least one intracellular membrane barrier (Schatz and Dobberstein, 1996). However, the transport of folded proteins into intracellular compartments is rarely observed. Instead, the folding process is often uncoupled from protein transport and is delayed until the protein reaches a non-cytosolic compartment. This is achieved via two different translation and translocation mechanisms:

- Co-translational translocation of nascent proteins, coupling the translation process to the transport of the polypeptide inside the endoplasmic reticulum
- Post-translational translocation of completed proteins that are kept in a folding-competent state by chaperones

Transport into most organelles occurs post-translationally, with the help of chaperones, that keep the nascent chain in an unfolded but folding-competent state. An exception is the transport of polypeptides into the endoplasmic reticulum, where translocation mostly occurs co-translationally and only a fraction of proteins, mainly in bacteria and yeast, is trafficked via post-translational pathways. In mammalian cells, only small polypeptides are targeted post-translationally to the ER (Lakkaraju et al., 2012) and the co-translational pathway is the major route of protein transport into the ER and beyond.

1.2. Co-translational SRP-dependent targeting into the ER

Co-translational SRP-dependent targeting into the endoplasmic reticulum is one way to transport proteins across the ER membrane. The protein is targeted and translocated *in statu nascendi* via the pore-forming membrane protein Sec61 to the luminal site of the endoplasmic reticulum, thus most of the folding process only occurs after the protein has reached the ER lumen. To ensure specific targeting, proteins contain molecular labels, mostly on an amino acid level, which direct them to the corresponding compartments. One of the most common labels is the N-terminal signal sequence or signal peptide for targeting to the endoplasmic reticulum. The signal sequence is not necessarily preserved after targeting. Especially for soluble proteins it is cleaved off by the signal peptidase. Although the sequence homology between different signal peptides is very low, they share several loosely defined properties (Nilsson et al., 2015a) (Heijne, 1990).

- The length of a signal peptide is between 16 – 30 amino acids
- The N-terminal region (n-region, 1 – 5 amino acids) is rich in lysines and arginines, and therefore positively charged.
- The middle part (h-region, 7 – 15 amino acids) is the hydrophobic core of the sequence. The hydrophobic amino acids are crucial for efficient targeting and translocation.
- The C-terminal region (c-region, 3 – 7 amino acids) has mostly polar amino acids and is important for signal cleavage. The (-3,-1)-rule states that amino acids in position one and three upstream of the cleavage site must be small, neutral residues. Importantly, the amino acid at position +1 must not be proline.

Signal sequences mediating co-translational targeting are recognized by the signal recognition particle (SRP). Even though the general function of SRP is highly conserved in all domains of life, its structural composition is different, with an increasing complexity from bacteria (Ffh) to yeast and mammals (Pool, 2005). Mammalian SRP is a ribonucleoprotein, consisting of a Y-shaped 7SL RNA and six protein subunits (SRP9, SRP14, SRP19, SRP54, SRP68 and SRP72). Yeast SRP consists of a 11S RNA, a SRP14 homodimer, a yeast specific SRP21, SRP68p, SRP54p and Sec65p (Pool, 2005). Both SRPs have in common that they can be divided into two

structural domains: The S-domain, which recognizes the signal peptide and the Alu-domain, which is thought to be responsible for translational arrest (Bacher et al., 1996; Siegel and Walter, 1986). A schematic representation of both SRPs is found in **Figure 3 A.** SRP54 plays an important role for co-translational targeting, since it is in close contact to the ribosomal exit site, enabling emerging signal peptides to bind to its methionine-rich M-domain (Keenan et al., 1998). The N-terminal part of SRP54 folds in a characteristic four-helix bundle, followed by a GTPase domain (NG-domain). M and NG-domain are connected via a highly flexible linker, allowing SRP to undergo major conformational changes during the targeting process (Nyathi et al., 2013). How SRP binds to the ribosome is shown in **Figure 3 B**. In mammalian cells, binding of SRP to the signal peptide leads to an SRP9/14-mediated elongation arrest, keeping the majority of the nascent chain in the chaperone-like environment of the ribosome tunnel, thus preventing the protein from premature folding and aggregation (Mason et al., 2000). In a next step, the complex is targeted to the endoplasmic reticulum by binding to the ER-bound heterodimeric SRP receptor (SR, FtsY in bacteria), consisting of a 70kDA SRα and a 30 kDA SRβ subunit. The C-terminal part of SRα shows high homology to the NG-domain of SRP54, including the GTPase functionality. The β subunit also has an Arf-like GTPase-domain (Miller et al., 1995), which facilitates strong SRα-β association in its GTP-bound state. Possibly, Sec61β acts as a nucleotide exchange factor, thus activating SRβ for binding with SRα (Helmers et al., 2003). Interaction between SRβ and the yeast Sec61 homologue Ssh1p has been verified, as well as targeting defects and growth defects upon disruption of this interaction, suggesting that targeting to the endoplasmic reticulum is dependent on the interplay of these two proteins ((Jiang et al., 2008), (Helmers et al., 2003)).

Interestingly, unlike classical GTPases, where activation and inactivation is triggered via binding and hydrolysis of GTP, the pairing of SRP and SR is regulated via a nucleotide-dependent dimerization cycle (Gasper et al., 2009). The mechanisms of dimerization and targeting were deeply analyzed in bacteria. Both, Ffh and FtsY, were found inactive, even in an open GTP-bound state, due to structural impairments. Only upon dimerization and targeting the GTPase activity is established over a cascade of conformational changes (Peluso et al., 2001). Importantly, the dimerization and activation process requires the presence of a ribosome-nascent chain complex (RNC) to be efficient, by on the one hand disfavoring the rearrangement of the GTPases to a

closed and activated state by forming a highly stabilized early targeting intermediate, and on the other hand by accelerating the formation of the RNC-SRP-SR complex 1000-fold (Peluso et al., 2000; Zhang et al., 2009). These stabilizing and accelerating effects are preventing abortive reactions of the targeting process. At the end of the targeting, the RNC is handed over to the translocon, resulting in a new cascade of conformational changes in the GTPases, leading to a so-called closed state. Intense biochemical and crystallization studies in bacteria demonstrated that amino acids in loops seven and nine of SecY are interacting with the ribosomal proteins L23 and L35, which are located at the ribosomal exit site. However, how the exact unloading mechanism of the RNC-SRP-FtsY(SR) complex to the translocon works is still under debate. Since the binding sites of the translocon to L23 and L35 are overlapping with the ribosomal binding sites for SRP, a strictly guided handover is possible, where SRP is stepwise exchanged for the translocon. A simplified representation of the targeting process is shown in **Figure 3 C**. The signal sequence has now engaged with the Sec61 translocon, anchoring the nascent chain in the ER membrane and translation continues to synthesize the polypeptide into the ER lumen.

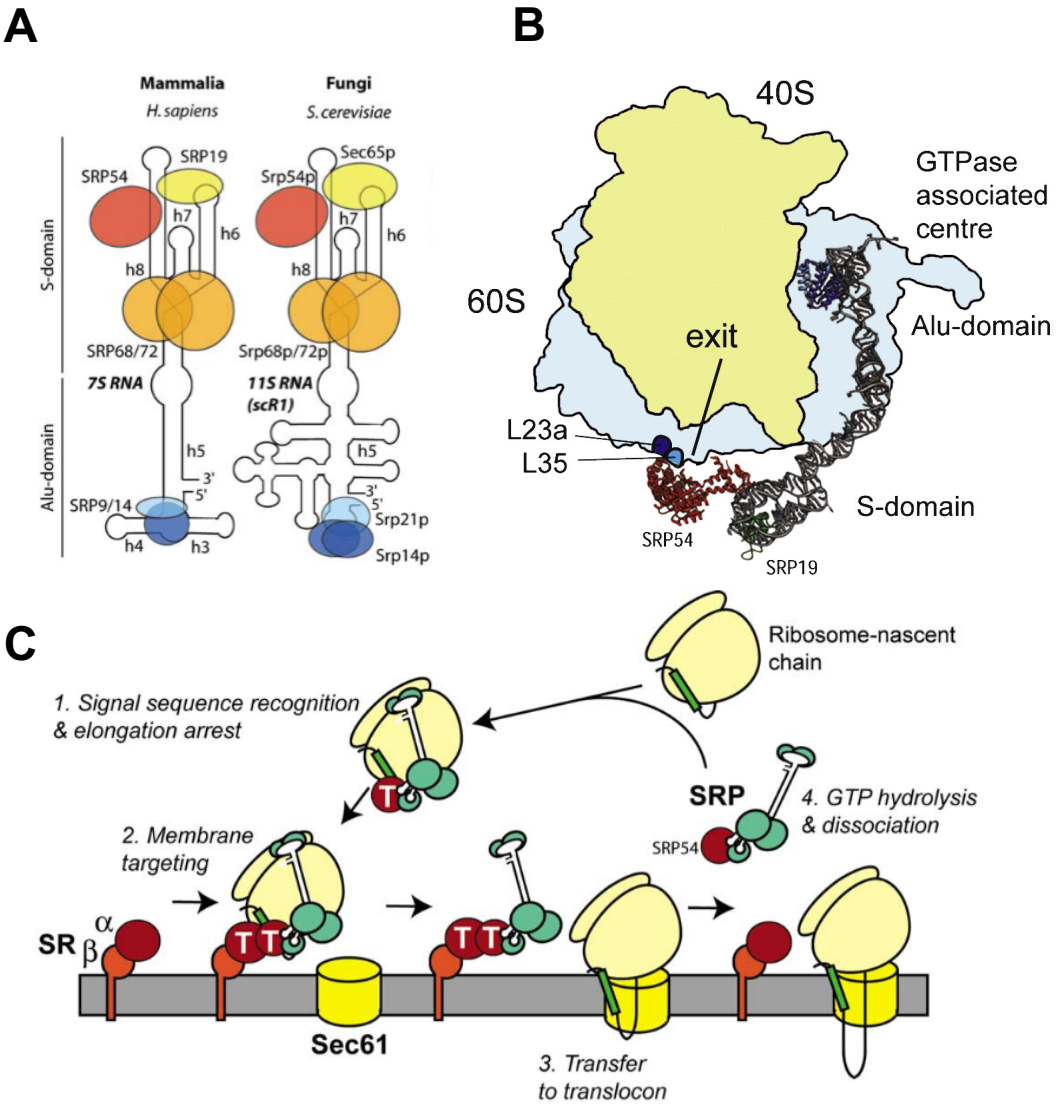


Figure 3: **A** Schematic representation of yeast and mammalian SRP. **B** Localisation of mammalian SRP on a translating ribosome. SRP54 is interacting with the ribosomal proteins L23 and L35, near the exit site, thus mainly interaction with the 60S subunit. The Alu-domain stretches to the interface between 60S and 40S, where the binding site of the elongation factor is located. **C** Co-translational targeting of the RNC complex. The signal sequence (green) is recognized by SRP, resulting in elongation arrest. Afterwards, the the complex is targeted to the SR, which is located on the ER membrane. How unloading of the RNC-SRP-SR complex to the translocon works is still under discussion. Upon unloading, hydrolysis of the active GTPases SR and SRP leads to dissociation of the proteins from the membrane (Nyathi et al., 2013; Pool, 2005).

1.3. Post-Translational Targeting and Translocation into the ER

For efficient post-translational sorting via the appropriate targeting route proteins have specific labels, either sequential or structural, which destine them for the correct targeting machinery. Post-translational targeting and translocation into the endoplasmic reticulum was well characterized in *Saccharomyces cerevisiae.* The hydrophobicity of post-translational signal sequences is reduced, compared to co-translational signal sequences. Thus, SRP most likely fails to capture the signal sequence, resulting in a protein that is fully synthesized into the cytosol (Ng et al., 1996; Zheng and Gierasch, 1996). To avoid aggregation in the wrong compartment, Hsp40/Hsp70 chaperones bind to the protein, keeping the protein in an unfolded, translocation-competent conformation (Ngosuwan et al., 2003). How post-translational substrates are targeted to the endoplasmic reticulum is not fully understood, however it is known that Sec61 and the subcomplex Sec62-Sec63-Sec71-Sec72 are required for efficient targeting (Deshaies et al., 1991; Panzner et al., 1995). The Hsp70-chaperone Ssa1 was proven to chaperone post-translational substrates, while the Hsp40-chaperone Ydj1 acts as a co-chaperone for Ssa, probably guiding it to the membrane (Becker et al., 1996). Lately, it was also shown that Ssa1 interacts with the ER-membrane associated Sec72, thus it is also likely that targeting to the translocon is mediated through Sec72 (Tripathi et al., 2017). Once the substrate has somehow engaged with the translocon and a part of the polypeptide reaches the ER lumen, the luminal chaperone Kar2p acts as a molecular ratchet to translocate the polypeptide into the ER lumen (Matlack et al., 1999). Kar2p is recruited to the ER lumen via the J-domain of the co-chaperone Sec63 (Misselwitz et al., 1999). While Kar2p is an ATPase, the nucleotide-exhange factors Sil1 and GRP170 ensure exchanging of ADP with ATP (Behnke et al., 2015; Tyson and Stirling, 2000). A simplified scheme of post-translational translocation is shown in **Figure 4** (Park and Rapoport, 2012).

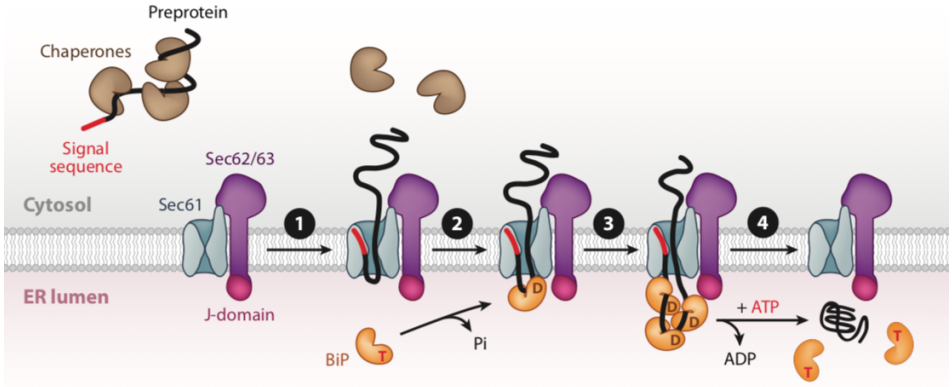


Figure 4: Post-translational translocation in eukaryotes. Cytosolic chaperones prevent aggregation of the protein and prime it to the ER membrane, where the signal sequence engages with Sec61, mediated by the Sec62-Sec63 complex (1). As soon as the polypeptide chain emerges from the luminal site of the translocon tunnel BiP (Kar2p) acts as a molecular ratchet to provide the vectorial force in order to translocate the polypeptide into the luminal side of the endoplasmic reticulum (2-4) (Park and Rapoport, 2012).

A fraction of membrane proteins is also integrated via a different mechanism that involves the proteins SND1, SND2 and SND3 (SRP-independent targeting). The mechanisms, regarding targeting and integration are not well understood. SND1 is believed to be cytosolic, ribosome-bound and thus may be responsible for the initial recognition of the signal sequence, which was found to be located more in the center of the synthesized proteins. SND2 and SND3 are membrane proteins which might associate with Sec62-Sec63, thus they might act as a receptor for cargo-carrying SND1. It was also shown that the SND-pathway not only exclusively directs SND-dependent proteins to the ER, but also acts as a rescue pathway, when proteins fail to be captured by SRP or the GET-pathway (Aviram et al., 2016).

Not all proteins have N-terminal signal sequences. Some proteins are also targeted to the endoplasmic reticulum by a C-terminal signal sequence, via the so-called GET-mediated tail-anchored protein-insertion pathway (Stefanovic and Hegde, 2007). This pathway involves a different targeting and translocation machinery that is acting independent from Sec61. The mechanism of the GET pathway in yeast is visualized in

Figure 5. Since the signal sequence for insertion via the GET pathway is C-terminal, targeting cannot occur co-translationally. Instead, the newly synthesized proteins are captured by the carboxy-terminal binding domain of Sgt2, which is part of a pre-targeting complex (Sgt2-Get4p-Get5p in yeast, Bag6-TRC35-Ubl4A in mammals) (Wang et al., 2010). In a subsequent step, the pretargeting complex loads its cargo onto Get3, a homodimeric ATPase, which transiently interacts with Get4 (Chartron et al., 2010). How transfer of the TA protein onto Get3 is facilitated is still under debate. It is proposed that the TA protein is transferred from Sgt2 to ADP-Get3, since the ADP-form of Get3 provides a hydrophobic grove for substrate binding (Mateja et al., 2009; Wereszczynski and McCammon, 2012). Interestingly, the TA binding domain (TABD) of Get3 shows similarity to the M-domain of SRP54, being rich in methionine and hydrophobic amino acids. The Get3-TA complex is further targeted to Get1-Get2, which are located on the endoplasmic reticulum membrane (Schuldiner et al., 2008). In Get2, long unstructured amino-terminal tails, on whose ends are short stretches of alpha-helices, capture Get3-TA, due to electrostatic interactions via a positively charged sequence in the Get2-tail and a negatively charged surface patch on Get3 (Mariappan et al., 2011; Stefer et al., 2011). After Get3 has been tethered to Get2, cytosolic coiled-coil domains of Get1 dock to Get3-TA, facilitating opening of the complex, and releasing of the TA protein (Kubota et al., 2012). How the TA protein inserts into the membrane is not well understood. The insertion step could either be spontaneous, or facilitated by Get1/2. An insertion mechanism that mimics the substrate tansport mechanism of ATP-binding cassette (ABC) transporters also was proposed (Denic, 2012).

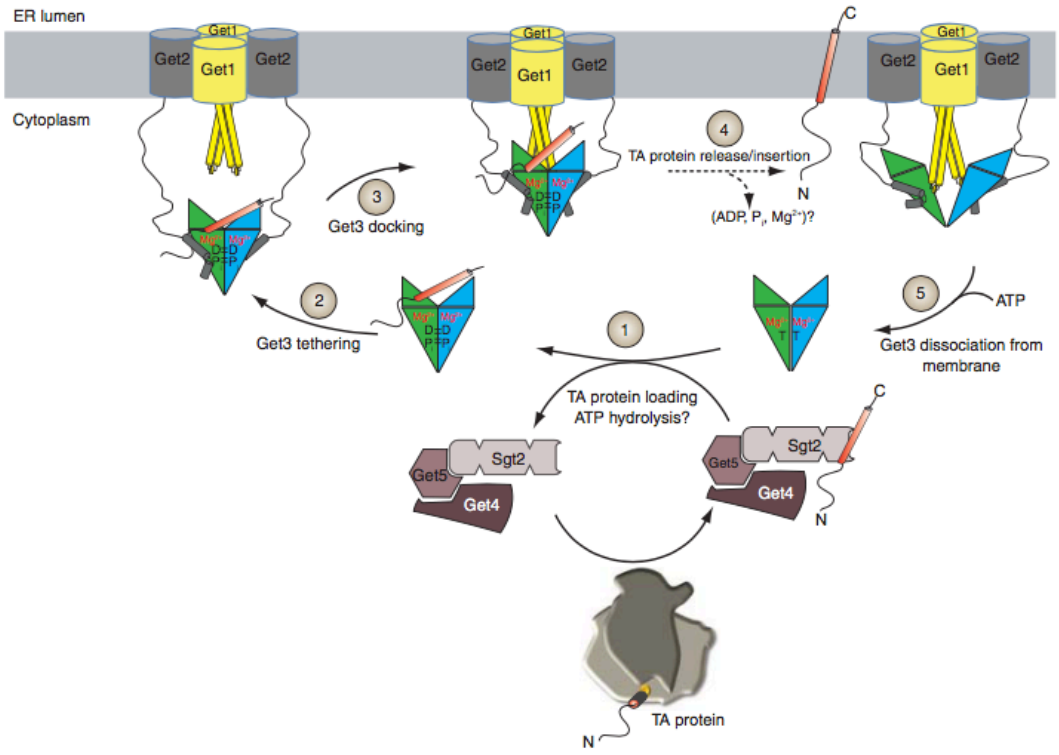


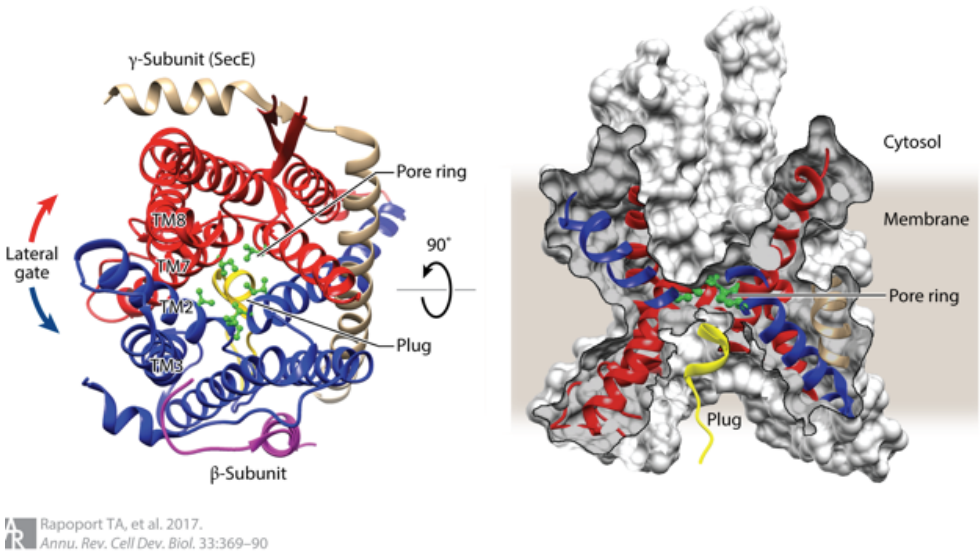
Figure 5: Illustration of the GET pathway in yeast. As soon as the signal of a TA protein emerges from the ribosome exit tunnel Sgt2 from the pretargeting complex binds to the TA domain. ATP-Get3 transiently associates with the loaded pretargeting complex, where the cargo is transferred to Get3 upon ATP hydrolysis (1). Loaded ADP-Get3 is targeted to the ER membrane via interacting with long N-terminal unstructured loops of Get2 (2). Docking to the membrane occurs via interacting with the coiled-coil domains of Get1 (3). This interacting leads to unloading of ADP-Get3 and the TA protein is somehow inserted into the membrane (4). After unloading Get3 dissociates from the membrane upon new binding of ATP (5) (Denic et al., 2013).

1.4. Structural insights into the Sec61 translocon machinery

Independent of the translational status of polypeptides, transport through and into the ER membrane is in most cases facilitated by the Sec61 translocon pore. The translocon is a multi-subunit membrane pore, whose overall structure is highly conserved within all domains of life (Park and Rapoport, 2012). The first structure of a translocon (from *Methanococcus jannaschii)* was published in 2004 (Van den Berg et al., 2004).

The core translocon machinery is a heterotrimeric complex, consisting of Sec61α, Sec61β and Sec61γ (Sec61p, Sbh1p and Sss1 in yeast, SecY, SecE and SecG in bacteria). Sec61 is a multi-spanning membrane protein, consisting of 10 transmembrane domains (TM domains) (**Figure 6**). The translocon can be visualized as a membrane protein with two bundles of five TM domains each (TM domain 1-5 and TM domain 6-10). The loop between TM domain 5 and 6 is acting as a hinge, connecting both halves. On the opposite side of the hinge, TM domain 2 and TM domain 7 are forming the lateral gate, which remains closed in the idle state. The idle translocon pore is obstructed by a so-called plug (also referred to as TM domain 2a). Mainly hydrophobic amino acids from TM domain 2,5,7 and 10 form a central constriction ring, which is narrowing the translocon channel, giving the pore an hourglass shape (Junne et al., 2010). The translocon is placed asymmetrically in the ER membrane, with the constriction ring located more on the cytosolic site of the membrane (Demirci et al., 2013). In eukaryotes, the nonessential Sec61β is a tail-anchored membrane protein, which only loosely contacts the α-subunit of Sec61. Sec61γ is a tail-anchored, clamp-like membrane protein. It has a long curved TM domain and is oriented diagonally in the membrane, thus contacting both halves of Sec61α (Rapoport et al., 2017). Although it has been shown that yeast Sec61γ is required for efficient translocation in yeast, the exact mechanism still remains elusive (Wilkinson et al., 2010).

Electron microscopy and crystallization studies revealed that upon binding of the ribosome on loop 8 the translocon undergoes a conformational change, unplugging the pore and opening the lateral gate to the ER bilayer (Raden et al., 2000). However, the exact timepoint of translocon opening is still under discussion.



Rapoport TA, et al. 2017.
Annu. Rev. Cell Dev. Biol. 33:369–90

Figure 6: Crystal strucute of the idle SecY translocon channel from Methanococcus jannaschii. Visualized on the left is a view onto the translocon from the cytosolic side. The right shows the translocon from the lateral side. The two halves of the translocon are shown in red and blue. The β-subunit is shown in purple, the γ-subunit is depicted in beige. The plug, which is sealing the pore, is shown in yellow (Rapoport et al., 2017).

In order to modulate efficient protein translocation, additional proteins form transient complexes with the translocon. The requirements, mechanisms and specificities of these complexes are organism-dependent and still under debate. Many auxiliary proteins lack detailed structural studies, thus their contact sites, interactions and conformational behavior upon translocation often can only be speculated upon.

The subcomplex **Sec62-Sec63-Sec71-Sec72** is of high importance. The essential Sec62 is a two TM domain membrane protein, contacting Sec61 near the lateral gate. Sec62 and the essential three-TM domain containing membrane protein Sec63 interact via their oppositely charged N- and C-terminal cytosolic tails (Harada et al., 2011). Even though earlier studies proposed that Sec62-Sec63 is especially required for post-translational translocation, recent findings demonstrated that a lack of Sec62-Sec63 results in translocation- and integration-deficiency also for co-translational translocation and for membrane protein topogenesis regulation (Jung et al., 2014). Sec63 can be characterized as a co-chaperone, recruiting **BiP** (Kar2p in yeast) to the translocon via its J-domain (Misselwitz et al., 1999). BiP is binding to the nascent chain upon ATP hydrolysis and facilitates translocation into the ER lumen by acting as a molecular ratchet and by preventing backsliding of the nascent chain (Matlack et al., 1999). Since BiP belongs to the ATPase family of chaperones nucleotide exchange factors are required to keep BiP in a functional state (Behnke et al., 2015). The non-

essential Sec71 and Sec72 have been proposed to be important for correct membrane protein insertion (Green et al., 1992). Structural studies revealed that the cytosolic chaperone Ssa1 and the ribosome-bound chaperone Ssb1 interact with Sec72, proposing a recruiting mechanism for post-translationally translocated nascent chains to the translocon via Sec72 (Tripathi et al., 2017).

The translocation process is often accompanied by several modification processes, like the cleavage of the signal sequence by the signal peptidase or the glycosylation of the polypeptide chain by oligosaccharyltransferase. For a detailed description of the signal peptidase see chapter 4.

The yeast **oligosaccharyltransferase** (**OST**) is a nine subunit complex (Ost1p, Ost2p, Ost3p, Ost4p, Ost5p, Ost6p, Stt3p, Swp1p and Wbp1p) that transfers dolichol-linked sugar $Glc_3Man_9GlcNAc_2$ on the Asparagine (Asn)-reside of the sequence Asn-X-Ser/Thr, with X being every amino acid except proline (Tai and Imperiali, 2001). The function of the subunits is not well understood, however, it was shown that Ost1p, Ost2p, Stt3p, Swp1p and Wbp1p are essential. The subunit SST3p contains the active site (Karaoglu et al., 1997).

TRAM (Translocating-chain associated protein) is a membrane protein with eight TM domains, which is involved into the translocation and integration of some proteins. It was shown that TRAM associates with the translocon and helps less hydrophobic TM domains to integrate into the membrane (Do et al., 1996; Görlich and Rapoport, 1993; Görlich et al., 1992).

TRAP (Translocon associated protein) is a hetero-tetrameric complex that binds directly to Sec61 and OST. It was shown that TRAP somehow accelerates the translocation of some substrates and it might also participate in the topogenesis of membrane proteins, with focus on moderating the positive-inside rule. However, the exact function of TRAP is still unknown (Sommer et al., 2013).

Recently, it was found that the membrane-bound metalloprotease **Ste24** acts as a safeguard of protein translocation, clearing clogged translocons by cleaving SRP-

independent membrane proteins which are stuck in the translocon due to misfolding on the cytosolic side. The exact mechanism is still unknown (Ast et al., 2016).

1.5. The ER quality control system

Correctly folded proteins are of high importance for the cell, since only functional proteins can fulfill their anticipated function. To ensure that proteins that have reached the endoplasmic reticulum are in a correctly folded conformation, the cell has developed a quality control mechanism, which tests the newly synthesized proteins on their proper folding and send proteins that fail to fold for degradation via the endoplasmic reticulum associated degradation (ERAD) pathway. To ensure proper folding, the endoplasmic reticulum contains various chaperones, like protein disulfide isomerases (PDI), which catalyze disulfide-bond formation, peptide prolyl isomerases, which catalyze the isomerization of peptidyl proline bonds between cis and trans, and several chaperones from the Hsp40, Hsp70, Hsp90 and Hsp100 family (Ellgaard and Helenius, 2003). Generally, Hsp70 and Hsp90 chaperones belong to the family of ATPases and bind to not-yet correctly folded proteins. Normally, hydrophobic segments of proteins are located in the core of a folded protein, however in misfolded proteins these hydrophobic sequences are exposed to the aqueous environment, which is recognized by chaperones.

The process of folding is further monitored by the calnexin/calreticulin-cycle via an Asn-linked glycosylation system. The OST complex transfers co-translocationally a $Glc_3Man_9GlcNAc_2$ branched glycan on Asparagin-residues of the consensus sequence Asn-X-Ser/Thr (X is not proline) (Kornfeld and Kornfeld, 1985). After glycosylation, the glycan structure is changed by several ER resident glycosidases. First, α-glucosidase I cleaves the terminal α-1,2-linked glucose, followed by cleavage of the now terminal α-1,3-linked glucose by α-glucosidase II (Trombetta et al., 1996). At this point, the calnexin/calreticulin cycle starts to chaperone the folding process. Calnexin and calreticulin are both chaperones of the lectin family, having calcium-ions bound in their globular domains near the carbohydrate binding site, which stabilizes the binding. The N-termini of both proteins together form a single carbohydrate binding-site which has high affinity to monoglucosylated glycans (Michalak et al., 2009; Schrag et al.,

2001). Both chaperones contain a central proline-rich domain (P-Domain) that transiently interacts with either ERp57 (a protein disulfide isomerase) or cyclophilin B (CyB, a peptidyl-prolyl isomerase), which in turn can assist in proper folding of the substrates (Kozlov et al., 2010; Oliver et al., 1997). If the substrate does not manage to fold correctly, α-glucosidase II will also cleave the last α-1,3-linked glucose, leaving an unglucosylated glycan on the substrate (Stigliano et al., 2011). In principle, the substrate can now follow three pathways: if the protein is folded correctly it will be released, its glycan further trimmed by ER mannosidases, and may be exported from the ER. If the protein is still misfolded, it is either re-monoglucosylated by UDP-glc:glycoprotein glucosyltransferase I (UGGT1), resulting in reinitiation of the calnexin/calreticulin-binding, or it can in addition undergo cleavages by mannosidases that direct the misfolded protein to the ERAD pathway. In general, a protein can undergo several calnexin/calreticulin cycles until it is either properly folded or sent to degradation (Quan et al., 2008; Taylor et al., 2004).

The ERAD pathway is crucial for cells homeostasis, sending misfolded proteins for degradation into the cytosol, where the proteins are finally degraded by the ubiquitin-proteasome system. The ERAD pathway through Hrd1 has been extensively studied in yeast and the results can also be extrapolated to mammals (Carvalho et al., 2006) The machinery involves a cascade of several proteins in different compartments. Since misfolded proteins can be very diverse, with unfolded domains either in the ER lumen, the ER membrane, or in the cytosol, several pathways emerged to ensure specific degradation: Substrates undergoing the ERAD-L pathway have misfolded domains in the ER lumen, while ERAD-M substrates have misfolded intermembrane domains. Both substrates are degraded via the Hrd1-E3-Ligase. ERAD-C substrate present un- or misfolded parts to the cytosol and are degraded via the Doa10-E3-ligase (Carvalho et al., 2006). It was also proposed that retrotranslocation can occur via the Sec61 channel, however direct involvement of Sec61 is still under debate (Römisch, 2017).
A schematic representation of the ER quality control system is shown in **Figure 7**. A misfolded protein is prone for degradation after mannosidases have trimmed the N-linked glycans, generating a terminal α-1,6-linked mannose residue which is recognized by the lectin Yos9p (Clerc et al., 2009; Xie et al., 2009). Probably Kar2p recruits the misfolded Yos9p-bound substrate to Hrd3p, a membrane protein that

additionally binds to unfolded regions of the substrate, activating it for degradation (Plemper et al., 1997). Hrd3p transfers the misfolded protein to the membrane-bound Hrd1 dimer, which is forming a pore and is facilitating retrotranslocation of the misfolded protein back into the cytosol, assisted by Der1. ERAD-M substrates might enter the Hrd1 pore laterally (Sato et al., 2009). Once the substrates enter the cytosol they are polyubiquitinated on lysine residues via Hrd1p, which recruits the ubiquitin-carrier Ubc7, an E2-konjugating enzyme. The retrotranslocated, polyubiquinated substrate is carried to the proteasome for degradation by cytosolic chaperones, like Bag6 (Ruggiano et al., 2014).

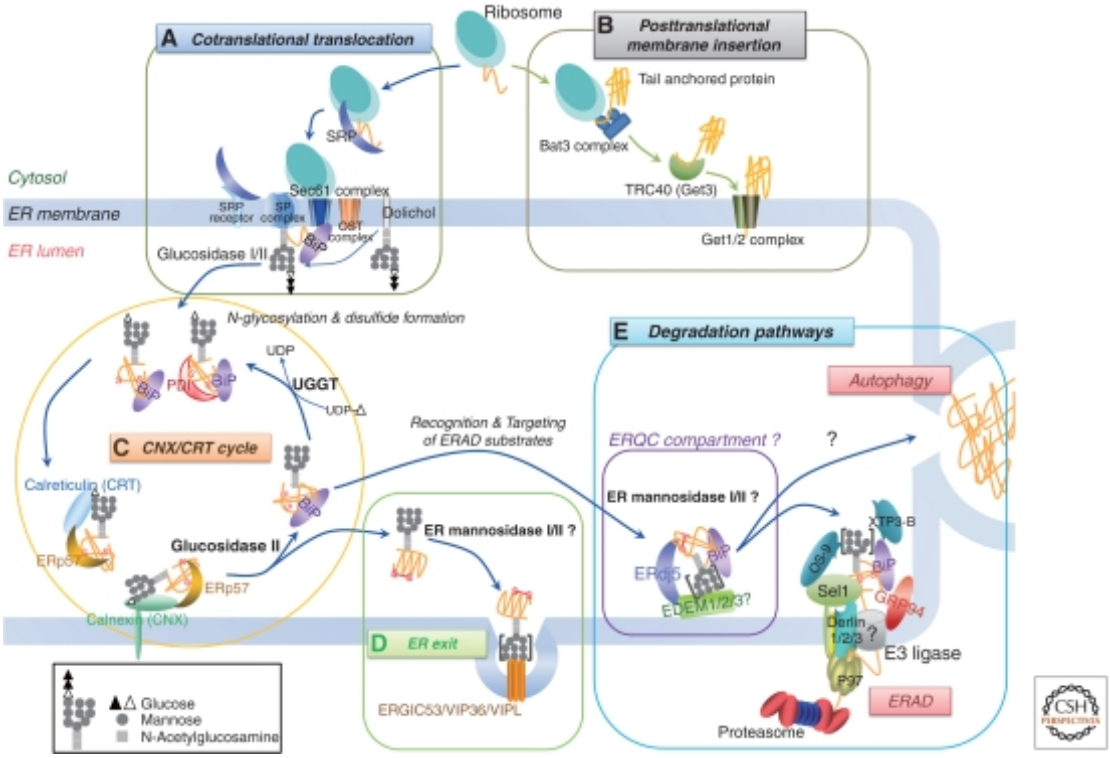


Figure 7: The ER quality control system (Araki and Nagata, 2011)**. A** Co-translational translocation of newly synthesized proteins into the endoplasmic reticulum. Upon entering of the nascent chain into the ER lumen the OST complex catalyzes the N-glycosylation of the polypeptide **B** Posttranslational translocation of tail-anchored proteins into the ER membrane via the GET pathway. **C** The calnexin/calreticulin cycle. The calnexin/calreticulin cycle recognizes N-glycans and test them on proper folding. **D** ER exit for correctly folded proteins. **E** The ERAD pathway for misfolded proteins. If a protein does not manage to fold during the calnexin/calreticulin cycle the N-glycans are trimmed by mannosidases, followed by retrotranslocation into the cytosol via an E3-ligase complex. The protein is then degraded via the ubiquitin proteasome pathway.

If the load of misfolded proteins in the endoplasmic reticulum starts to increase and reaches a certain threshold level, the cell starts to react by the well conserved unfolded protein response (UPR). The UPR tries to meet the ER stress by upregulating the ER-resident chaperone synthesis and at the same time attenuate translation. If the UPR does

not manage to lower the level of misfolded proteins it aims towards apoptosis (Shore et al., 2011). In general, three ER resident membrane proteins are ER stress sensors and signal to the cytosol or to the nucleus for countermeasures: PERK, Ire1 and ATF6. Among the many functions of BiP one is also the blocking of the three ER stress sensors, by binding to these receptors. Only when the threshold of misfolded proteins reaches a certain level, BiP disassembles from the ER stress sensors to bind to hydrophobic patches of the misfolded proteins. This unbinding activates the sensors. Upon unbinding of BiP, PERK homodimerizes, followed by autophosphorylation of its kinase domain, which in turn induces phosphorylation and thus inactivation of the elF2α, a transcription factor. This leads to reduced translation of new proteins, thus helping the ER to clear its lumen from misfolded proteins by ERAD (Kebache et al., 2004; Raven et al., 2008).

The IRE1 pathway shows similarity to PERK. Dimerization of IRE1 happens after unbinding of BiP, followed by autophosphorylation. This phosphorylation activates the endoribonuclease activity of IRE1, which cleaves the XBP1-mRNA, resulting in a frameshift variant of XBP1 that acts as a transcriptional factor to modulate gene expression of chaperone and ERAD genes (Hetz and Glimcher, 2009).

ATF6 exits the endoplasmic reticulum after BiP dissociates from its N-terminus by vesicular transport to the Golgi apparatus. Once at the Golgi apparatus, ATF6 is cleaved two times by the intramembrane proteases S1P and S2P. The soluble cytosolic ATF6-domain is trafficked into the nucleus, where it increases transcription of genes, encoding for chaperones, like BiP or PDIs (Chen et al., 2002; Shen et al., 2002; Yoshida et al., 2001).

1.6. The biogenesis of membrane proteins at the Sec61 translocon

The biological membranes that separate different compartments in a cell mainly consist of phospholipids, sphingolipids, sterols and membrane proteins (Harayama and Riezman, 2018). One of the major characteristics of the lipid bilayer membrane is the hydrophobicity of the acyl chains, which are oriented towards the membrane core, while the polar head group is exposed to the hydrophilic environment of the cell. The thickness of the hydrophobic core of a membrane generally is around 30 Å (Watson, 2015). While an organelle provides a separate environment with specialized functions, membrane proteins are crucial for the inter- and intracellular communication between these different enclosed cell compartments. About 30% of all proteins are membrane proteins (Krogh et al., 2001). Generally, two major types of membrane proteins can be discriminated: α-helical membrane bundle proteins and β-barrel membrane proteins. The latter are only found in the outer membrane of mitochondria or chloroplasts. The topologies of mature helix-bundle proteins can be quite divers, not only when it comes to stoichiometry, but also in number and orientation of the TM domains. Their cytosolic and luminal loops can bear several interaction motifs, which are often required for intracellular communication. Despite this heterogeneity, membrane proteins can basically be subdivided into only three classes ((Higy et al., 2004), **Figure 8 A**). This simple model only considers the orientation of the first TM domain. In secreted proteins the signal sequence is cleaved of by the signal peptidase (a). In single-spanning membrane proteins the signal sequence is either a part of the mature protein, or it is also cleaved of and the protein is anchored in the membrane by an internal TM domain (b, Type I). Proteins with non-cleaved signal sequences can attain two topologies: while a normal signal-anchor results in a protein with N(in)-C(out) topology (c, Type II), a reversed signal-anchor leads to a protein with N(out)-C(in)-topology (d, Type III). Multispanning membrane proteins basically can be described by combination of the abovementioned 4 elements (e-g). A signal-anchor domain in N(in)-C(out) orientation is followed by a stop-transfer domain (N(out)-C(in)), which is succeeded by a re-integration domain (N(in)-C(out)) (**Figure 8 B**).

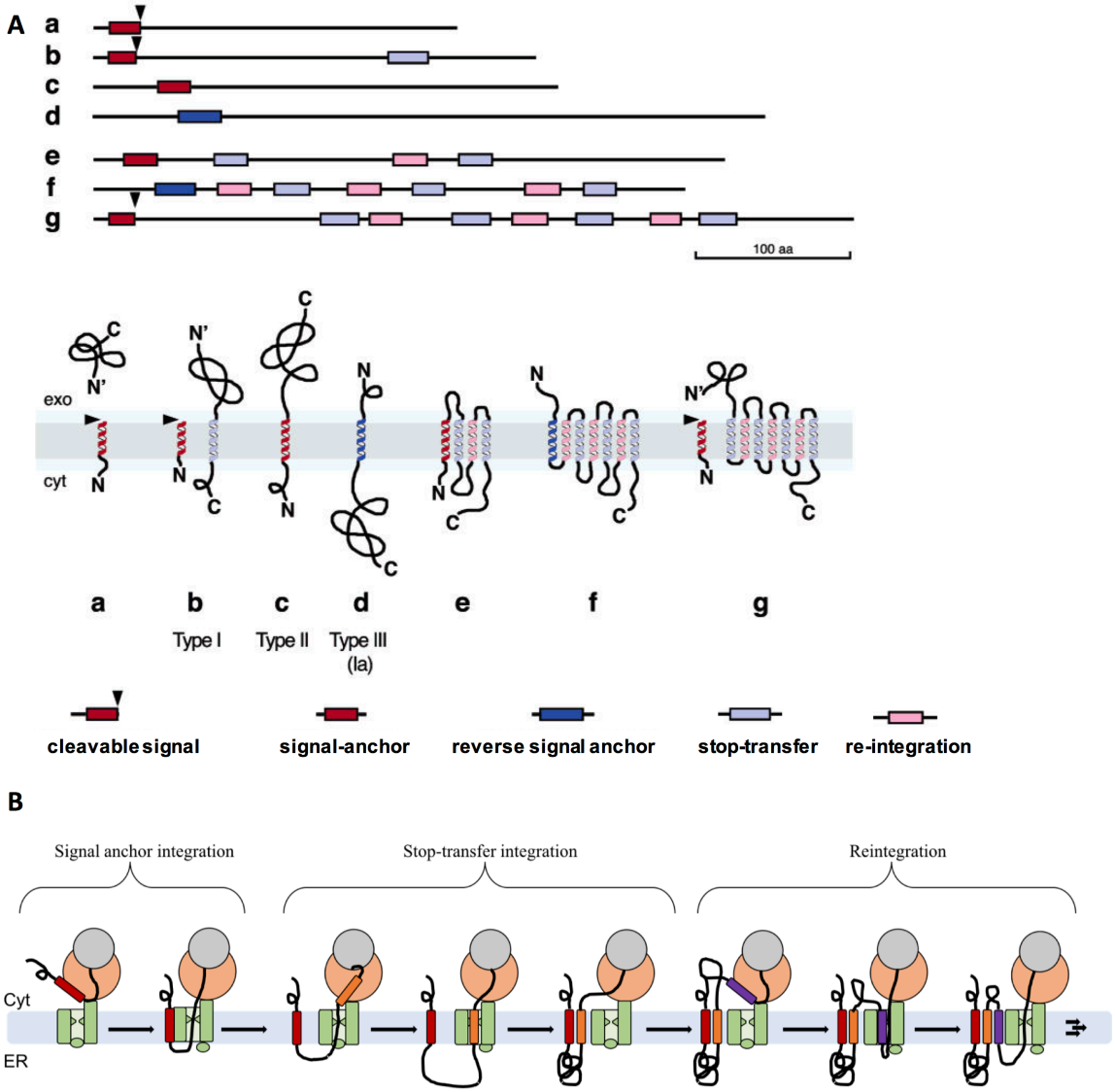


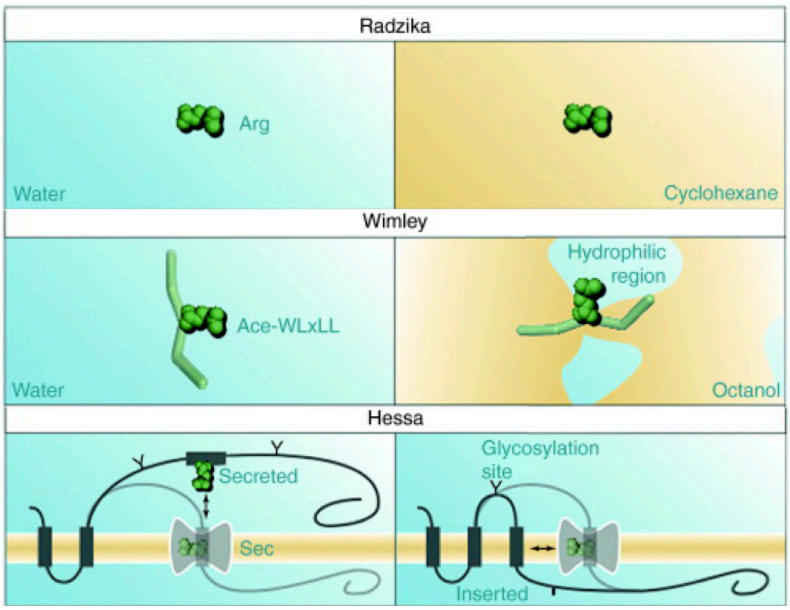
Figure 8: A Membrane proteins can be classified into several classes. a Cleavable signals (red rectangle with black arrowhead) results in a soluble protein. b Cleavable signal with an additional transmembrane domain results in a Type I membrane protein in N(out)-C(in)-orientation. c Uncleavable signal-anchor, resulting in a Type II membrane protein in N(in)-C(out)-orientation. d Reversed signal-anchor (blue rectangle), resulting in a Type III membrane protein in N(out)-C(in)-orientation. e – f Examples for multispanning membrane protein topologies. Adapted from (Higy et al., 2004). **B** The three integration steps for TM domains. The red TM domain is a signal-anchor domain. The stop-transfer domain (orange) moves through the translocon pore and exits the translocon through the lateral gate, if it overcomes the required hydrophobicity threshold. The re-integration domain (purple) is located on the cytosolic site. How it enters the translocon pore is still under debate.

In contrast, a reversed signal-anchor (N(in)-C(out)) is followed by a re-integration domain, thus, the third TM domain is a stop-transfer domain. The overall understanding of the mechanisms leading to efficient integration is based mainly on the concept of sequence autonomous, hydrophobicity-dependent integration: Independent of its sequence context, a domain integrates into the ER membrane if it overcomes a

certain hydrophobicity threshold, otherwise it will either be translocated into the ER lumen (for ST-TM domains) or it will remain in the cytosol (for RI-TM domains). Thus, hydrophobicity was soon known to be the major driving force for membrane protein integration. Several studies engaged the question how much hydrophobicity is required for a domain to efficiently integrate into a membrane. To do so, hydrophobicity scales were created that determine the contribution of each amino acid to the overall hydrophobicity of a domain (MacCallum and Tieleman, 2011).

A first and simple hydrophobicity scale was proposed by Radzicka and Wolfenden. This scale is based mainly on the partitioning of single amino acids between an aqueous phase and cyclohexane (Wolfenden, 2007). The hydrophobicity scale developed by Wimley and White is based on a similar approach, however using a pentapeptide (Ace-WLXLL, X=each amino acid) and measure the equilibration of each of these pentapeptides between water and 1-octanol (Wimley and White, 1996).
(Hessa et al., 2005) established a 'biological hydrophobicity scale', where the contribution of each amino acid to the integration threshold of a domain into the lipid bilayer was determined. To do so, a model protein derived from leader peptidase (Lep) was created, containing a mildly hydrophobic segment (H-segment) in stop-transfer orientation that consists of 19 alanine host residues. The H-segments hydrophobicity was modulated by replacing individual alanines with one of the other 19 amino acids. Insertion of the domain into the lipid bilayer of dog pancreas-derived rough microsomes was determined from the glycosylation pattern. The apparent free energy for each amino acid was determined similarly to the free energy calculations of the other hydrophobicity scales. A comparison of the different hydrophobicity scales is shown in **Figure 9**.

A



B

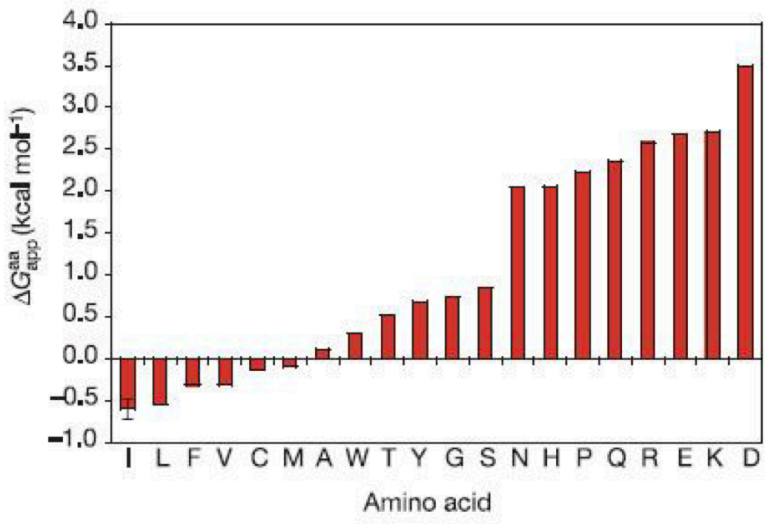


Figure 9: A Schematic representation of the equilibration principles behind the hydrophobicity scales from Razdika, Wimley and Hessa. Adapted from (MacCallum and Tieleman, 2011) **B** Biological hydrophobicity scale from Hessa et al. (2005). The one letter code for each amino acid is shown together with the ΔG_{app} values.

Besides the hydrophobicity of a domain several other properties of a polypeptide were identified that influence topogenesis. Back in 1986 Gunnar von Heijne described a phenomenon called “positive-inside rule” by observing a charge imbalance in signal sequences of proteins in the bacterial inner membrane (Heijne, 1986). This rule was later further verified in eukaryotes (Hartmann et al., 1989). It states that the orientation of a signal sequence in the membrane is dependent on the difference of flanking charges: positively-charged regions are oriented to the cytosol, while negatively charged regions are oriented to the outside (in bacteria) or to the ER lumen (in eukaryotes). These findings provided a first tool to predict the topology of membrane proteins and the linear-insertion model arose from these results: the first TM domain orients itself in accordance with the positive-inside rule, thus fixing the orientation of all following TM domains. It was shown that the topology of Type III membrane proteins can be inverted by exchanging the flanking charges (Beltzer et al., 1991). The positive-inside rule applies to some extend also to multispanning membrane proteins (Heijne, 2006). The positively charged amino acids lysine and arginine are more commonly found on the cytosolic side, while on the extracellular site their occurrence is significantly reduced.

This model was soon shown to be insufficient to explain the topology of some membrane proteins. It was revealed that folding of the N-terminal part of the sequence can occur co-translationally when exposed to the cytosol, thus preventing insertion of the folded N-terminal domain into the translocon, resulting in a N(in)-C(out)-topology, regardless of the charge distribution (Denzer et al., 1995). Futhermore, N-terminal signal sequences enter the translocon pore in a head-first orientation and invert co-translationally (Devaraneni et al., 2011; Goder and Spiess, 2003). Another finding concerning the orientation of the signal sequence was that length and hydrophobicity of the h-region is affecting the topology: a long hydrophobic h-region favor translocation of the N-terminus, while a short h-region translocates the C-terminus into the ER lumen (Wahlberg and Spiess, 1997).

Interestingly, signal-anchors do not necessarily have to be located near the N-terminus, but can also be localized in the internal part of a protein sequence. It was demonstrated that long N-terminal hydrophilic sequences do not affect protein targeting to the translocon and promote translocation of the C-terminus independent of the signal-

anchor's hydrophobicity (Kocik et al., 2012). Thus, for internal signal-anchors the signal sequence does not enter head-first, but orient themselves before pore-insertion.

While for single-spanning membrane proteins the final topology is already not easy to predict, due to the abovementioned intrinsic properties, the complexity increases even more for multispanning membrane proteins, where at least one additional transmembrane domain is complicating the translocation and integration mechanism.
The hypothesis that hydrophobicity of a potential transmembrane domain is the main driving force for membrane integration has been further challenged by examining the amino acid composition of several natural transmembrane domains. Especially TM domains from ion channels and aqueous pores were identified to have a low overall hydrophobicity, thus prediction programs often do not recognize these TM domains (De Marothy and Elofsson, 2015). To efficiently integrate charged TM domains into the membrane the charged residues of arginine and lysine can `snorkel` into the hydrophilic environment of the lipid's polar head groups, thus minimizing the energetic cost of integrating a charged residue into the lipid bilayer (Li et al., 2013). Another possibility to overcome the hydrophobicity threshold required for efficient integration is cooperative membrane integration. In this model mildly hydrophobic TM domains do not integrate alone, but together with neighboring TM domains (Hedin et al., 2010). Lately, it was shown by that the integration of a potential stop-transfer TM domain is in general not only dependent on the hydrophobicity of the domain, but also on a sequence up to 100 amino acids downstream of the TM segment (Junne and Spiess, 2017). These findings suggest that topogenesis is a far more complex process, which is influenced not only by hydrophobicity, but also by intrinsic determinants in the neighboring sequences of a potential TM domain.

Part I: Biogenesis of multispanning membrane proteins: membrane integration of N(in)-C(out) transmembrane domains at the Sec61-translocon in *Saccharomyces cerevisiae*

2. Aim

The topogenesis of membrane proteins in the lipid bilayer of the ER is of high complexity. The definition of a biological hydrophobicity scale suggested a rather simple mechanism based on sequence-autonomous, mainly hydrophobicity-dependent equilibration between the membrane and aqueous compartments. Previous studies (in particular Junne and Spiess, 2017) already indicated that sequences outside of potential transmembrane domains may affect transmembrane domain integration.

The focus of this thesis was to dissect the requirements and mechanisms that drive re-integration. To address this questions we mainly investigated the hydrophobicity threshold of a potential transmembrane domain in N(in)-C(out)-orientation for integration into the ER membrane. We discovered that this threshold critically depends on the properties of the preceding cytosolic polypeptide sequence. By changing this loop sequence, we investigated the features facilitating or inhibiting transmembrane re-integration, including loop folding and chaperone binding.

3. Results

3.1. The hydrophobicity threshold for re-integration is modulated by the preceding cytoplasmic loop length.

For translocation and integration studies, we have used a model protein derived from yeast dipeptidyl-aminopeptidase B (DPAPB or DAP2), a SRP-dependent type II membrane protein. The integration efficiency of generic stop-transfer domains was analyzed by replacing residues 170-378 with a mildly hydrophobic domain (H-segment), flanked by GGPG/GPGG hydrophilic helix-breaking insulators. The host sequence of the H-segment consists of 19 alanines. To modulate hydrophobicity, leucine guest residues of variable numbers (0-19) were replacing host alanines (for a list of all the H-segments used in this work see Appendix B). Since a major part of the native protein was deleted and additional residues were added to the protein, the resulting construct is unable to obtain a native fold (which according to the prevailing model of transmembrane integration is no concern). The construction pathway from the native DPAPB to our model protein is shown in **Figure 10 A**. To discriminate between the translocated (N) and integrated (I) form, glycosylation sites present in the protein made it possible to dissect the topology by analyzing the glycosylation pattern (**Figure 10 B**). Due to the non-native nature of the model protein, degradation via ERAD might influence the outcome of the experiment when looking at steady-state by western blot. To avoid this problem radiolabeling was used to examine the non-equilibrium ratio of the two topologies immediately after synthesis. The hydrophobicity threshold was determined after short pulse labeling with [^{35}S]methionine for 5 min, followed by immunoprecipitation against a C-terminal HA-Tag that is present in all model proteins, gel electrophoresis, and autoradiography. *The hydrophobicity threshold is by definition the amount of leucines that are required in an H-segment for more than 50% membrane integration.* For ST-DP this value was found to correspond to ~4 leucines in the H-segment by Junne et al. (2010) and Demirci et al. (2013) (**Figure 10 C and D, left side**). This is generally interpreted as the result of thermodynamic equilibration of the stop-transfer H-segment, between translocon pore and the lipid membrane after the segment has entered the translocon.

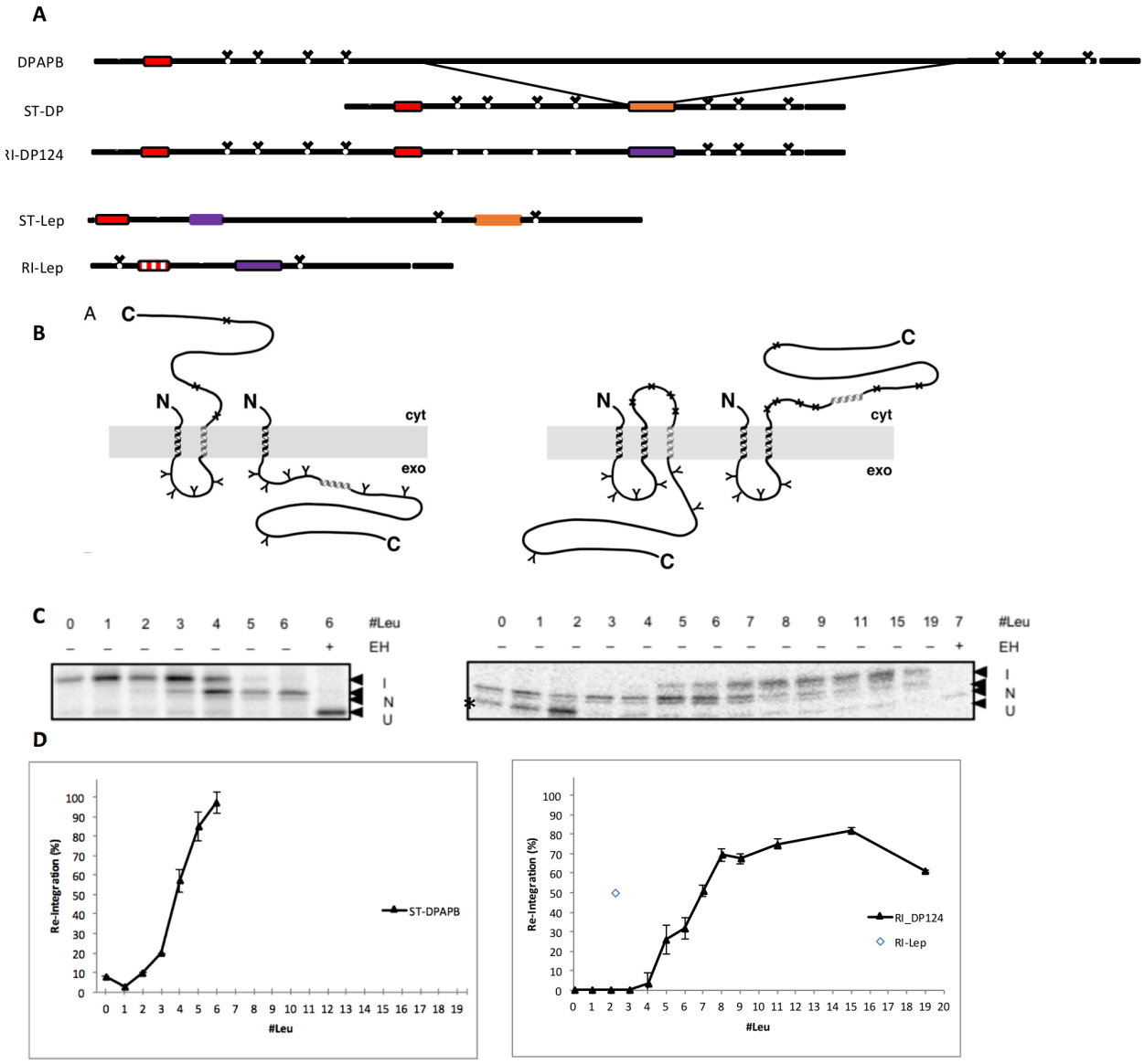


Figure 10: **A** The cloning pathway for ST-DP and RI-DP. The signal-anchor domain is marked in red, ST-domains are orange, re-integration domains violet. The Lep-constructs from Lundin et al. are shown to visualize the difference between the two model proteins. The reverse signal-anchor of RI-Lep is visualized with a red/white pattern. Glycosylation sites are indicated with an Y. **B** Schematic representation of the topologies for ST-DP and RI-DP. As indicated with a Y, fully integrated ST-DP has four out of seven glycosylation sites glycosylated. On the contrary, for RI-DP the fully integrated form has seven glycosylated sites in the lumen, while partial glycosylation only occurs for the partially inserted RI-DP **C** Autoradiographs for ST-DP (Demirci et al., 2013)and RI-DP. Wild-Type yeast cell are producing the model proteins with 0 – 6 and 0 – 9, 11, 15, 19 leucines in the H-segment. The different glycosylation patterns lead to a difference in molecular weight, which is visualized after metabolic labeling with [^{35}S]methionine, immunoprecipitation and SDS-gel electrophoresis. The band related to an integrated H-segment is marked in both gels with an I (Integrated). N stands for non-integrated for the re-integration constructs. Concerning the ST-DP, T stands for translocated. U marks the unglycosilated band in both model proteins. The asterisk indicates bands of a truncated version of the protein, produced due to internal initiation (see Appendix A) **D** Quantitation of the experiments are shown in C. For the re-integration graph the value determined by Lundin et. al (2008) is shown as a rectangle

.

In contrast, the mechanism of re-integration is more complicated. The re-integration domain is synthesized into the cytosolic vestibule between the ribosome and the translocon, where it needs to engage with the translocon to initiate membrane integration and C-terminal translocation similar to a signal-anchor sequence, or to follow the preceding loop into the cytosol (see **Figure 8 B)**.
Lundin et al. (2008) were the first to analyze this re-integration step using a Leader peptidase (Lep)-derived model protein in yeast and found the molecular code for these N(in)-C(out) transmembrane sequences to parallel that for stop-transfer integration (Lundin et al., 2008). Surprisingly, the integration threshold was shifted to a lower hydrophobicity of only ~2 leucines for re-integration, compared to ~3 leucines for stop-transfer with their Lep-derived model protein (Hessa et al., 2005).

To determine the hydrophobicity threshold required for efficient re-integration in our system, a RI-DP model protein was created by cloning the DPAPB codons 1 – 146 in front of ST-DP, resulting in a protein with an H-segment for potential integration in an N(in)-C(out) topology. The integrated topology of the H-segment (I) thus has seven sites glycosylated, while for the non-integrated, cytosolic topology (N) glycosylation only occurs on the four sites in the loop connecting the SA domain with the ST domain. The RI-DP124 constructs were created with 0 – 9, 11,15, and 19 leucines in the H-segment. The integrated band appeared with 5 leucines, becoming more prominent the more leucines were present in the H-segment. Interestingly, the re-integration efficiency reached a maximum at ~15 leucines with ~80% re-integration and decreased to ~60% for 19 leucines. The reduction of re-integration efficiency for strong hydrophobic H-segments could be seen in several construct series in this study. Upon digestion with endoglycosidase H (EndoH) the bands representing the glycosylated forms collapsed into the band representing the unglycosylated protein. About 7 leucines were required for 50% re-integration, in contrast to only 4 leucines for stop-transfer integration. Notably, the bands in the re-integration constructs appeared as double bands. This indicates that glycosylation on one of the glycosylation sites in the front part of our model protein is not 100% efficient. However, this did not affect the general interpretation of the data, since the integration (I) band and non-integration (N) bands in the autoradiographs were still distinguishable. The lower band in **Figure 10 C**, which is indicated by an asterisk is not an unglycosylated form of our model protein,

but represents a truncated form of our model protein which was expressed due to internal initiation of a methionine start codon (for details, see Appendix A).
The discrepancy of the re-integration efficiencies of 2 leucines for RI-Lep vs. 7 leucines for RI-DP124 led us to the conclusion that re-integration efficiency is not only dependent on hydrophobicity, but may be influenced by additional determinants. Junne and Spiess (2017) already observed for stop-transfer integration that the hydrophobicity threshold is not only dependent on the number of leucines, but also on a sequence of up to 100 amino acids downstream of the H-segment . Thus, the concept of sequence-autonomous, mainly hydrophobicity dependent integration may not be generally valid. Up- or downstream sequences may influence transmembrane integration efficiency. A comparison of the RI-Lep and the RI-DPAPB model proteins shows the following differences: RI-Lep consists of a reverse signal-anchor, a 37-amino acid linker, and a RI-H segment. RI-DP consists of a normal signal-anchor, a stop-transfer TM domain, a 124- amino acid linker, and a RI-H segment.

To dissect, whether the differences in re-integration efficiency of the abovementioned two model proteins can be explained by the differences in linker length, four DPAPB construct series with shortened linker lengths were created. A former study showed that positively charged amino acids in the regions flanking the H-segment can influence its integration behavior (Lerch-Bader et al., 2008). To avoid different flanking regions the first 16 amino acids upstream of the H-segment were kept unchanged. The deletions were carried out via PCR, reducing the linker lengths from 124 amino acids, to 96, 71, 46, and 19. (**Figure 11 A**)

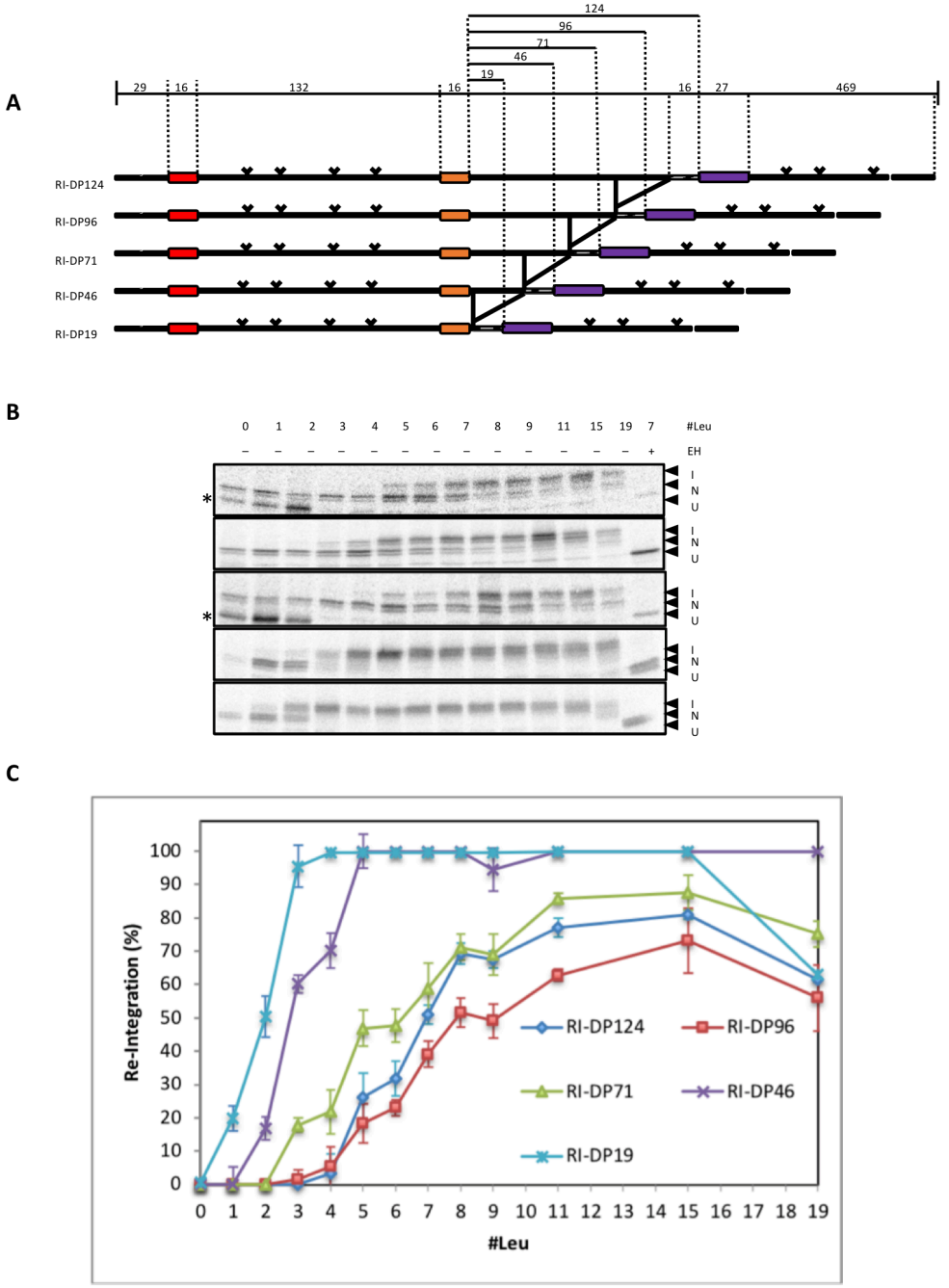


Figure 11: **A** Schematic representation of the model constructs, derived from DPAPB. They consist of a signal-anchor (red rectangular), a stop-transfer domain (orange) and a RI-H segment (purple). Glycoslyation sites are indicated with a Y. The linker between ST-TM domain and RI-H was truncated, letting the first 16 amino acids untouched (strived sequence). **B** Autoradiographs of the constructs series shown in A. The glycosylation pattern of the constructs was analyzed in yeast cells after pulse labeling with [^{35}S]methionine, immunoprecipitation and SDS-gel electrophoresis.The asterisk indicates bands of a truncated version of the protein, produced due to internal initiation (see Appendix A) **C** Quantitation of the autoradiographs. The re-integration efficiency increases for short linker lengths.

The results demonstrated a strong linker-length dependency, with an increased re-integration efficiency for small linkers. The re-integration threshold in relation to the linker length is shown in **Figure 12.**

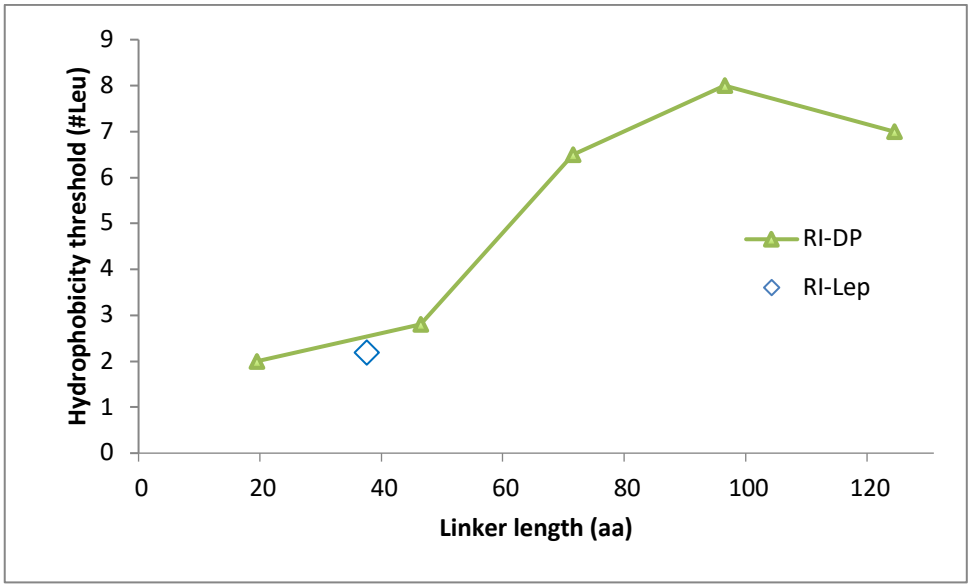


Figure 12: The re-integration threshold plotted against the linker length for the RI-DP constructs. The value increases with growing linker length, thus re-integration efficiency declines. As comparison the re-integration threshold from Lundin et al. (2008) is shown as a blue rectangle. The re-integration threshold of the RI-Lep fits to the curve we obtained for the linker-length dependency.

Upon reduction of the linker length, the hydrophobicity threshold decreased from initially 7 leucines for 124 amino acids to only 2 leucines for 19 amino acids. An exception was the construct DP96, which showed the highest hydrophobicity threshold (8 leucines) among the tested constructs, even though it is shorter than DP124. Only the construct series DP19 and DP46 showed a re-integration efficiency of 100% for more than 4 and 5 leucines, respectively. The abovementioned effect of inefficient integration for very hydrophobic H-segment (19 leucines) occurrd in all construct series, except RI-DP46.

Even though an important observation of this experiment was the finding that re-integration becomes more efficient for short linker lengths, re-integration efficiency and linker length cannot easily be correlated. The relation seemed to be true for RI-DP19, RI-DP46 and RI-DP71, however, the construct series RI-DP96 was less efficient in re-integration, compared to RI-DP124. Moreover, the findings explain the results Lundin et al. (2008) obtained from their integration studies with RI-Lep, which has a relatively small linker with 37 amino acids and showed a hydrophobicity threshold of

2.2 leucines. As can be seen in **Figure 12** (blue rectangle) the hydrophobicity threshold for RI-Lep fited to the curve we obtained, providing evidence that the differences in re-integration efficiency can be explained by the linker-length.

3.2. Reduction of translation speed leads to increased re-integration efficiency

The outcome of the abovementioned experiment demonstrated that the length of the cytosolic loop of RI-DP124 has a great influence on the re-integration efficiency of the RI-H segment. The results showed an overall tendency towards a high re-integration efficiency for small loops, while the re-integration of H-segments preceded by longer loops was less efficient. This length-dependency could have several reasons. Time is one of the most obvious factors influenced by the loop length: a shorter loop needs less time to be synthesized than a long linker. In yeast, the common translation speed is 3.0 – 10 aa/s (Karpinets et al., 2006), hence the smallest linker (19 aa) needs 1.7 - 5.6 seconds to be synthesized, compared to 12.4 - 41.3 seconds for the longest linker (124 aa). During synthesis of the loop, the dynamics of the translocation system might change. It was shown that neighboring TM domains of multispanning membrane proteins show cooperativity: a TM domain which integrates into the lipid bilayer is for some time still localized near the Sec61 lateral gate, thus interactions with a following TM domain helps both domains to overcome the hydrophobicity threshold necessary for integration.

An increased loop length which is always accompanied with an increased time passing until the RI-TM domain emerges from the ribosome exit tunnel, might lead to diffusion of the preceding stop-transfer TM domain into the cytosol, making it unable to aid in re-integration of the following RI-TM domain. *According to this hypothesis, an increased time results in a decreased re-integration efficiency, since the RI-H segment has to integrate into the membrane without assistance of the previous ST-TM domain.*

To modulate time-dependency the translation speed was reduced by using cycloheximide (CHX), a known protein-synthesis inhibitor, which slows down translational elongation. The constructs RI-DP19-L3, RI-DP46-L3 and RI-DP124-L7 were radiolabeled, supplemented with 0, 1 and 2 µg/ml cycloheximide, reducing the translation rate by 0, ~2, and ~4-fold, respectively. The outcome of the experiment is shown in **Figure 13**.

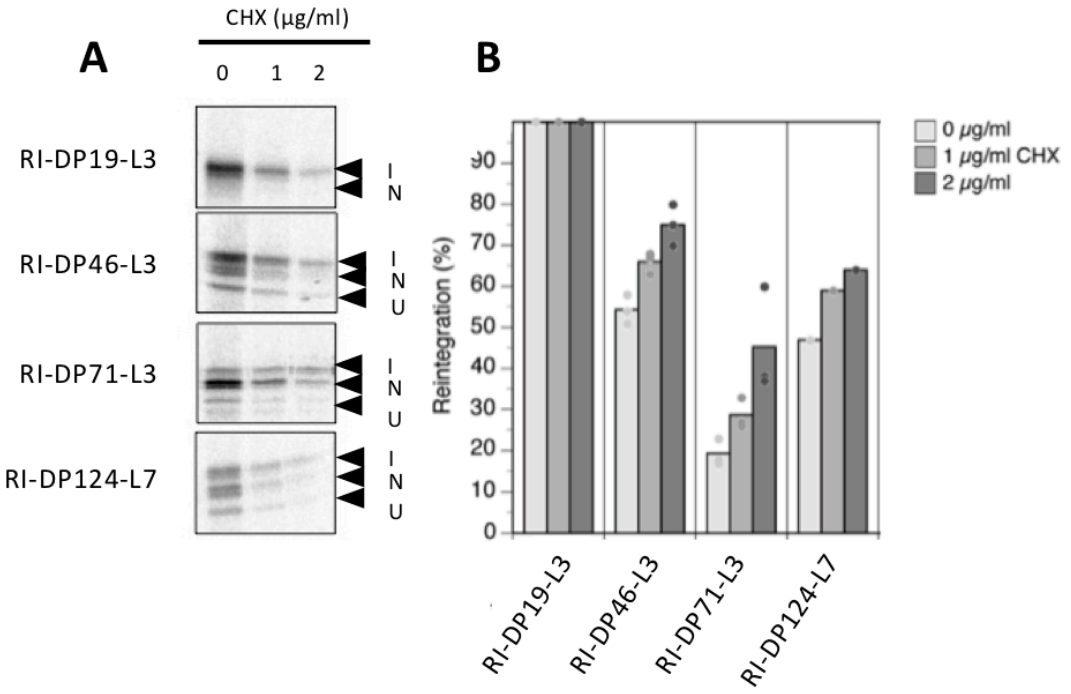


Figure 13: Slowing down of translation speed increased the re-integration efficiency **A** Autoradiographs of the constructs DP19-L3, DP45-L3, DP71-L3 and DP124-L7, supplemented with 0, 1 or 2 µg/ml CHX. The constructs were expressed in yeast and the glycosylation pattern (I = Fully integrated form, N = Partially integrated form) was analyzed after metabolic labeling as described in **Figure 10**. **B** Quantification of the re-integration efficiencies. Re-integration efficiency increases upon treatment with CHX.

Surprisingly, reduction of translation speed led to an increased re-integration efficiency for all constructs, which is in contrast to the scenario described above. Re-integration efficiency increased from 54% for 0 µg/ml CHX to 75% for DP46-L3, from 19% to 45% for DP71-L3 and from 47% to 64% for DP124-L7.

The result of these experiments disproved the hypothesis that the linker-length dependency observed in the previous chapter is caused by the time needed to synthesize the loops. For an explanation of this effect we refer to the discussion.

3.3. Cytosolic loops with naturally unfolded domains enhance re-integration

The outcome of the linker length reduction assays performed in chapter 3.1 demonstrated the importance of including the properties of the upstream sequence when estimating a RI-TM segment's integration efficiency. However, the reduction of the linker length alone did not provide an explanation for the effect, observed for the two construct series RI-DP124 and RI-DP96. To further analyze the underlying mechanisms, it was important to reconsider which variables are changed via linker length reduction, since the deletion of several amino acids in a given cytosolic loop does not only make it shorter, but also alters its intrinsic properties, including protein folding and affinity to chaperones.

To minimize the side effects of folding and chaperone binding, the DPAPB-derived linker sequence was substituted with a generic glycine-serine repeat sequence that is highly flexible and water soluble, and unable to be recognized by chaperones. To directly compare the outcome of this experiment with the experiment performed in 3.1 the linkers were chosen to have a similar length than the original DPAPB linker and its truncated derivatives. Linkers of 31, 53, 76, 99, and 122 amino acids consisting mainly only of glycine-serine repeats were created and cloned into RI-DP124, exchanging codons 196 – 321 with the abovementioned GS repeats. The GS repeat sequences are shown in **Figure 14 A**.

A

```
RI-GS122  LKGASSGSGSGSGSGSGSGSGSGSGSGSSGSGSGSGSGSGSGSGSGSGSGSSGSGSGSGSGSGSGSGSGSGSGSSGSGSGSGSGSGSGSGSGSGSGSSGSGSGSGSGSGSGSGSGSGSGSTS
RI-GS96   LKGASSGSGSGSGSGSGSGSGSGSGSGSSGSGSGSGSGSGSGSGSGSGSGSSGSGSGSGSGSGSGSGSGSGSGSSGSGSGSGSGSGSGSGSGSGSGSTS
RI-GS71   LKGASSGSGSGSGSGSGSGSGSGSGSGSSGSGSGSGSGSGSGSGSGSGSGSSGSGSGSGSGSGSGSGSGSGSGSTS
RI-GS46   LKGASSGSGSGSGSGSGSGSGSGSGSGSSGSGSGSGSGSGSGSGSGSGSGSTS
RI-GS19   LKASSGSGSGSGSGSGSGSGSGSGSGSGSTS
```

B

C

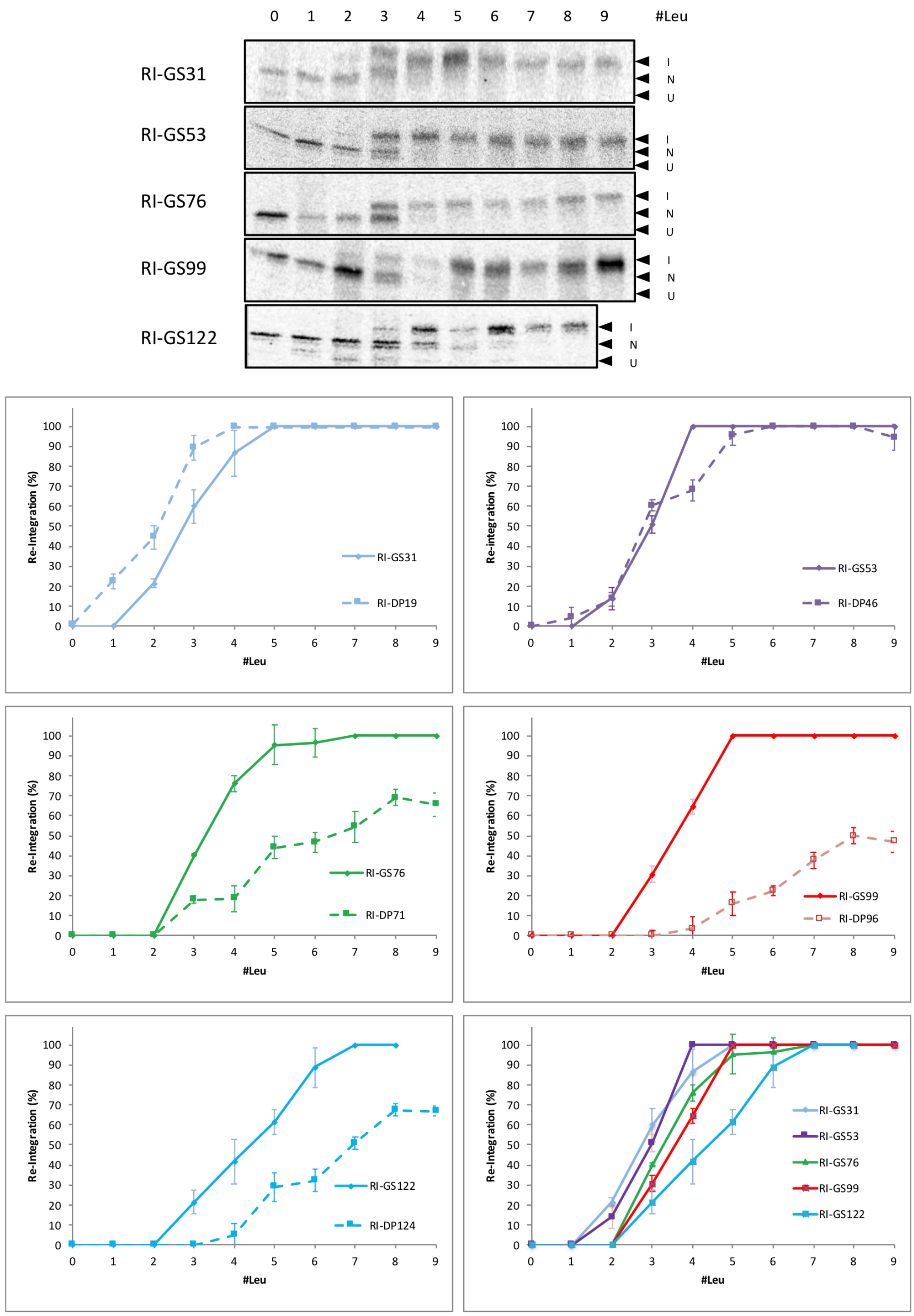


Figure 14: The non-folding GS repeat constructs integrate more efficiently than the original DP sequences. **A** The GS repeat sequences used as cytosolic loop. **B** The glycosylation pattern (I = Fully integrated, N = partially integrated, U = unglycosylated) was determined after metabolic labeling as described in **Figure 10**. **C** shows the re-integration graphs of the GS-repeat constructs (solid line) and compares them to the DP linker constructs (dashed line). In the lower right graph the curves for all GS-repeat constructs are blotted.

The autoradiographs and the quantifications are shown in **Figure 14 B** and **C.** The quantifications compare the re-integration efficiency of the DPAPB natural sequence with the GS-repeat sequences. For the short linker lengths (DP17/GS31 and DP46/GS53), the differences between natural and generic linkers were not significant. However, the construct pairs DP71/GS76, DP96/GS99 and DP124/GS122 showed a significant difference in integration efficiency. In general GS sequences were more efficient and the length dependency was largely abolished. **Figure 15** shows the re-integration threshold for the GS-repeat constructs and correlates them to the linker length. For comparison the linker length dependency of the RI-DP constructs is shown.

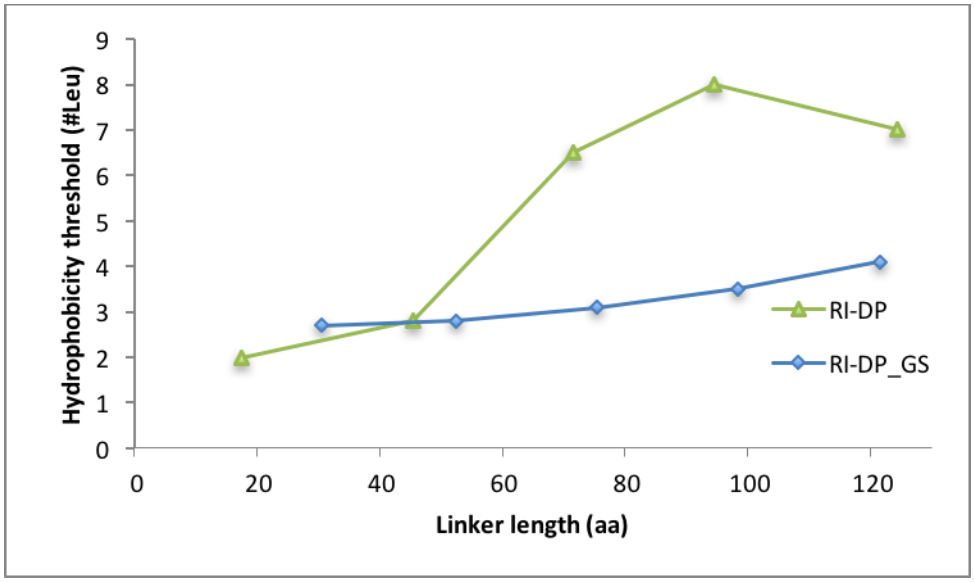


Figure 15: Comparison of the linker length dependency for RI-DP (green) and RI-GS (blue). The re-integration threshold is blotted against the linker length.

A weak linker-length dependency was also observed for the GS-repeat constructs, spanning from 2.7 leucines for RI-GS31 to 4.4 leucines for RI-GS122. However, this effect was not as strong as for the native linker, where a range of 2 – 8 Leucines was observed. The result demonstrated that the apparent linker-length dependency for the natural sequences was not caused by the length reduction, but by changing other intrinsic features of the sequences. These differences only result in an altered re-integration efficiency for loops longer than 50 amino acids.

The GS repeat sequences were used to create cytosolic loops that are hydrophilic, flexible and do not fold. Such intrinsically disordered domains can also be found in

natural proteins. One example is nucleoporin, a protein which is involved in nuclear import (Peleg and Lim, 2010). To test if naturally unfolded (or disordered) domains behave similar to the generic GS repeat sequence, we replaced the GS sequence with a part of nucleoporin. We cloned codons 614 – 711 from yeast Nup1 (Uniprot entry: P20676) into RI-GS99 replacing the cytosolic loop by the Nup1 sequence to test how naturally intrinsically disordered domains influence the re-integration efficiency of a following re-integration TM domain. The result of this experiment is shown in **Figure 16**. RI-Nup102 is compared to RI-GS99, which has a similar loop length.

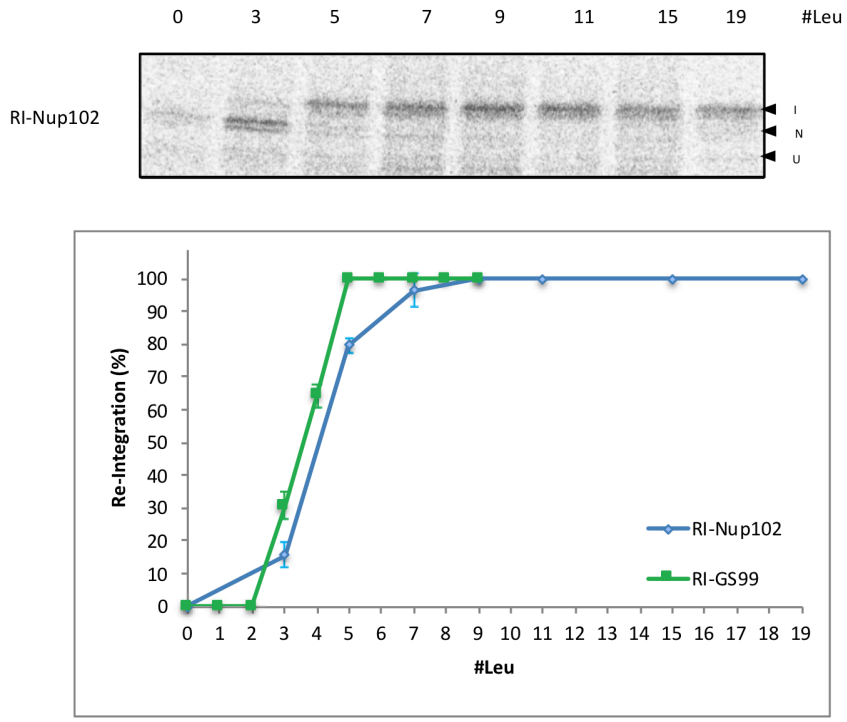


Figure 16: Analyzing the integration behavior of constructs with an intrinsically disordered domain in front of the H-segment. **A** Autoradiograph of the construct RI-Nup102 after expression of the construct in yeast and metabolic labeling as described in **Figure 10**.. **B** Quantification and comparison of RI- NUP102 (blue) with RI-GS99 (green). Both construct series are in a similar efficiency range.

The hydrophobicity threshold for RI-Nup102 was found to be ~4.0 leucines and is very similar to the threshold for RI-GS99, which was ~3.5 leucines. Interestingly, the decrease in re-integration efficiency for H-segments with 19 leucines was not observed for RI-Nup102.

The similarity of the re-integration efficiencies of totally different amino acid sequences provided first evidence that naturally disordered domains in the cytosolic loop have an influence on the re-integration behavior of the H-segment.

3.4. Polypeptide folding affects re-integration efficiency

The outcome of the abovementioned experiments led us to the conclusion that the linker-length effect was mainly caused by the different conformational properties of the linker sequences. While GS-repeats and Nup102 are naturally disordered, the DP-derived sequences are likely to form partial secondary structures. However, the loop sequences of the DP constructs were taken out of the context of the native protein (see **Figure 11**), so they are unable to attain a native fold. Instead they are likely to form molten globule intermediates.

To analyze how domains with a defined structure affect the re-integration efficiency, we introduced wild-type zincfinger motifs and their folding-deficient counterparts into the cytosolic loop. The zincfinger motif consists of only 29 amino acids with the sequence KPYPCGLCNRCFTRRDLLIRHAQKIHSGN. Previous studies demonstrated that zinc fingers can already form in the ribosome exit tunnel (Nilsson et al., 2015b). The cysteine residues at position 5 and 8 are coordinating a zinc ion, together with the histidines at position 21 and 26 (marked in red). Upon mutation of the histidine at position 21 to a proline, zinc coordination is lost and the motif cannot fold into its native structure. Several constructs were created as shown in **Table 1**.

Table 1: Schematic description of the zincfinger constructs used to analyze the effect of folding on re-integration Zf means zincfinger motif, while mZf corresponds to the mutated zincfinger motif..

Construct Name	Schematic cytosolic loop sequence	Total linker length
RI-ZG61	LKGASS-(Zf)-GS-$(SG)_{11}$-STS	61 aa
RI-PG61	LKGASS-(mZf)-$(SG)_{11}$-STS	61 aa
RI-4xZf103	LKGASS-(Zf)-GSS -(Zf)-GSS-(Zf)-GSTS	103 aa
RI-4xPf103	LKGASS-(mZf)-GSS-(mZf)-GSS-(mZf)-GSTS	103 aa

For RI-ZG61/PG61 the position of the zincfinger motif is cloned on the N-terminal end of the cytosolic loop and is followed by the generic GS repeat sequence to ensure that the zincfinger motif has enough time to fold correctly and does not interact with other amino acids in the cytosolic loop. The constructs RI-4xZf103/4xPf103 were cloned to analyze how accumulations of several rapidly folding domains affect the re-integration efficiency. The autoradiographs are shown in **Figure 17 A** and **B**, together with the quantifications.

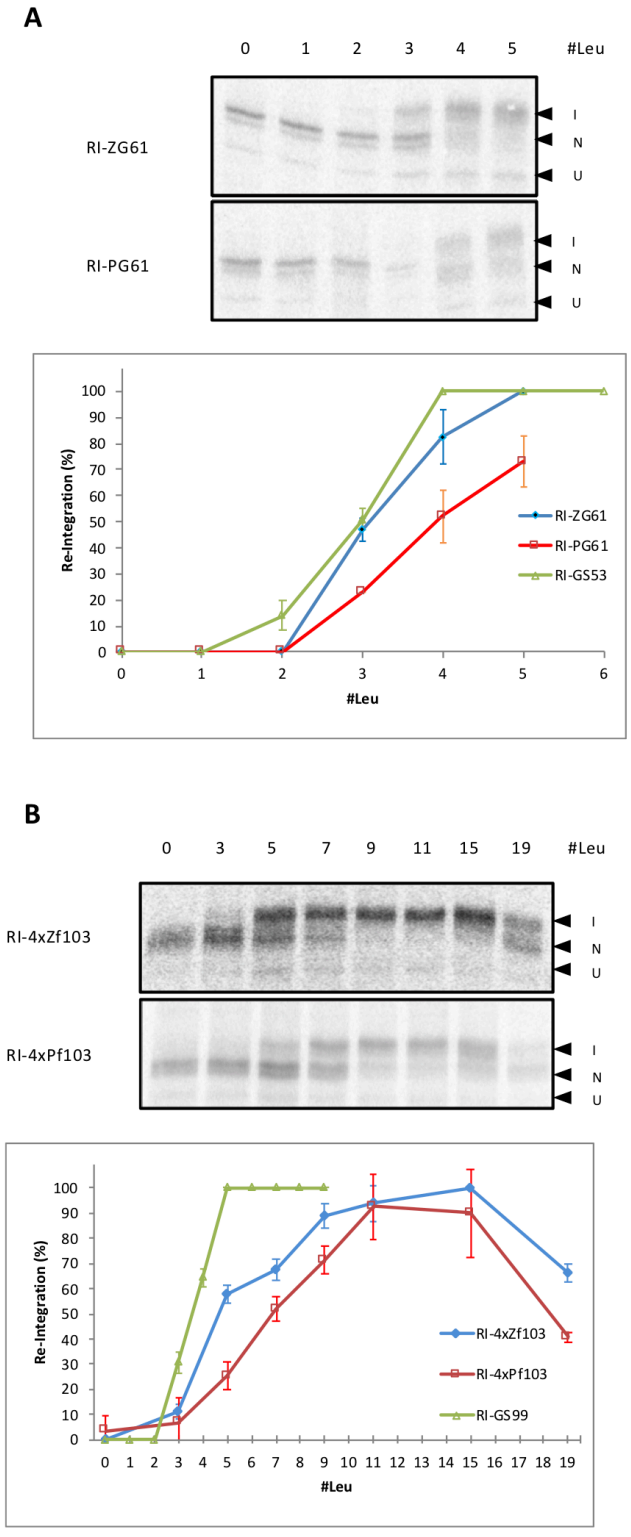


Figure 17: Rapidly folding zincfinger domains influence the integration efficiency of RI-H segments. Wildtype zincfinger (blue) motifs in front of the H-segment result in a higher re-integration efficiency compared to a mutated zincfinger (red) motif. GS repeat constructs with comparable loop lengths are shown in green **A** Autoradiographs and quantification of RI-ZG61 and RI-PG61. **B** Autoradiographs and Quantification of RI-4xZf103 and RI-4xPf103. The glycosylation pattern (I = Fully integrated form, N = partially integrated form, U = unglycosylated) was analyzed after expression of the constructs in yeast, followed by metabolic labeling.

Apparently, RI-ZG61 with a rapidly folding zincfinger domain was almost as efficient as RI-GS53, while for the folding-deficient zincfinger mutant RI-PG61 the re-integration efficiency was decreased. The constructs RI-4xZf103/4xPf103 showed a similar result, however the folding-effect was not accumulative and RI-GS99 was clearly more efficient than RI-4xZf103. As observed for several other construct series RI-4xZf103/4xPf103 showed a decreased re-integration efficiency for the 19-leucine H-segment. Whether all four zincfingers in such closely spaced arrangement are able to fold is not certain.

Spectrin domains are also able to rapidly fold in the ribosome exit-tunnel (Nilsson et al., 2017), providing a second opportunity of testing how isolated folding motifs influence the re-integration efficiency of the H-segment. An advantage of the spectrin domain is, that only one bigger domain is folding into a defined structure, thus the possibility that the folding of several consecutive folding domains is inefficient is avoided. The following spectrin sequence was cloned into RI-DP, replacing the cytosolic loop, resulting in the construct RI-Sp125:

…LKGASSSSKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDL
TGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEEIQQR
LAQFVDHWKELKQLAAARGQRLEEHRLGSTS…

The spectrin domain is shown in green, linker sequences up- and downstream of the spectrin domain are marked in red. The total loop length is 125 amino acids.

As can be seen in **Figure 18** the re-integration efficiency for RI-Sp125 and RI-4xZf103 were in a similar range. However, both were less efficient than RI-GS99. We concluded from these experiments that the folding of the cytosolic loop affects the re-integration efficiency in a way that rapidly folding to a stable structure favors re-integration. In contrast, H-segments preceded by sequences that fail to fold into a native structure are less efficient in re-integration, since the folding-deficient mutants are worse in providing an environment suitable for efficient re-integration. The mutated zincfinger sequence cannot obtain its native structure, however it is unlikely that it will stay totally unfolded in the cytosol. Thus, it will obtain a non-native structure that we believe can be described as a molten globule. It remains to be shown that an H-segment after a folding-deficient spectrin domain is also less efficient in re-integration.

The comparison of the rapidly folding zincfinger and spectrin motifs, the folding-deficient mutant and the GS repeat constructs leads to the following order, with respect to re-integration threshold of the H-segment: GS repeat < folded zincfinger < folded spectrin > folding-deficient zincfinger. This led us to the conclusion that GS repeat constructs and rapidly folding domains are somehow providing a similar situation, promoting re-integration, while folding-deficient zincfinger motifs are less favorable for a re-integration domain.

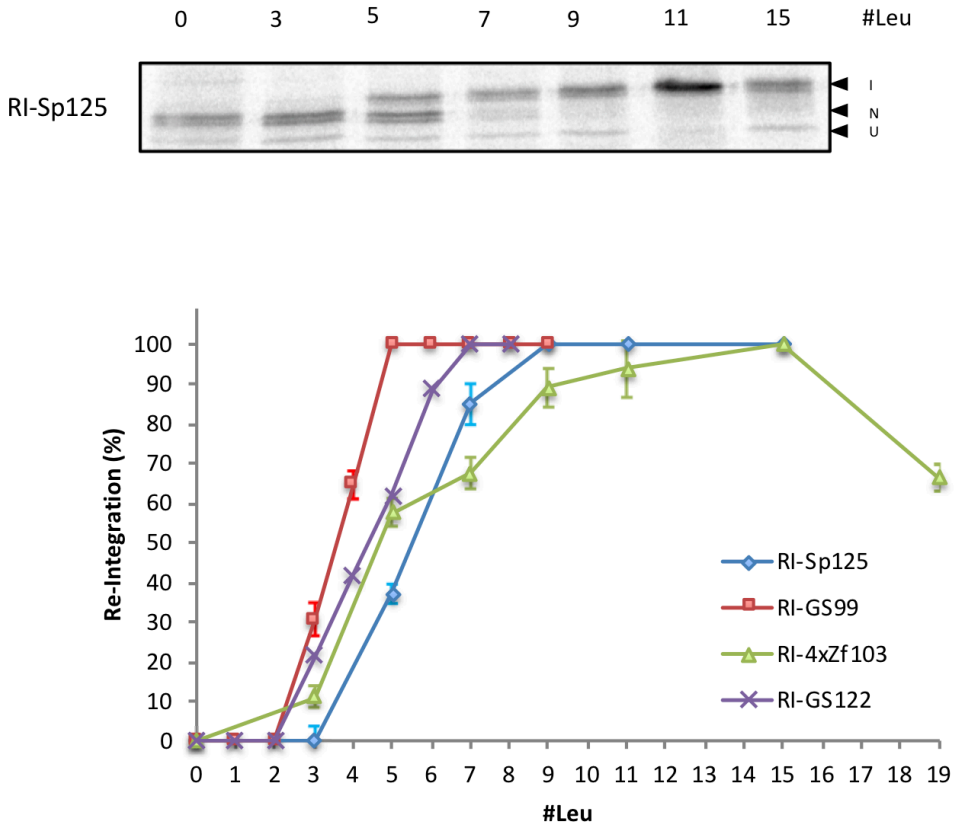


Figure 18: Analyzing the integration behavior of a rapidly folding spectrin domain. Above: Autoradiograph of RI-Sp125. Below: Quantification of RI-Sp125 (blue) and comparison with the other rapidly folding domain construct RI-4xZf103 (green), which has four zincfinger motifs in a row. The intrinsically disordered RI-GS99 (red) and RI-GS122 purple) are shown for comparison. All construct series integrate efficiently, with the GS repeat construct being the most efficient.

3.5. Chaperone binding motifs increase the re-integration efficiency

Intrinsic features of the loop sequence which might affect the efficiency of re-integration can be numerous: Folding properties and how folding of the loop affect the re-integration have been studied in previous chapters. Another intrinsic feature of the loop sequence is its ability to bind chaperones. Chaperone binding is common for most folding proteins *in statu nascendi*, To dissect how co-translational chaperone binding to the cytosolic loop affects the re-integration efficiency, we used sequences that are known for their high or low affinity to Hsp70 chaperones. Such sequences were identified for the ER resident Hsp70 chaperone BiP by affinity panning of libraries of bacteriophages (Blond-Elguindi et al., 1993). A loose consensus sequences for chaperone binding, which has the sequence Hy(W/X)HyXHyXHy (Hy = hydrophobic amino acid, W = Tryptophan, X = any amino acids), was determined. Consistent with the broad variety of different sequences that have to be recognized by chaperones, the consensus sequence shows a high redundancy. Since most cytosolic chaperones involved in co-translational folding are also of the Hsp70 family we assumed that the chaperone binding affinity to the detected sequences is similar (Clerico et al., 2015) .

The following sequences with high and low chaperone-binding affinity were chosen:

- Good chaperone binding: HWDFAWPW and FWGLWPWE
- Low chaperone binding: AGEYYAAL

In addition to these sequences a scrambled sequence was created out of the good chaperone binding motif, where the overall composition of the loop remains unchanged, while the order of the amino acids was scrambled in a way that they do not match the consensus sequence. This scrambled sequence was also considered to be a low chaperone binding motif.

The cytosolic loop of RI-DP was substituted by these sequences. In each case the loop contained 6 good or poor chaperone binding sequences, separated by the hydrophilic spacer ASRSGDSA. A scheme of the loops amino acid sequences is shown in **Table 2**.

Table 2. The loop sequences cloned into RI-DP. The good chaperone binding sequences are highlighted in blue, the spacers are marked in black. Bad chaperone binding sequences are shown in red and orange.

Construct series	Loop sequence
RI-Chp109: Good chaperone binding affinity	LKGASTQESGKSAHWDFAWPWASRSGDSAFWGLWPWEA QESGKSAHWDFAWPWASRSGDSAFWGLWPWEAQESGKS AHWDFAWPWASRSGDSAFWGLWPWEAQESGSTS
RI-Scr109 Bad chaperone binding affinity	LKGASTQESGWKWSAHSWRFAPALSWSDGAFEWAPDWG QESGWKWSAHSWRFAPALSWSDGAFEWAPDWGQESGW KWSAHSWRFAPALSWSDGAFEWAPDWGQESGSTS
RI-Scr109 Bad chaperone binding affinity	LKGASTQESGKSAAGEYYAALASRSGDSAAGEYYAALAQ ESGKSAAGEYYAALASRSGDSAAGEYYAALAQESGKSAA GEYYAALASRSGDSAAGEYYAALAQESGSTS

The outcome of the experiment is shown in **Figure 19**. The autoradiographs and their quantifications confirmed our hypothesis that differences in chaperone binding affinity in the cytosolic loop result in clearly different re-integration efficiencies. RI-Chp109 showed an increased efficiency, compared the non-chaperone binding constructs RI-nChp109. Similarly, the scrambled constructs RI-Scr109 were in a comparable inefficient range to RI-nChp109. This demonstrates that the re-integration efficiency of RI-Chp109 is caused by the high affinity to chaperones and that destruction of the chaperone-binding sequences by scrambling the amino acids is most likely due to poor chaperone binding. Furthermore, RI-Chp109 and RI-GS99 had an almost identical re-integration threshold (3.4 leucines to 3.5 leucines), while it was significantly lower for RI-nChp109 (6.2 leucines) and RI-Scr109 (6.8 leucines).

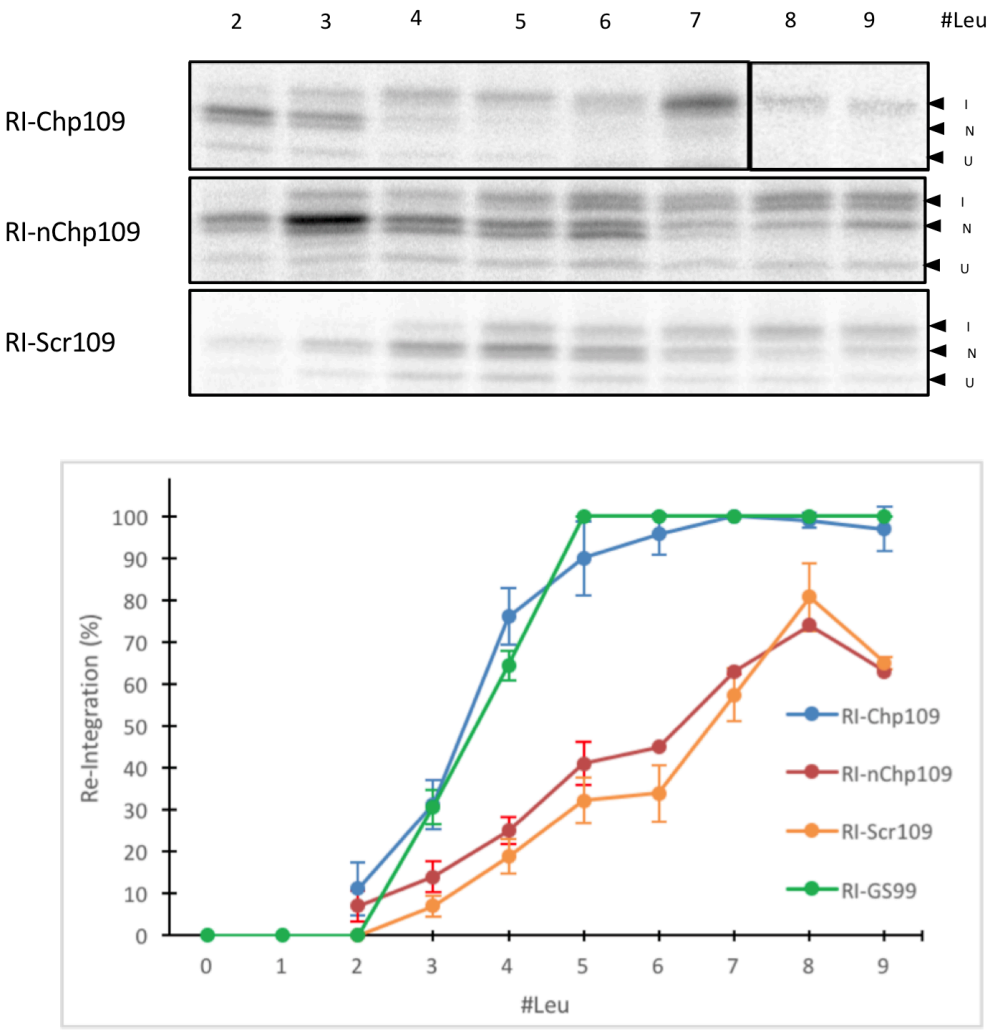


Figure 19: The effect of good and bad chaperone binding sequences on the re-integration efficiency of a following RI-H segment. **Above:** Autoradiographs of the constructs RI- Chp109, RI-nChp109 and RI-Scr109 after metabolic labeling as described in **Figure 10**. **Below:** Quantifications of the autoradiographs for RI-Chp109 (green), RI-nChp109 (red) and RI-Scr109 (orange). Re-integration is very efficient for RI-Chp109, while RI-nChp109 and RI-Scr109 are less efficiently integrated. Both are in a similar range, suggesting that the effect observed is indeed due to chaperone binding.

The results indicate that efficient chaperone binding to the cytosolic loop facilitates re-integration. Cytosolic chaperones keep the nascent chain in an unfolded and folding-competent state, thus the cytosolic loop *in statu nascendi* is mimicking the intrinsically disordered loop of the GS repeat constructs. In contrast to this, RI-nChp109 and RI-Scr109 have a poor affinity to chaperones. However, due to the generic nature of both sequences they cannot fold into defined structures, leaving the formation of a molten globule as the most likely outcome. Formation of a molten globule decreases the re-integration efficiency as it was already pointed out for the RI-DP constructs.

3.6. Cytosolic loops of natural membrane proteins lead to different re-integration efficiencies

The findings in the previous experiments demonstrated that the efficiency of integration is not only dependent on the hydrophobicity of a domain alone, but is also affected by upstream sequence properties. However, under physiological conditions a cell needs to synthesize membrane proteins efficiently. Thus, cytosolic loops of natural membrane proteins should be evolutionary selected to fold into domains that ensure proper re-integration of the following RI-TM domain.

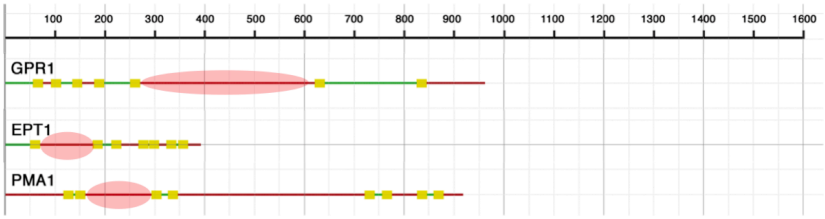
We were interested in testing cytosolic loops of natural membrane proteins, to analyze how they influence the integration efficiency of the following re-integration TM domain. To do so we chose cytosolic loops of several membrane proteins, listed in **Table 3.**

Table 3: List of natural proteins chosen for the assay.

Membrane protein	Uniport ID	Loop number	Loop length
GPR1 G protein-coupled receptor GPR1	Q12361	5	348 aa
EPT1 Choline/ethanolaminephosphotransferase 1	P22140	1	109 aa
PMA1 Plasma membrane ATPase 1	P05030	2	133 aa

The cytosolic loops were cloned into our model protein RI-DP, using a yeast PCR approach. The autoradiographs and the quantifications are shown in **Figure 20**.

A



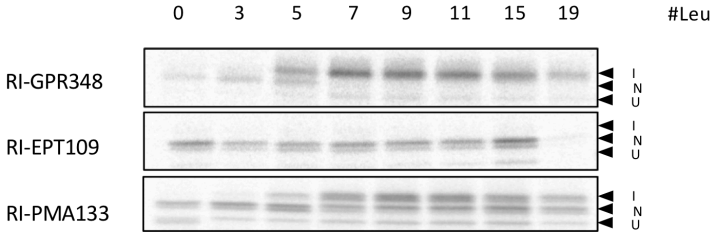
B

```
GPR348    LKSESKRIKAQIGDFNHNVLEEEKEKKKLFGLGHWGKAKWYFRSYFKLPLLHLLRNLKNFFTISFIDPNEETDDSGSSNG
          TFNFGESSNEIPTLFRKTNTGSDENVSASGGVRLLDYNSAKPLDMSKYAMSEQPDLERNNPFDCENDITLNPSELVSKQK
          EHKVTFSVENEGLDTRKSSMLGHQTFSCQNSLESPLAMYDNKNDNSDITSNIKEKGGIINNNSNNDDDDNNNNNDNDNDN
          NNSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNSNNIKNNVDNNNTNPADNIPTLSNEAFTPSQQFSQERVNNNADRCEN
          SSFTNVQQHFQAQTYKQMKKRRAQIQKN

EPT109    LKGASSDPNLNTDTPRWTYFSYALGVFLYQTFDGCDGVHARRINQSGPLGELFDHSIDAINSTLSIFIFASETGMGFSYN
          LMLSQFAMLTNFYLSTWEEYHTHTLYLSE

PMA133    LKQEFQAGSIVDELKKTLANTATVIRDGQLIEIPANEVVPGEILQLESGTIAPADGRIVTEDCFLQIDQSAITGESLAAE
          KHYGDEVFSSSTVKTGEAFMVVTATGDNTFVGRAAALVGQASGVEGHFTEVLN
```

C



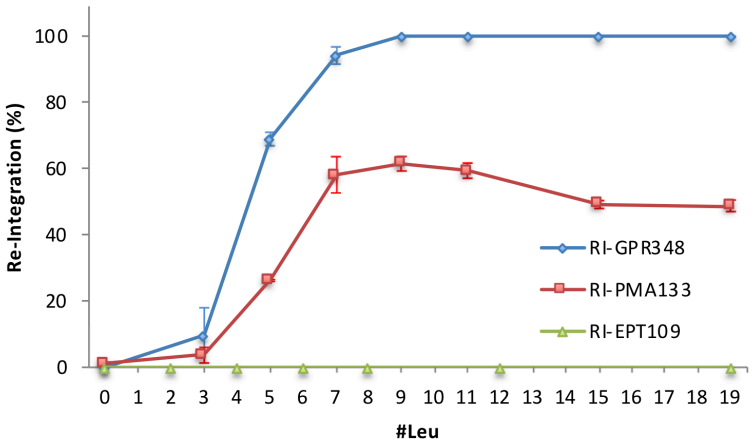


Figure 20: Selected natural loops were cloned into the model protein RI-DP to dissect how randomly chosen cytosolic loops affect the re-integration efficiency of a subsequent H-segment. **A** Schematic overview of the chosen natural proteins. TM domains are indicated in yellow. Lumenal domains are shown as a green line, cytosolic domains as a red line.The cytosolic loops cloned into RI-DP are marked with a red cycle **B** Primary sequences of the natural loops cloned into the model protein.. **C** Autoradiographs and quantifications for RI- GPR348, RI-EPT109 and RI-PMA133. RI-GPR348 (blue) re-integrates very efficiently, compared to RI-EPT109 (green) which does not re-integrate at all. RI-PMA133 re-integrates only maximal 60% for 9 leucines.

Interestingly, the loops influenced the re-integration behavior of the following H-segment very differently. While the GPR348 loop allowed for very efficient re-integration, the PMA133 loop only produced an intermediate integration level, and EPT109 did not integrate even with a 19-leucine H-Segment.

The wide range of results may appear surprising, since one might have expected that natural loops should be conducive top efficient re-integration. Instead, it seems that a re-integration-prone environment is only established by RI-GPR348.

From the experiments in the previous parts we concluded that efficient folding of isolated and rapidly folding domains as well as intrinsically disordered loops facilitate re-integration. The efficient re-integration of RI-GPR348 suggests that the loop matches one of these two criteria: It is either rapidly folding into an isolated folding domain, or it is providing an intrinsically disordered region, which has been shown to facilitate efficient re-integration. We used the prediction tool IUPred2A (https://iupred2a.elte.hu) predict the disorder of GPR348. The prediction is shown in **Figure 21**.

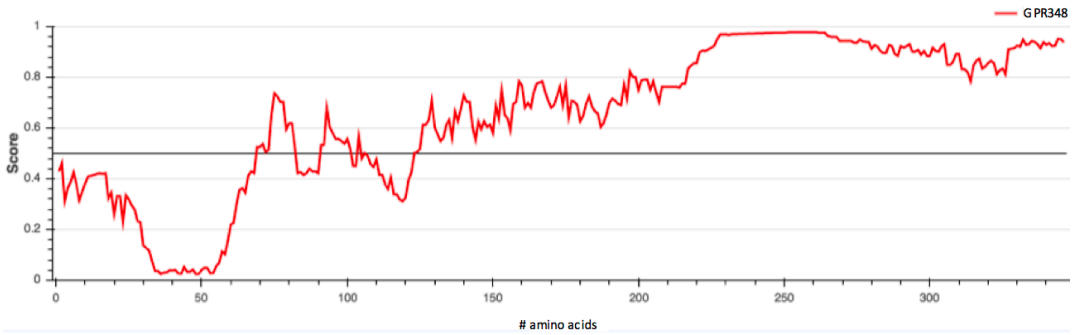


Figure 21: Prediction of the intrinsically disorder of RI-GPR348. The score provided for each amino acids can range from 0 to 1, corresponding to the likelihood of the amino acid of being part of an intrinsically disordered domain.

Interestingly, the result points towards a high likelihood that the loop is intrinsically disordered. Especially the C-terminal part of the loop has a score of 0.8 – 1, indicating that this part of the sequence is most likely disordered. This explains the high re-integration efficiency of RI-GPR348 and is in accordance with the previous results.

The inefficiency of RI-EPT109 might be due to the fact that we have taken the cytosolic loop out of context into another protein sequence It might be that for efficient folding additional sequences in other parts of the native EPT protein are necessary and that the isolated EPT loop does not provide an environment which is suitable for re-integration of a following RI-TM domain. The same issue may cause the low efficiency for RI-PMA136.

3.7. Modulation of the translocon pore's hydrophobicity by mutation of the constriction ring

When stop-transfer TM domains integrate into the lipid bilayer of the endoplasmic reticulum, the domain moves from the ribosome exit tunnel to the inside of the translocon, where it leaves the translocon through its lateral gate into the hydrophobic environment of the ER membrane. In contrast to this, the insertion mechanism for re-integration TM domains is still elusive. Unlike stop-transfer TM domains, re-integration TM domains need to insert together with a part of their downstream sequence in a hairpin-like conformation. Recently, a different model of insertion was proposed (Cymer et al., 2015): the hydrophobic RI domain does not enter the translocon pore itself, but associates directly with the surface of the ER membrane in the vicinity of the translocon, so that only the downstream sequence of the substrate is fully moving through the translocon pore (**Figure 22 A**).

To contribute to this question, we modulated the hydrophobicity of the translocon pore. The translocon mutants Sec61-6Ala and Sec61-6Ser, which were already established in our lab (Junne et al., 2010), were used, where the 6 residues that constrict the translocon pore were substituted by 6 alanines or 6 serines. For Sec61-6Ser the pore is more hydrophilic, compared to wildtype, while the Sec61-6Ala mutant provides a less hydrated and thus more hydrophobic pore (Demirci et al., 2013). If the RI-TM domain indeed does not enter the pore but only associates with the ER membrane one would expect an increased re-integration efficiency for the hydrophilic mutant Sec61-6Ser. The hydrophilic downstream sequence of the substrate's TM domain would translocate more easily through the translocon, enhancing the re-integration efficiency. In contrast, if the potential RI-TM domain inserts into the pore a hydrophilic Sec61-mutant would decrease the re-integration efficiency of the hydrophobic RI-TM domain. The construct series RI-DP124 was used to examine how an RI H-segment behaves when exposed to a translocon with an altered pore structure.

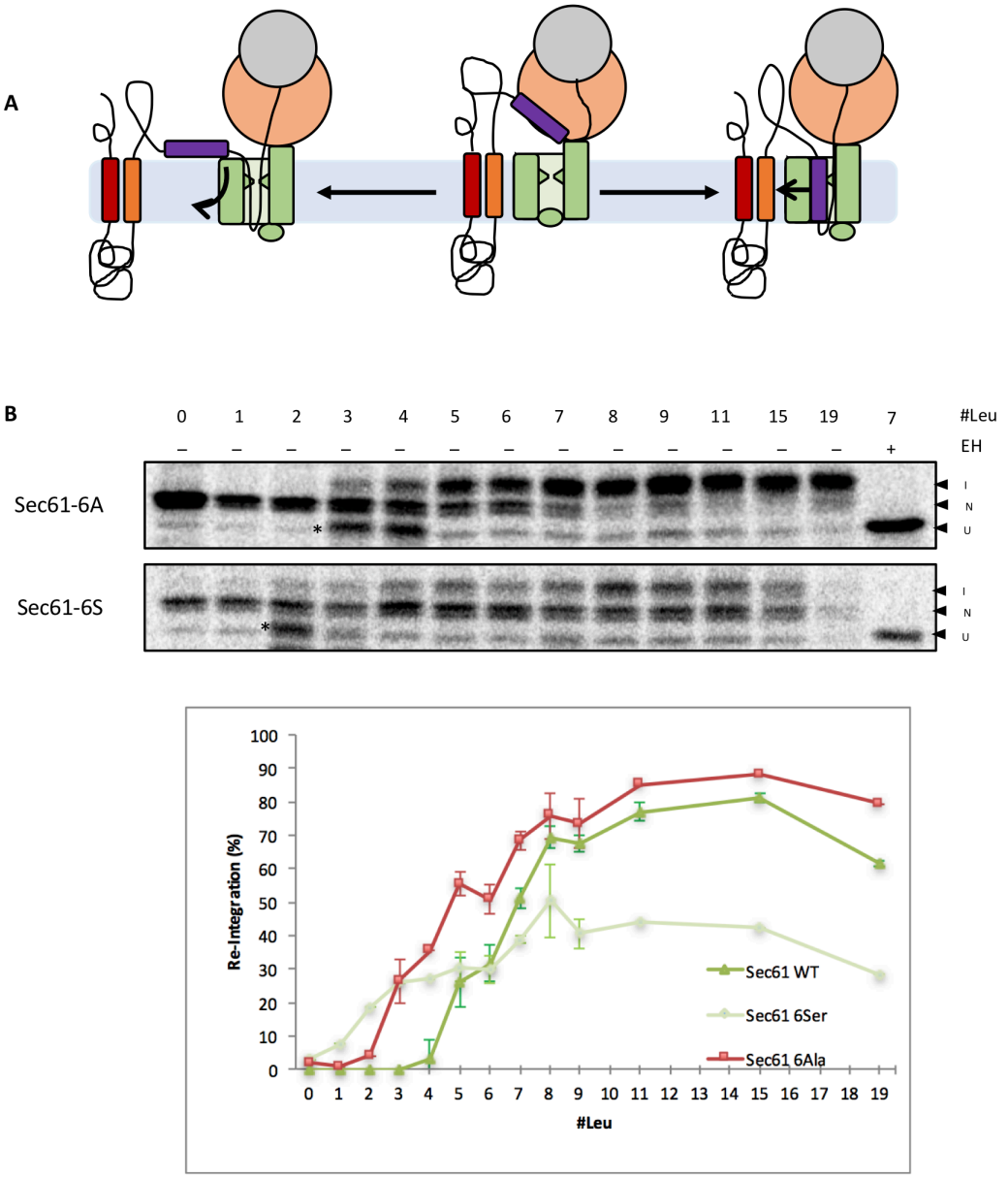


Figure 22: The interaction of the H-segment with the translocon pore **A (left)** Model for re-integration, proposed by (Cymer et al., 2015). The re-integration domain does not enter the pore, but places itself on the membrane instead, only translocating the downstream sequence. **(right)** The conventional model, where the RI-TM domain inserts into the translocon pore. **B** Autoradiographs and quantifications of the RI-DP124 construct series, expressed in Sec61-6Ala (red) or Sec61-6Ser (light green) yeast mutants, followed by metabolic labeling as in **Figure 10**. The quantifications of RI-DP124 expressed in Sec61wildtype is shown for comparison (dark green). While RI-DP124 re-integrates a little bit more efficiently for Sec61-6Ala it is clearly less efficient for Sec61-6Ser.

The autoradiographs and quantifications are shown in **Figure 22 B**. In general, re-integration was found to be less efficient for the hydrophilic Sec61-6Ser mutant, compared to wild-type Sec61, whereas for the Sec61-6Ala mutant translocon, re-integration efficiency was slightly increased. However, for low numbers of leucine in the H-segment the order of the re-integration efficiencies changed.

We concluded from this experiment that the hydrophobic H-segment is indeed inserting into the translocon pore and is not associating with the ER membrane. The results suggest that the hydrophobic H-Segment is repelled by the hydrophilic environment of the mutant Sec61-6Ser, decreasing the re-integration efficiency. This can be explained by the conventional model of pore-insertion of a RI-TM domain, where a potential RI-segment wants to insert into the translocon pore but is hindered due to the hydrophilic environment of the Sec61-6Ser mutant pore. Accordingly, the hydrophobic mutant Sec61-6Ala leads to increased re-integration efficiency, since hydrophobic H-segments can insert into the hydrophobic pore more easily.

3.8. General Discussion & Model

We are still in the beginning of understanding how the topogenesis of multispanning membrane-proteins work and how the processes of translation, integration and co-translational folding act together and influence each other. In this study we examined the influence of the cytosolic loop upstream of a re-integration TM domain on its integration behavior. Therefore, we cloned several sequences into the cytosolic loop of our model protein RI-DP, using intrinsically disordered loops, rapidly folding motifs, natural loops and loops that form a molten globule-like structure.

We propose a re-integration mechanism that is based on the co-translational interaction between the potential TM domain and either the translocon or the upstream cytosolic loop (**Figure 23**)

(1) As soon as a part of the cytosolic loop is synthesized and is experiencing the hydrophilic environment of the cytosol, it will not stay unfolded, but will minimize its free energy by collapsing to a molten globule. The process of rapidly collapsing can be best described as a hydrophobic collapse (see chapter 1.1). The re-integration TM domain can either participate in this structure of the cytosolic loop or it can insert and integrate into the ER membrane. In other words, the molten globule provides a hydrophobic environment competing with that of the translocon and the membrane for the TM domain. Different cytosolic loops offer more or less of this competing environment. Concerning our model protein, longer DPAPB fragments generally produce a more efficiently competing environment than shorter ones. This is in accordance with the finding, that truncated versions of RI-DP124 integrate more efficiently. However, depending on the sequence context the efficiency of molten-globule formation can be different. This explains the reversed order of the construct series RI-DP124 and RI-DP95 (**Figure 11**). The loop of RI-DP95 forms a molten globule in which the H-segment can more easily participate with, compared to RI-DP124, thus the re-integration efficiency is lower for RI-DP95. The behavior of an TM segment when exposed to the different environments that the loops provide is shown in **Figure 23 B**

(2) The intrinsically disordered loops tested in this work remain unstructured and do not provide an alternative mildly hydrophobic phase, but thus favor insertion into the

ER membrane. Interestingly, differences between the molten-globule forming RI-DP and the intrinsically disorderd GS repeat constructs only appear for linker lengths longer than 50 amino acids. This might also explain why the construct from Lundin et. al (2008) shows a similar re-integration efficiency as RI-DP45.
(3) Similarly, rapidly folding motifs hide their hydrophobic residues inside their folds and also do not interact with a potential RI-TM domain. However, mutant zincfingers that cannot attain the native fold are likely to collapse to a molten globule capable of interacting with an H-segment and therefore reducing re-integration efficiency.
(4) Chaperone binding motifs within the cytoplasmic loop are expected to recruit chaperones keeping the polypeptide in an unfolded and soluble state, preventing collapse to a molten globule. For the scrambled version of the same sequence as well as for the loop composed of non-chaperone-binding motifs, the hydrophobic residues will again collapse and compete with membrane re-integration.

Importantly, the rapidly folding of domains into isolated structures (e.g. zincfingers, spectrin domains), the binding of chaperones and the formation of a molten-globule are dynamic processes, that can influence each other. For example, the folding of the zincfinger domain might not be 100% efficient, resulting in a folding-deficient fraction of the wild-type zincfinger, that influences the re-integration behavior of the potential TM domain similarly to the real zincfinger mutant. A fraction of the folding-deficient mutant might be also bound to chaperones, preventing molten-globule formation. On the other hand, chaperone-binding sequences might not bind chaperones 100% efficient, resulting in the partial collapse of the sequence into a molten-globule. We tried to examine these processes as isolated as possible. However, one has to be careful when evaluating the integration efficiency of totally different molten-globule forming sequences, since each sequence has its own intrinsic capacity to collapse into a molten-globule. The inversed integration efficiencies of the constructs RI-DP124 and RI-DP95 demonstrate this problem. The natural cytosolic loops we tested are another example. Only the construct RI-GPR348 has a high re-integration efficiency, which is due to the intrinsically disordered state of a major part of the sequence. The other sequences are taken out of their natural context, thus it remains elusive whether they can fold into their native structure in the context of our model protein. Maybe additional parts of the up- and downstream sequences are required for efficient folding and without those they collapse into a molten-globule.

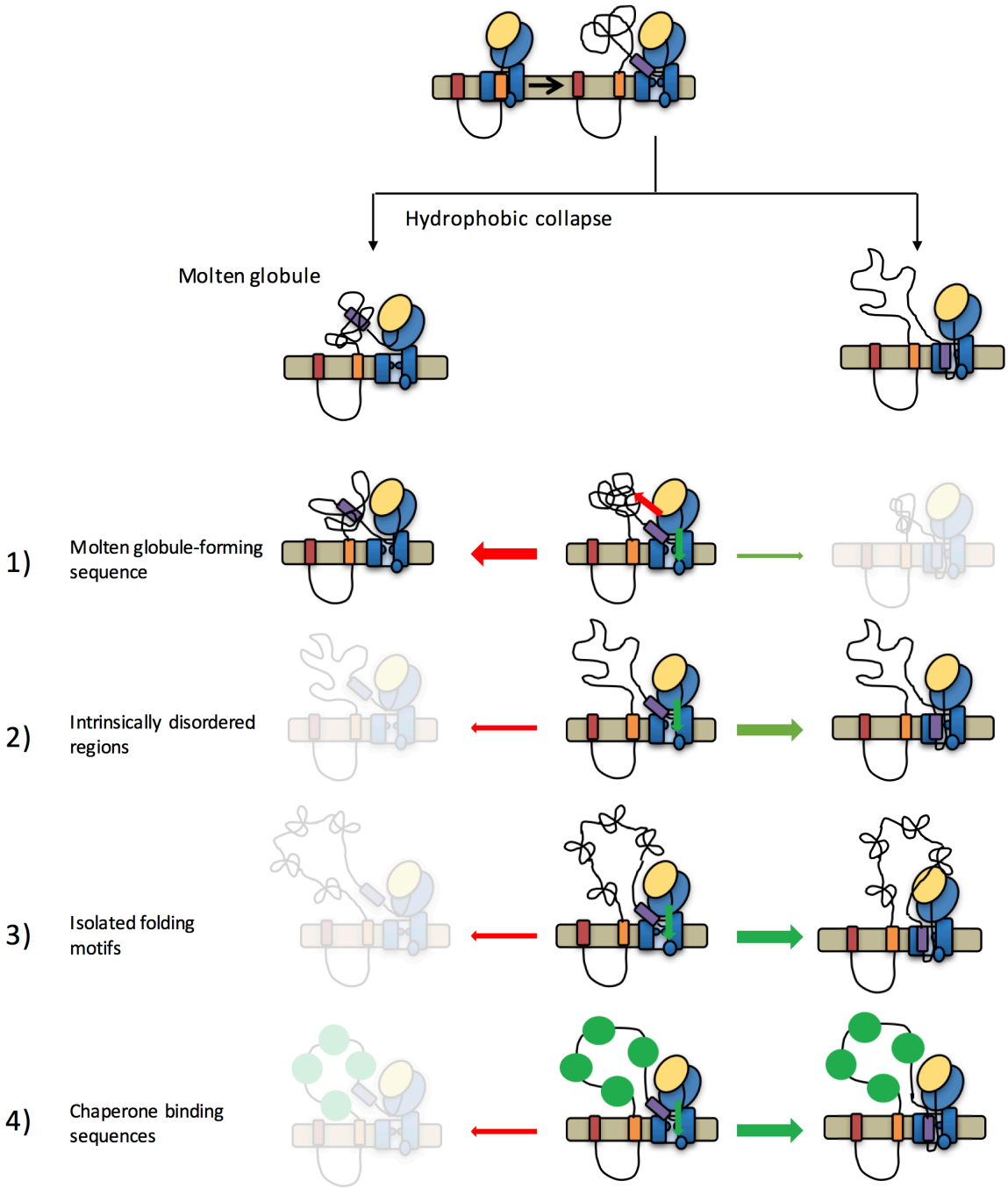


Figure 23: Model for re-integration. When the RI TM domain emerges from the ribosome exit tunnel it can either engage with the translocon, which leads to integration (right arrow) or it can participate in the formation of a molten globule. When a sequence collapses into a molten globule, the H-segment can participate in this formation, reducing re-integration efficiency (1). For intrinsically disordered regions, isolated folding motifs and chaperone binding sequences such an interaction is not possible or strongly reduced (2-4). Since no molten globule is formed the H-segment can inserts into the translocon to reduce its free energy.

In some cases we observed a decreasing re-integration efficiency for H-segments with 19 leucines. In the beginning, this was surprising for us and we still do not have a solution to this problem. Our hypothesis is that the strong hydrophobicity of the 19 leucines H-Segment favors molten-globule participation since pore insertion is less favorable. For re-integration the H-segment has to enter the pore in a hairpin-like structure together with the hydrophilic downstream sequence. The formation of this hairpin might be energetically disfavored for very hydrophobic H-segment. It seems that for constructs with intrinsically disordered loops the reduction does not occur (RI-GPR348 and RI-Nup109). This might be due to the fact that for intrinsically disordered loops no molten-globule is formed, thus the only possibility for the H-segment to reduce its free energy is to insert into the ER membrane.

How does the cycloheximide experiment fit into this model? CHX slows down the translation rate, but the folding processes are of course not directly affected by CHX. We think that there is an indirect influence: the possibility for the H-segment to participate in the molten-globule formation is reduced. To effectively participate in a molten-globule the sequence downstream of the H-segment is required to have a certain length, to allow the H-segment to diffuse away from the ribosomal exit tunnel into the cytosol. A short downstream sequence can be seen to act as a `leash` that is tethered to the ribosomal active site, preventing the H-segment to diffuse away by holding it back in close vicinity to the translocon. Since CHX slows down the translation process the C-terminal `leash´ grows longer more slowly. Thus, the H-segment spends more time in the intermediate area between the ribosome exit tunnel and the translocon pore, resulting in an increased likelihood that it inserts into the pore.

Taken together, we suggest a model in which the process of re-integration is not only dependent on the hydrophobicity of the potential TM domain, but is also strongly influenced by the characteristics of the preceding cytosolic loop. Integration (into the ER membrane) or participation (with a molten-globule) is dependent on an interplay between the hydrophobicity of a potential TM segment and the capacity of the cytosolic loop to collapse into a molten-globule. Preventing of the hydrophobic collapse into a molten-globule – either by chaperone-binding or isolated folding - increases the re-integration efficiency, since participation of the potential TM domain with the molten-globule cannot occur.

The model we propose needs further validation. To do so, we plan to test more intrinsically disordered domains. Independent of the sequence context intrinsically disordered domains should always provide an environment that favors integration. In addition to that, we plan to destroy the rapidly folding spectrin domain in construct RI-Sp125. Upon destruction of the folding motif, a molten globule should be formed, thus re-integration efficiency should be lower. For the chaperone binding constructs we plan to quantify the amount of chaperones bound to the cytosolic loops. The most likely interactors are the cytosolic chaperones Ssa and Ssb. Unfortunately, co-immunoprecipitations unfortunately failed so far.

PART II: THE NATURAL PRODUCT CAVINAFUNGIN SELECTIVELY INTERFERES WITH ZIKA AND DENGUE VIRUS REPLICATION BY INHIBITION OF THE HOST SIGNAL PEPTIDASE

4. Introduction

The proper translocation and processing of polypeptides through membranes is a crucial step in the secretory pathway of pro- and eukaryotes. In principle, this makes the translocon a potential target for antibiotic interference, since selective inhibition of this step in prokaryotes would provide a useful tool to fight bacterial infections .

However, the machinery that translocates polypeptides through the plasmamembrane of bacteria (SecYEG) is well conserved and shows homology to the eukaryotic counterpart (Sec61αβγ), which is required for the translocation of polypeptides into the endoplasmic reticulum (Park and Rapoport, 2012). This makes it unlikely that a compound selectively blocks the prokaryotic translocon. However, other proteins that participate in the translocation process are less conserved. While the signal peptidase is functionally conserved, it is sufficiently different in structure to be a potential target for antibiotic interference (Auclair et al., 2012).

Signal peptidases (SPase) cleave the signal sequence of a variety of polypeptides during translocation, a crucial step in polypeptide processing. The *E.coli* signal peptidase (Leader peptidase) is among the most studied and best characterized peptidases (Dalbey, 2013). It was found to be a monomeric membrane protein, cleaving the signal sequence via a catalytic dyad, involving the amino acids serine and lysine and is as such classified as a serine-protease (Dalbey, 2013). In contrast, eukaryotic signal peptidases are multi-subunit complexes. The exact stoichiometry still remains elusive. The yeast signal peptidase involves at least four subunits (Sec11p, Spc1p, Spc2p, Spc3p), with Sec11p as the catalytic subunit. The mammalian counterpart is composed of five subunits (SPC25, SPC22/23, SPC21, SPC18 and SPC12). SPC12 and SPC18 are isoforms and homologous to the yeast Sec11p, thus also contain the active site, which

was identified to be a serine-protease, acting over a Serine-Histidine-Aspartate tryad (Walker and Lively, 2013). In mammals, impaired cleavage of hormones has been shown to lead to diseases, like diabetes insipidus (Ito et al., 1993). While most protease inhibitors lack the ability to block signal peptidase activity, the bacterial signal peptidase was found to be inhibited by β-lactam derivatives (Therien et al., 2012). Interestingly, signal peptides itself have been found to be competitive inhibitors of the signal peptidase, like the signal sequences of mutant forms of the hormones preproinsulin and preproparathyroid. In addition to that, it was found a mutant of the maltose-binding protein, where the cleavage site at position +1 is mutated to a proline can also inhibit the prokaryotic signal peptidase (Barkocy-Gallagher and Bassford, 1992). For the mammalian signal peptidase no chemical inhibitors had been identified prior to this study (Dalbey, 2013; Walker and Lively, 2013).Not only compounds against bacterial targets are of great interest, but also compounds against viral infections. Viruses hijack cells and exploit the cellular machinery for replication. One of the most dangerous viral infections is caused by the mosquito-borne dengue virus, which is a flavivirus, like the zika virus. The dengue virus is responsible for more than 20 000 deaths per year (Bhatt et al., 2013). Until now, neither vaccines nor antiviral compounds are available.

The 11kb viral mRNA of the dengue virus is translated into a single polyprotein by the cellular ribosome in the cytosol. To mature, an interplay between viral and host proteases cleave the polyprotein at several internal cleavage sites into the final proteins (**Figure 24**) (Perera and Kuhn, 2008). While the structural proteins C, prM and E form the virion, the other 7 proteins (NS1, NS2A/B, NS3, NS4A/B, NS5) fulfill different cellular functions in hijacking the cell and are in total responsible for the cellular replication of the viral RNA (Norazharuddin and Lai, 2018). This demonstrates, that a functional signal peptidase is crucial for the correct processing of the viral polyprotein and that a compound that inhibits processing or leads to apoptosis of infected cells is of therapeutic interest.

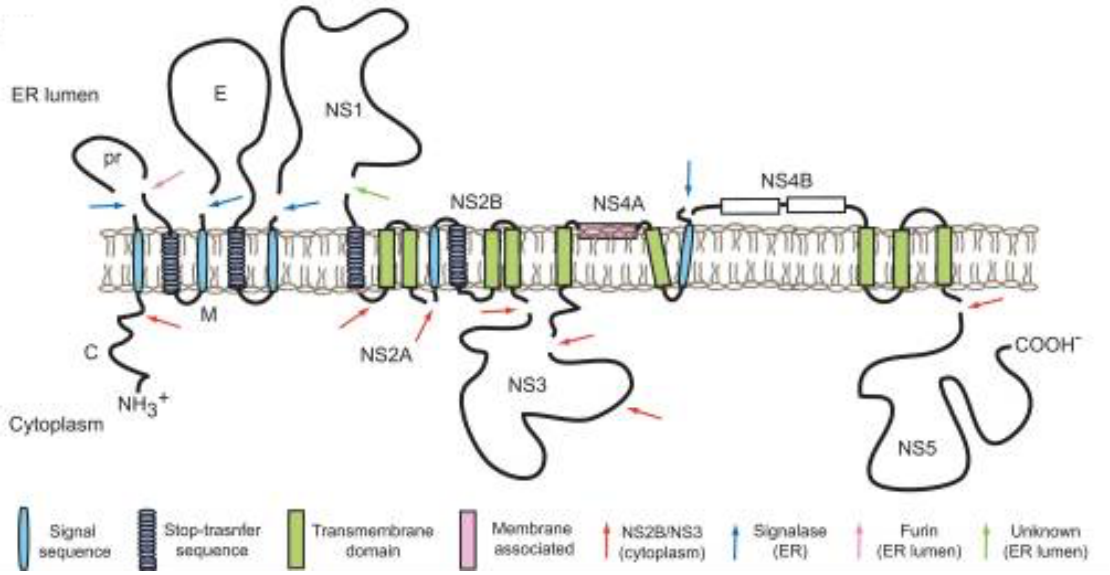


Figure 24: Topology and processing of the dengue virus-polyprotein. The polyprotein is processed by viral and cellular signal peptidases. The 4 cleavage sites for the ER signal peptidase are denoted by a blue arrow. The viral protease NS2B-NS3 cleaves at sites indicated with a red arrow. (Perera and Kuhn, 2008).

5. Aim

To find new therapeutic leads Dominic Höpfner and colleagues (Novartis) screened several compounds for their inhibitory activity against dengue- and zika-virus infected cell lines. The lipopeptide cavinafungin showed an up to 100-fold stronger activity in infected cell lines compared to uninfected cells. To identify the target of cavinafungin, chemogenomic assays in mammalian cells and yeast were performed. In both cases the host's signal peptidase was identified as a possible target.

The aim of this part was to verify whether cavinafungin indeed targets the signal peptidase. To do so, we used a metabolic radiolabeling approach in mammalian cell lines, testing the ability of the signal peptidase to cleave the signal sequences of pre-proinsulin and pre-provasopressin as well as some precursors of the dengue virus-polyprotein when exposed to cavinafungin in a dose-dependent manner.
We could verify the inhibition of the signal cleavage upon treatment with cavinafungin for host and viral proteins, demonstrating that the target of cavinafungin indeed is the signal peptidase. Whether cavinafungin also inhibits the signal peptidase of bacteria still has to be elucidated.

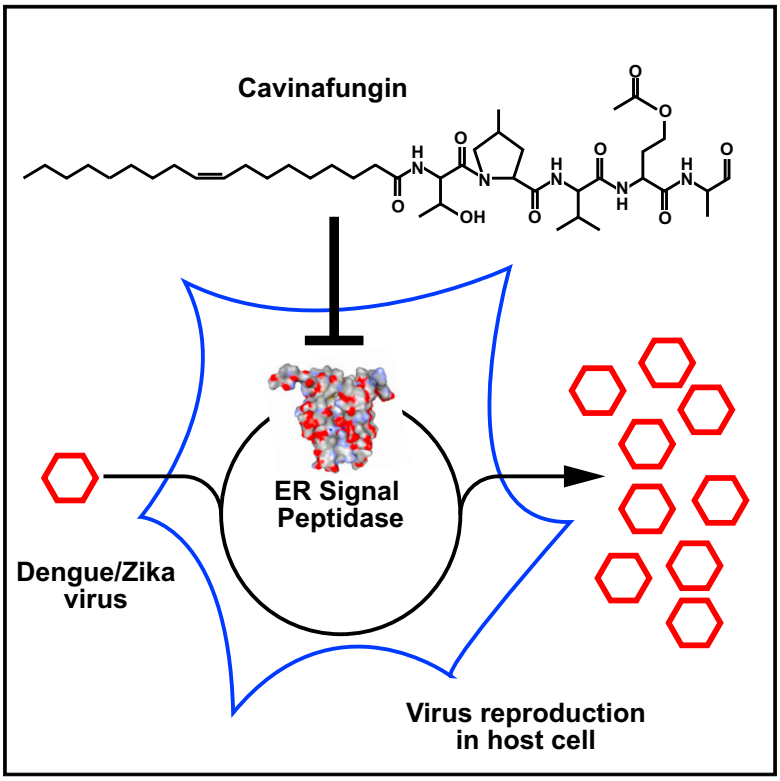
The results were published 2017 in *Cell reports*. The following chapter contains the publication, with the parts done in our lab highlighted in yellow.

Report

Cell Reports

The Natural Product Cavinafungin Selectively Interferes with Zika and Dengue Virus Replication by Inhibition of the Host Signal Peptidase

Graphical Abstract



Authors

David Estoppey, Chia Min Lee, Marco Janoschke, ..., Tewis Bouwmeester, Ghislain M.C. Bonamy, Dominic Hoepfner

Correspondence

bonamgh1@gmail.com (G.M.C.B.), dominic.hoepfner@novartis.com (D.H.)

In Brief

Recent outbreaks and lack of effective treatments against dengue and Zika virus have caused public concerns. Estoppey et al. have identified cavinafungin as exerting potent and selective antiviral activity by targeting the signal-binding cleft of the catalytic subunit of the endoplasmic reticulum signal peptidase.

Highlights

- Cavinafungin has potent and selective antiviral activity
- CRISPR/Cas9 profiling identifies the ER signal peptidase as the target of cavinafungin
- Resistance profiling maps the putative binding site of cavinafungin on SEC11
- Biochemical assays validate cavinafungin inhibition of viral and host protein processing

Estoppey et al., 2017, Cell Reports *19*, 451–460
April 18, 2017
http://dx.doi.org/10.1016/j.celrep.2017.03.071

CellPress

Cell Reports
Report

The Natural Product Cavinafungin Selectively Interferes with Zika and Dengue Virus Replication by Inhibition of the Host Signal Peptidase

David Estoppey,[1,5] Chia Min Lee,[2,5] Marco Janoschke,[3] Boon Heng Lee,[2] Kah Fei Wan,[2] Hongping Dong,[2] Philippe Mathys,[1] Ireos Filipuzzi,[1] Tim Schuhmann,[1] Ralph Riedl,[1] Thomas Aust,[1] Olaf Galuba,[1] Gregory McAllister,[4] Carsten Russ,[4] Martin Spiess,[3] Tewis Bouwmeester,[1] Ghislain M.C. Bonamy,[2,*] and Dominic Hoepfner[1,3,6,*]
[1]Novartis Institutes for BioMedical Research, Novartis Pharma AG, Forum 1 Novartis Campus, 4056 Basel, Switzerland
[2]Novartis Institute for Tropical Diseases, 10 Biopolis Road, Chromos #05-01, 138670 Singapore, Singapore
[3]Biozentrum, University of Basel, Klingelbergstrasse 50/70, 4056 Basel, Switzerland
[4]Novartis Institutes for BioMedical Research, 250 Massachusetts Avenue, Cambridge, MA 02139, USA
[5]These authors contributed equally
[6]Lead Contact
*Correspondence: bonamgh1@gmail.com (G.M.C.B.), dominic.hoepfner@novartis.com (D.H.)
http://dx.doi.org/10.1016/j.celrep.2017.03.071

SUMMARY

Flavivirus infections by Zika and dengue virus impose a significant global healthcare threat with no US Food and Drug Administration (FDA)-approved vaccination or specific antiviral treatment available. Here, we present the discovery of an anti-flaviviral natural product named cavinafungin. Cavinafungin is a potent and selectively active compound against Zika and all four dengue virus serotypes. Unbiased, genome-wide genomic profiling in human cells using a novel CRISPR/Cas9 protocol identified the endoplasmic-reticulum-localized signal peptidase as the efficacy target of cavinafungin. Orthogonal profiling in *S. cerevisiae* followed by the selection of resistant mutants pinpointed the catalytic subunit of the signal peptidase SEC11 as the evolutionary conserved target. Biochemical analysis confirmed a rapid block of signal sequence cleavage of both host and viral proteins by cavinafungin. This study provides an effective compound against the eukaryotic signal peptidase and independent confirmation of the recently identified critical role of the signal peptidase in the replicative cycle of flaviviruses.

INTRODUCTION

Flavivirus infections, including Zika and dengue virus infections, are mosquito-borne viral diseases with more than 100 million infections per year. Geographic distribution of infections is focused in the tropics and subtropics, a region covering 128 countries and inhabited by more than 3.9 billion people (Bhatt et al., 2013; Brady et al., 2012). The recent reports of Zika virus outbreaks and suspicion of detrimental effects on embryonic development, in particular cerebral development (Cugola et al., 2016), raise awareness of the fact that to date, there is no US Food and Drug Administration (FDA)-approved vaccination or specific treatment available.

In the search for novel antiviral intervention points, there have been three recent independent reports where genome-wide genetic screens investigated host factors that are required for efficient replication of the dengue virus (Ma et al., 2015; Marceau et al., 2016; Savidis et al., 2016; Zhang et al., 2016). They all identified endoplasmic reticulum (ER)-associated functions, including protein processing, maturation, and modification as essential host factors for the viral life cycle. In particular, the study by Zhang and colleagues outlined the importance of the host ER-associated signal peptidase (SPase) for processing of the prM and E structural proteins common to all flaviviruses (Zhang et al., 2016). However, whereas the signal peptidase is an established target for antibacterials with several reported inhibitor classes, to the best of our knowledge, no cellular active and selective eukaryotic signal peptidase has been reported to date.

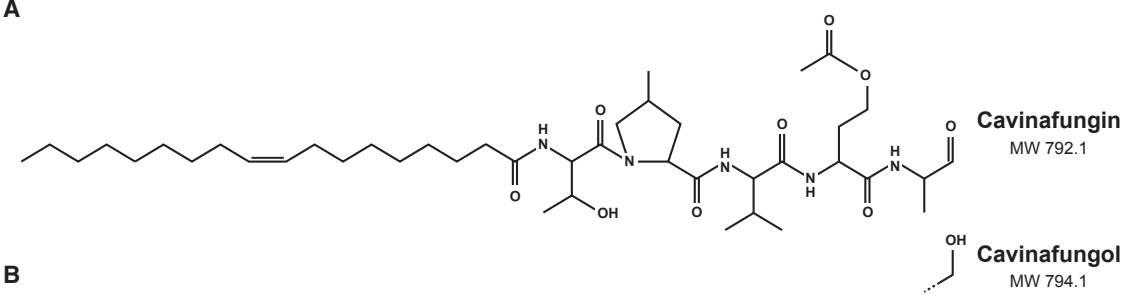
Here, we report that a recently identified alaninal-containing lipopeptide of fungal origin, cavinafungin (Ortíz-López et al., 2015), potently inhibits growth of dengue- and Zika-virus-infected cells with significant selectivity over non-infected cells. We established an unbiased genome-wide CRISPR/Cas9 chemogenomic profiling approach and identified the ER-associated SPase as the likely efficacy target of cavinafungin. This was verified by haploinsufficiency profiling in yeast, raising of resistant mutants, biochemical testing, and direct analysis of signal sequence processing in mammalian cells. This study provides the first compound against the eukaryotic signal peptidase and pharmacological confirmation of the critical role of the SPase in the replicative cycle of flaviviruses.

RESULTS

Cavinafungin Inhibits Zika- and Dengue-Virus-Infected Cells

To identify novel therapeutic leads, we tested compounds of various origins in a dose-response fashion for inhibitory activity against dengue- and Zika-virus-infected cells. The recently

A



B

	Virus	Cell Line	IC_{50} / CV (nM) / (%)	CC_{50} / CV (nM) / (%)
Cavinafungin	DV1	A549	4 / 57	800 / 36
	DV2	A549	5 / 71	800 / 36
	DV2*	HUH7	4 / 49	2'000 / n.d.
	DV2°	PBMC	1 / n.d.	18'000 / n.d.
	DV3	A549	3 / 22	800 / 36
	DV4	A549	3 / 27	800 / 36
	ZV	A549	150 / 23	1'650 / 49
	CHIKV	HUH7	900 /24	2'000 / n.d.

C

	Virus	Cell Line	IC_{50} / CV (nM) / (%)	CC_{50} / CV (nM) /(%)
Cavinafungol	DV1	A549	9'000 / 37	30'000 / 48
	DV2	A549	1'000 / 110	30'000 / 48
	DV2*	HUH7	3'000 / 78	50'000 / 39
	DV3	A549	2'000 / 69	30'000 / 48
	DV4	A549¨	1'000 / 95	30'000 / 48
	CHIKV	HUH7	40'000 / 18	50'000 / 39

D

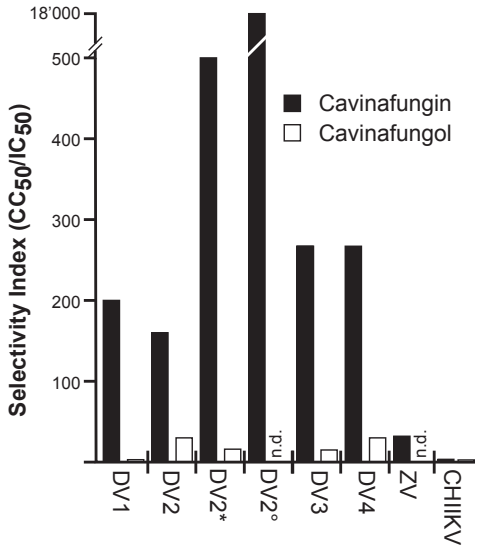


Figure 1. Cavinafungin Is a Potent and Selective Inhibitor of Cells Infected by Dengue Virus

(A) Chemical structure of cavinafungin and its inactive alcohol derivative.

(B) Inhibition and cytotoxicity concentrations of cavinafungin and cavinafungol on different dengue virus serotypes (DV), Zika virus (ZV), and chikungunya virus (CHIKV) in different cell lines. Note that chikungunya is not a flavivirus family member, and cavinafungin did not score significant antiviral activity against this virus.

(C) Inhibition and cytotoxicity concentrations of cavinafungol on different dengue virus serotypes in different cell lines. For those values with a CV indicated, at least two independent biological replicates were measured.

(D) Selectivity index as calculated by the log2 ratio of antiviral activity over cytotoxicity.

IC, inhibitory concentration; CC, cytotoxicity concentration; CV, coefficient of variation; n.d., not determined.

identified linear lipopeptide cavinafungin isolated from the fungus *Colispora cavincola* (Ortíz-López et al., 2015) (Figure 1A) displayed potent cellular activity against all four dengue serotypes with more than 100-fold selectivity over uninfected cells (Figure 1B). This activity was consistent across multiple cell types, in particular in human peripheral blood mononuclear cells (hPBMCs), which contain monocytes, thought to be one of the clinically relevant cells for dengue (Jessie et al., 2004). Expanding to other flaviviruses also revealed potent activity against Zika virus, with ~30-fold selectivity compared to the cytotoxic effect of the compound (Figure 1B). Cavinafungin was however inactive against other types of viruses, like chikungunya virus (Figure 1B), cytomegalovirus, and herpes simplex virus (data not shown). Cavinafungol, the reduced alcohol derivative of the cavinafungin aldehyde, revealed significantly reduced activity and a collapse of the therapeutic index, indicating that the aldehyde moiety was important for antiviral activity (Figures 1C and 1D).

Chemogenomic Profiling Using CRISPR/Cas9 Identifies Subunits of the Signal Peptidase

Inhibition of viral components or host cell processes can account for the observed antiviral activity. To identify a possible host target, we profiled cavinafungin in a genome-wide compound/CRISPR experiment (Figures 2A and 2B). We hypothesized that mutations in the target and pathway directly affected by cavinafungin would result in selective hypersensitivity and fitness defects upon exposure to sublethal doses of cavinafungin as recently shown for an inhibitor of NAMPT (Estoppey et al., 2017). Plotting the redundant siRNA activity (RSA) p value (a gene-level measure for conserved depletion of its respective guides) against Q1 (a gene-level effect size corresponding to the RSA p value) (König et al., 2007) identified only few genes that separated from the pool (Figure 2C). These hits were conserved across both experiments as visible when plotting them against each other (Figure 2D). Among the hits, we

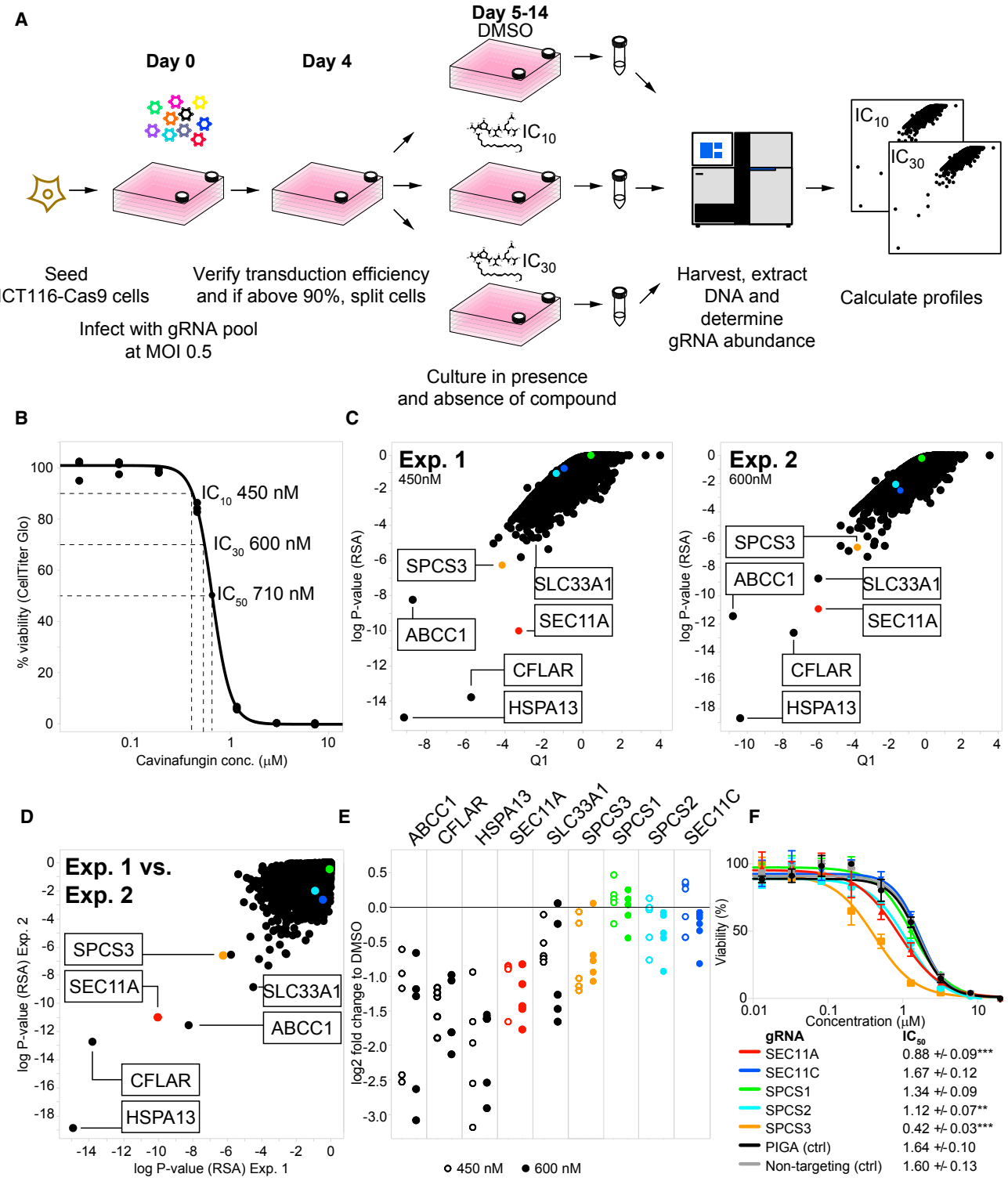


Figure 2. Genome-wide CRISPR/Cas9 Profiling of Cavinafungin in HCT116 Cells Identifies Two Subunits of Signal Peptidase

(A) Schematic representation of the experimental setup.

(B) Potency determination of cavinafungin against HCT116 cells by dose-response testing using two biological replicates.

(C) Profiles of genome-wide CRISPR/Cas9 profiling experiment of cavinafungin at IC_{10} and IC_{30} concentrations, respectively. Graphs show relative abundance of gRNAs against DMSO control, averaged to gene-level resolution as calculated by logP RSA (a gene-level measure for conserved depletion of its respective

(legend continued on next page)

identified CFLAR, a regulator of apoptosis, and ABCC1, encoding an ABC transporter involved in pleiotropic drug resistance (Hipfner et al., 1999), which are likely to affect cavinafungin effects indirectly, and SLC33A1, which is annotated as a probable acetyl-coenzyme A (acetyl-CoA) transporter. Noticeably, three additional hits act in a pathway recently reported to be required for efficient viral replication (Zhang et al., 2016): HSP13A, an HSP70 family chaperone involved in folding of secretory proteins (Yang et al., 2015), and SEC11A and SPCS3, encoding essential functions for the catalytic activity of the ER-associated SPase complex (Böhni et al., 1988; Fang et al., 1997; Shelness et al., 1993). The SPase complex also contains two additional regulatory subunits, SPCS1 and SPCS2, and a presumed paralog of SEC11A, SEC11C (Mullins et al., 1996). Guide RNAs (gRNAs) for these subunits did not separate from the pool, but analyzing the results of each of the five individual gRNAs showed dose-dependent cavinafungin effects on all five gRNAs (Figure 2E), suggesting at least some moderate effect. Previous expression analysis in HCT116 cells revealed significantly higher expression of SEC11A than of the presumed paralog SEC11C, thus likely accounting for the predominant dependency on SEC11A (Junne et al., 2015). Testing the best-scoring gRNAs individually in cavinafungin dose-response experiments supported significant hypersensitivity of cells edited for SEC11A, SPCS3, and to a lesser extent SPCS2 but no significant effects on the other SPase subunits (Figure 2F).

Chemogenomic Profiling in Budding Yeast Supports SPase as Cavinafungin Site of Action

Since the SPase complex is highly conserved in eukaryotic cells and cavinafungin was reported to also have antifungal activity (Ortíz-López et al., 2015) (Dataset S1), we decided to intersect the mammalian data with another unbiased, genome-wide approach: haplosinsufficiency profiling (HIP) and homozygous profiling (HOP) in fungal *S. cerevisiae* cells (Giaever et al., 1999; Hoon et al., 2008; Lee et al., 2014). HIP identifies proteins and/or pathways directly affected by the compound, whereas HOP can reveal synthetic lethal effects and delineate compensating factors and/or pathways. The results are visualized by plotting the relative growth reduction of individual HIP/HOP strains by the compound (sensitivity) versus a measure of significance (*Z* score) (Hoepfner et al., 2014). We tested cavinafungin at 30% inhibition concentration (IC_{30}) in two independent experiments (Figures 3A, 3B, S1A, and S1B). Interestingly, cavinafungin scored IC_{50} values in a range between 0.5 and 3.5 μM (Figures 3D and S1C), similar to the ranges observed for mammalian cells (Figures 1B and 2B). In both unbiased, genome-wide HIP experiments, strains carrying deletions in *SEC11* and *SPC2* encoding for the catalytic and a regulatory SPase subunit scored as the top hits. Additional gene hits included *SPC3*, another SPase subunit essential for catalytic activity, *SEC21*, encoding a component of the coat protein complex I (COPI) coat required for trafficking from ER to Golgi (Gaynor and Emr, 1997), and *ERD2*, coding for the His-Asp-Glu-Leu c-terminal tetrapeptdide motif (HDEL) receptor mediating retention/retrieval of ER proteins (Gaynor and Emr, 1997; Semenza et al., 1990). HOP profiling identified the nonessential *SPC2* SPase subunit gene as well as *HAC1* and *IRE1*, which are involved in the ER stress response, and *SIL1*, which is associated with an ER Hsp70 chaperone, and all are involved in ER quality control (Figure 3B). All hits were validated in single-strain dose-response experiments (Figure S1C). Hypersensitivity of the *SPC2* knockout (KO) strain in the HOP experiment ruled out this component as the primary pharmacological target of cavinafungin, pin-pointing the catalytic SPase subunit Sec11 as the likely candidate. Conserved action against the fungal and mammalian catalytic SPase subunit was assessed by replacing the *Saccharomyces cerevisiae* (*S.c*). *SEC11* gene with the mammalian SEC11A cDNA. The resulting yeast strain was viable, demonstrating functional complementation of the human SEC11A protein, and ~10-fold more susceptible to cavinafungin, with an IC_{50} almost identical to that measured in mammalian cells (Figure S2A). Cavinafungol did not exert any growth inhibitory effect against both strains up to the highest tested dose (Figure S2B).

Mutations in the Catalytic SPase Subunit Cause Cavinafungin Resistance

To corroborate the SPase target hypothesis, we deployed a focused mutagenesis approach. The cDNAs for *S.c. SEC11*, *SPC1*, *SPC2*, and *SPC3* were amplified by error-prone PCR, cloned into low-copy-number plasmids under the control of the endogenous promoter and terminator elements, and transformed into yeast cells heterozygous for the corresponding gene (Huang et al., 2013). 10^7 plasmid-harboring cells per gene were plated onto growth-inhibitory concentrations of cavinafungin. No resistant colonies could be isolated from the *SPC1*, *SPC2,* and *SPC3* plates. In contrast, 28 colonies conferring resistance to high doses of cavinafungin were isolated from the *SEC11* plates, and nine distinct mutations in *SEC11* were identified (Figure 3C). To validate genotype-phenotype linkage, these mutations were introduced into wild-type cells and subjected to dose-response testing. All mutations resulted in notable resistance up to the highest dose tested (100 μM), but to different degrees (Figure 3D). Resistance to cavinafungin was dominant over the wild-type copy of SEC11 still present in the cells. The majority of mutations (seven out of nine) localized to or near a triple valine motif in the peptidase domain around position 38–40, one to a triple valine motif at position 71, and one further downstream in the peptidase domain. To investigate if these sites outline a distinct structural motif, we used the published crystal structure of *S. pyogenes* SPase, the closest

guides) and Q1 (a gene-level effect size corresponding to the RSA p value) (König et al., 2007). Signal peptidase subunits are labeled by the colors referenced in (F).

(D) Alignment of the logP values of both experiments illustrate reproducibility and robustness.

(E) Log_2-fold changes of individual gRNAs of selected hits from the two genome-wide profiling experiments.

(F) Single gRNA validation of SPase subunits. The best-scoring gRNA for each gene identified in the genome-wide experiment was individually tested. Curves were fitted on data from three biological replicates. The number of asterisks indicates the number of SDs separating the curve from that of the non-coding gRNA experiment. A gRNA against the SPase unrelated ER protein PIGA was used as a second control.

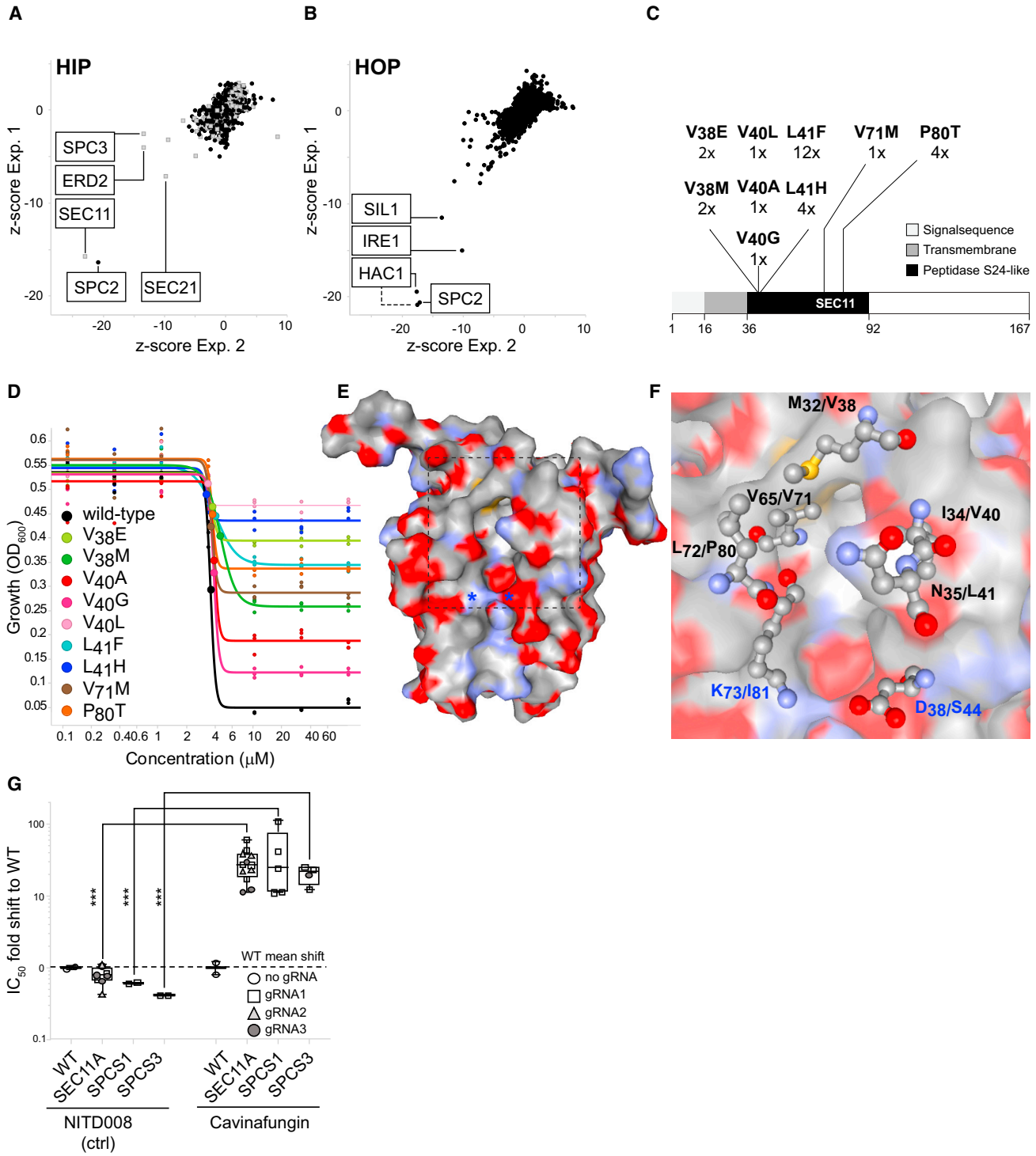


Figure 3. Chemogenomic Profiling of Cavinafungin in *S. cerevisiae* Supports the Catalytic SPase Subunit Sec11 as a Primary Target

(A) Alignment of two independent cavinafungin haploinsufficiency profiling (HIP) experiments. Three out of four components of the SPase complex (SEC11, SPC2, and SPC3) and key components in the secretory pathway (SEC21 and ERD2) reproducibly scored as hits.

(B) Alignment of two independent homozygous profiling (HOP) experiments reproducibly identified the ER stress response components HAC1/IRE1/SIL1 to be synthetic lethal with cavinafungin action, supporting disruption of proteostasis at the level of the ER. Deletions in essential genes are labeled by gray boxes and deletions in nonessential genes by black dots. The dotted line to HAC1 represents the YFL032W deletion strain, an open reading frame not yet identified to encode a protein but overlapping the HAC1 gene.

(legend continued on next page)

yeast and mammalian SPase homolog for which a structure was available (Young et al., 2014). Mapping the yeast mutations into the 4N31 crystal structure of the related SPase of *S. pyogenes* (Young et al., 2014) outlined the signal-peptide-binding cleft adjacent to the protease active site (Figures 3E and 3F; Dataset S1). We hypothesized that the identified mutations interfere with cavinafungin binding into the peptide-binding cleft.

Cavinafungin Shows Differential Inhibition in SPase KO Cells

To verify that Cavinagfungin was exerting its efficacy on dengue through the inhibition of the host SPase, we evaluated whether the compound anti-dengue activity was affected by editing of SPase components via CRISPR/Cas9. We evaluated the IC_{50} of Cavinagfungin in cells upon transient editing of SEC11A, SPCS1, and SPCS3 using up to three different gRNAs for each gene. NITD008, a nucleoside shown to directly inhibit the RNA-dependent RNA polymerase activity of dengue virus and thus a mechanism independent of SPase, was used as control (Yin et al., 2009).

Cells where SPase components were edited showed a consistent >10-fold decrease of cavinafungin in vitro efficacy compared to untreated wild-type control (Figure 3G). Conversely, the efficacy of the nucleoside NITD008 remained unaffected (<2-fold) in cells treated with the same gRNAs (Figure 3G). This lack of effect from a genetic KO of the various components of SPase on the efficacy of NITD008 is expected considering that the compound inhibits the viral polymerase directly. Together, the loss of efficacy measured by an increase in the IC_{50} fold change was significantly different ($p < 0.0001$) for cavinafungin compared with NITD008 in cells where SEC11A, SPCS1, or SPCS3 was targeted by CRISPR/Cas9.

Cavinafungin Inhibits Viral Polyprotein Processing and Signal Peptide Cleavage of Host Proteins

The dengue virus encodes a single polyprotein that has to be processed co- and post-translationally by viral and host proteases to yield three structural proteins (Figure 4A) (Lindenbach et al., 2007). Four cleavages (prM, E, NS1, and NS4B) are performed by host SPase. Thus, SEC11A inhibition could very well explain the observed antiviral activity of cavinafungin and is supported by recent genetic experiments (Zhang et al., 2016). In order to study the effect of cavinafungin on SPase cleavage of the viral precursor protein, we transfected cells with constructs C_{14}-prM-E and 2K-NS4B-V5 (Figure 4A). Western blot analysis revealed that 1 μM cavinafungin strongly reduced the levels of (C_{14}-)prM and E proteins released from the C_{14}-prM-E construct and enhanced the level of the full-size precursor (C_{14}-)prM-E (FL; recognized on western blots only by the anti-E antibody) (Figure 4B, left), indicating inhibition of SPase cleavage at site 2. No effect was detected with 1 μM cavinafungol. Inhibition by cavinafungin was dose dependent, with a weighted mean IC_{50} of 0.73 ± 0.08 μM (Figure 4B, right; Table S1). Similarly, upon expression of 2K-NS4B-V5, the presence of 1 μM cavinafungin, but not cavinafungol, induced the vast majority of products to shift to a higher molecular weight corresponding to the uncleaved 2K-NS4B-V5 (Figure 4C, left). Accumulation of the uncleaved protein was dose dependent, with a weighted IC_{50} mean of ~0.08 ± 0.01 μM (Figure 4C, right; Table S1).

We also analyzed inhibition of SPase cleavage by metabolic labeling to detect the immediate translation products rather than steady-state levels. Available antibodies allowed us to immunoprecipitate and distinguish the precursor (C_{14}-)prM-E (FL) and the products of cleavage at site 2 alone, C_{14}-prM, or at both sites 1 and 2, prM (Figure 4D). As summarized in Table S1, cavinafungin added during the labeling period inhibited cleavage, with a weighted IC_{50} mean of 0.49 ± 0.11 μM at site 1 and 0.17 ± 0.01 μM at site 2. Expression of 2K-NS4B-V5 (Figure 4E) revealed a considerable amount of uncleaved product, even in the absence of inhibitor, most likely the result of inefficient ER targeting of the viral protein fragment and not detectable in the western blot because of rapid degradation. Cavinafungin inhibited cleavage at site 4, with an IC_{50} of 1.8 ± 0.01 μM. Taken together, the viral SPase cleavage sites were inhibited with similar efficiencies by cavinafungin, with a weighted IC_{50} mean of 0.87 ± 0.01 μM.

SPase inhibition is expected to also compromise processing of host proteins. We assessed cleavage of two host secretory proteins at increasing cavinafungin concentrations, pre-provasopressin and pre-proinsulin, both of which are known to produce a detectable shift in electrophoretic mobility when the signal peptide is retained (Beuret et al., 1999; Cui et al., 2015). Transfected HeLa cells producing pre-provasopressin and INS-1E insulinoma cells endogenously expressing pre-proinsulin were metabolically labeled, immunoprecipitated, and analyzed by SDS-gel electrophoresis and autoradiography (Figures 4F and 4G). Cavinafungin caused accumulation of the larger, uncleaved protein forms, with an IC_{50} of ~0.91 ± 0.15 μM for pre-provasopressin and 3.29 ± 0.3 μM for pre-proinsulin. As summarized in Table S1, cavinafungin inhibits cleavage at the viral and host cleavage sites at similar concentrations and close to the IC_{50} values measured for growth and cytotoxicity of different cell lines (Figure 1B). The alcohol form of cavinafungin, in contrast, only produced minimal signal cleavage inhibition at much higher concentrations, in agreement with the observed minimal inhibition of cell growth (Figure 1C). These experiments

(C) Induced mutagenesis followed by selection on growth-inhibitory concentrations of cavinafungin identified mutations in the peptidase domain of the SEC11 gene that confer resistance. Position, amino acid changes, and frequencies are indicated.
(D) Validation of resistance by cloning of the identified mutations into fresh wild-type strains and measuring dose-response curves of three technical replicates.
(E) Structure 4N31 of *S. pyogenes* SipA, the closest yeast/human SPase homolog with a solved crystal structure. The proposed active site is labeled by blue asterisks.
(F) The homologous residues identified to give cavinafungin resistance are labeled in black, and the catalytic dyad residues of the protease are depicted in blue. The first letter depicts the *Streptococcus pyogenes* (*S.p.*). SipA residue and the second the corresponding *S.c.* Sec11p residue. The dotted box delimits the area enlarged in (E).
(G) Effect of knocking out components of the SPase complex on the in vitro efficacy of cavinafungin or on the RNA-dependent polymerase inhibitor NITD008. The different shapes represent different gRNAs used per gene. ***$p < 0.0001$ using a two-way multiple comparisons ANOVA with Sidak corrections.

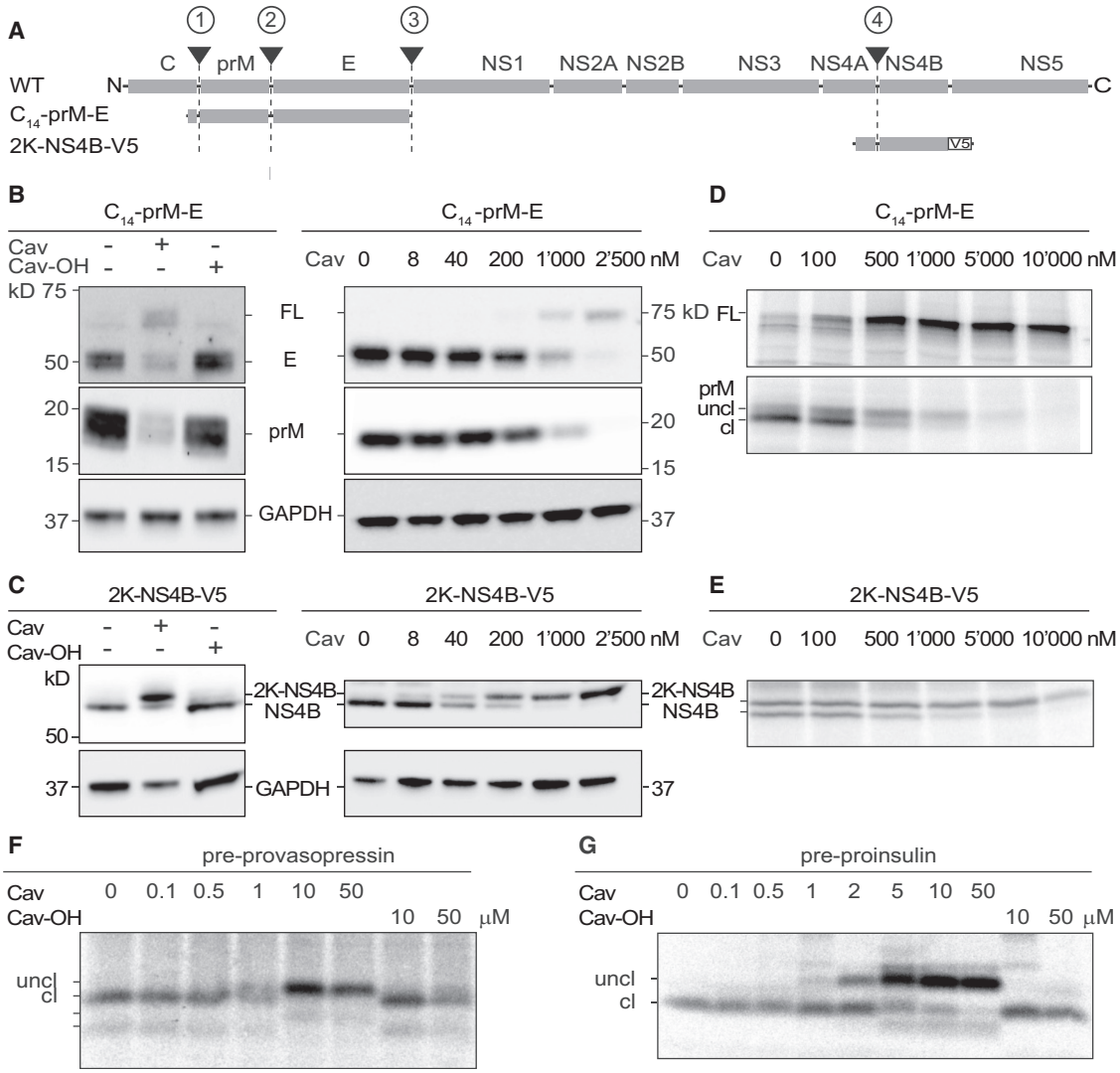


Figure 4. Biochemical Confirmation of Signal Peptide Cleavage Inhibition by Cavinafungin of Viral and Human Substrates

(A) Schematic representation of the dengue polyprotein and of constructs used to investigate inhibition of processing by cavinafungin (Cav) and cavinafungol (Cav-OH). SPase cleavage sites are indicated by black triangles and numbered.

(B and C) Effect of Cav and Cav-OH on HEK293T cells transfected with C_{14}-prM-E (B) or 2K-NS4B-V5 (C) as identified by western blot. The blots represent one out of two to four independent experiments. 17- to 19-kDa bands are indicated as various forms of prM, and E and uncleaved prM-E (FL) are detected as 50- and 70-kDa bands, respectively. GAPDH was probed as control.

(D and E) Effect of Cav on cells transfected with C_{14}-prM-E (D) or 2K-NS4B-V5 (E) as determined by labeling experiments.

(F and G) Effect of Cav and Cav-OH on human pre-provasopressin (F) and INS-1E insulinoma cells endogenously expressing pre-proinsulin (G) as determined by labeling experiments.

provide biochemical evidence for inhibition of the SPase activity by cavinafungin as predicted by the chemogenomic profiling experiments.

DISCUSSION

In phenotypic screens against Zika and dengue virus, we have discovered the recently identified natural methylproline-containing lipopeptide compound cavinafungin (Ortíz-López et al., 2015) to be a selective antiviral agent. Using a suite of genetic and finally proteomic and biochemical assays, we have identified cavinafungin to be a selective, cellular active inhibitor of the ER SPase. Although there are a variety of compounds in the literature that are reported to target the bacterial enzyme (Craney and Romesberg, 2015), to the best of our knowledge, this is the first identification of a selective eukaryotic SPase inhibitor.

SPase is an essential, membrane-bound serine protease complex involved in cleavage of the signal peptides of newly synthesized secretory and membrane proteins at the ER. SEC11A and SPCS3 are necessary for the cleavage activity (Fang et al., 1997; Meyer and Hartmann, 1997), but only SEC11A so far was

identified to contain a signal-sequence-binding domain and classic serine protease dyad motif (Fang et al., 1997). To date, no crystal structure of any eukaryotic SPase subunit is available that would allow for predictive cavinafungin in silico docking approaches. However, it is plausible that membrane anchoring of cavinafungin via its lipophilic tail orients the oligopeptide moiety to be recognized as a signal peptide in the SEC11A peptide-binding cleft and engaging the catalytic serine of the protease by the reactive aldehyde via a nucleophilic attack. This results in formation of a covalent hemiacetal intermediate, the classic serine protease inhibition mechanism (as reviewed by Hedstrom, 2002). This hypothesis is supported by several lines of experimental evidence presented in this report: (1) the aldehyde group of cavinafungin is critical for activity (Ortíz-López et al., 2015), (2) the signal-peptide-binding subunit SEC11A was the only one for which resistance conferring mutations could be isolated, (3) the identified resistance-conferring residues outline the signal peptide binding cleft, and (4) cavinafungin has a peptidic moiety that fulfills the −1 and −3 rules (Ala and Val) of a signal peptide (von Heijne, 1990)

In addition to reporting the first mammalian SPase inhibitor, we also demonstrate that cleavage by SPase is critical for flavivirus polyprotein processing. In particular, we provide support that the inhibition of dengue by cavinafungin is directly mediated via inhibition of the SPase by showing that the compound loses in vitro efficacy in cells where the SPase complex is genetically targeted by CRISPR/Cas9. We also we provide experimental support that the 2K domain upstream of NS4B and the C_{14} domain upstream of the prM protein, two transmembrane domains, are cavinafungin sensitive and thus likely processed with involvement of the SPase, as previously suggested (Lindenbach et al., 2007).

The prM-E junction shows the highest degree of inhibition by cavinafungin, with an IC_{50} of 0.19 ± 0.01 μM, compared to the inhibition of pre-proinsulin or the cytotoxicity measured of ∼1.34 ± 0.18 μM, which corresponds to a selectivity of ∼7-fold. The effect of prM-E would therefore not explain on its own the selectivity observed. We attribute the hypersensitivity of infected cells to the fact that virus production and virion assembly require four SPase cleavages (host proteins only one site), that partially cleaved fragments are unstable, as observed for unprocessed C14-prM-E (Figure 4B), and that non-stoichiometric amounts of viral proteins might exert a dominant-negative effect on viral replication and assembly; for example, prM acts as a folding chaperone for E (Roby et al., 2015). Folding stress is likely potentiated by the sheer load of virus production, far exceeding the normal volume of endogenous secretory load in uninfected cells.

Based on the tight homology of the sequences across multiple viruses in this family (Figure S3), the activity of cavinafungin is likely to impact other members of the flavivirus family. The sequence alignment shows that these regions are highly conserved (Figure S3), particularly the small, uncharged residues required at −1 and −3 sites and the upstream hydrophobic stretch that are predictors for SPase cleavage (von Heijne, 1990). This notion is indeed supported by our data showing that Zika virus replication is also inhibited by cavinafungin, thus matching previously reported effects on both dengue and Zika virus replication when SPase was compromised by genetic means (Zhang et al., 2016).

Our study demonstrates the power of unbiased, genome-wide CRISPR/Cas9 compound profiling in mammalian cells for identifying the mechanism of action of compounds. In particular, the intersection with chemogenomic profiling data from yeast and genetic follow up experiments in this organism allowed conclusive target identification of the first cellular active, selective, eukaryotic SPase inhibitor. This chemical biology dataset corroborates recent genetic findings that highlight the host SPase complex as an Achilles heel of flaviviruses (Zhang et al., 2016) and provides a research compound for further dissection of the viral life cycle, in particular to gain a better understanding of polyprotein processing. Finally, the observed differential antiviral activity raises the hope that this node can be therapeutically exploited for urgently needed, novel antiviral therapies.

EXPERIMENTAL PROCEDURES

Compound Isolation and Handling

Cavinafungin was isolated from cultures of an unclassified fungal strain and purified according to a previously described method (Ortíz-López et al., 2015). Compound structure was confirmed by high-resolution mass spectroscopy and nuclear magnetic resonance (NMR). Sample purity was found to be >90% according to high-performance liquid chromatography (HPLC) analysis. The compound was stored as powder at 4°C until use and dissolved in 100% DMSO to a stock concentration of 10 mM. Solutions were kept at 4°C for up to 6 months.

High-Content Imaging Cell-Based Flavivirus Immunodetection Assay

In brief, 7×10^3 A549 cells per well in a 384-well plate were infected at MOIs of 0.5, 0.3, 2, and 1 DENV-1(MY97-10245), DENV-2(MY97-10340), DENV-3(MY05-34640), and DENV-4(MY01-22713), respectively, in the presence of 3-fold 10-point serial dilutions of the compound. 8×10^3 HUH7 cells per well in a 384-well plate were infected at an MOI of 2 for DENV-2 (My97-10340) in the presence of 3-fold, 10-point serial dilutions of compounds. At 48 hr postinfection (p.i.), the infected wells were washed, fixed, and labeled with Dylight-488-conjugated 4G2 Antibody (Genescript). Cells were stained with Draq5 (Thermo Fisher Scientific) to determine toxicity and images were acquired by Opera imaging system (PerkinElmer). A dose-response curve was plotted using sigmoid regression analysis, and IC_{50} and CC_{50} were calculated.

Inhibition Shift Assay upon Genetic KO

HUH7 cells stably expressing Cas9 were dispensed into 384-wellplates at 4×10^3 cells/well and were treated with individual lentiviral packaged gRNAs at an M.O.I. of ∼5. Cells were maintained under puromycin selection pressure at 2μg/ml for 5 days. After this initial CRISPR/Cas9 mediated knock out, cells were infected with DENV2 (MOI2) in the presence serial dilutions of compounds. 48 hr later plates were processed and IC_{50} calculated as described above in in the high-content imaging cell-based flavivirus immunodetection (HCI CFI) assay. IC_{50} fold-shift for each drug/gRNA combination was calculated by dividing mean IC_{50} gRNA-treated with wild-type, mock-treated cells. A two-way ANOVA allowing for multiple comparisons with a Sidak correction was performed between NITD008 and cavinafungin for each gRNA-treated sample using the logged IC_{50} fold-shift obtained in GraphPad Prism v7.0.

Dengue Viral Titer Reduction Assay

5×10^5 cells per well of hPBMCs were seeded into 96-well plates, followed by treatment with 4G2 antibody. The antibody-enhanced hPBMCs were infected at an MOI 1 of DENV-2 (MY97-10340), and the infected cells were immediately treated with 5-fold serial dilutions of the compounds. 1×10^5 cells per well hHepIPS were seeded into 96-well plates and infected at an MOI 4 DENV-2

(MY97-10340) 5 days post-seeding. The infected cells were then treated with 3-fold serial dilutions of the compound. Culture fluids from DENV infections were collected at 48 hr p.i. The viral titers in the collected samples were quantified using a plaque assay.

Testing Other Viruses

Zika virus testing was performed at University of Texas Medical Branch (UTMB) by the team of Pei-Yong Shi using a live reporter virus as reported previously (Shan et al., 2016). Chikungunya testing was performed using a live GFP-expressing infectious clone, LR-3′-GFP-CHIKV, obtained from EVA (European Virus Archive) as previously described (Tsetsarkin et al., 2006).

Genome-wide CRISPR/Cas9 Profiling Experiment

A previously described protocol (Estoppey et al., 2017) was followed. Cavinafungin was tested at two sublethal doses, IC_{10} (450 nM) and IC_{30} (600 nM), based on the best results from related chemogenomic profiling experiments in yeast (Giaever et al., 1999; Hoepfner et al., 2014) and CRISPR pilot experiments with other compounds (Estoppey et al., 2017). Detailed protocols for genomic DNA extraction, sequencing of the gRNA, and processing of the data have been reported previously (Estoppey et al., 2017).

HIP HOP Profiling

Chemogenomic profiling and data processing of cavinafungin was conducted as described previously (Hoepfner et al., 2014).

Focused Mutagenesis Approach

The focused mutagenesis screen was performed as described previously (Huang et al., 2013), with slight modifications. In short, 10^7 yeast BY4743Δ8 cells (Hoepfner et al., 2012) carrying randomly mutagenized, plasmid-based *SEC11*, *SPC1*, *SPC2*, or *SPC3* variants, respectively, were plated on synthetic minimal SD medium complemented with 150 μM cavinafungin and incubated for 3 days at 30°C. Growing colonies were restreaked onto selective media. Stably resistant clones were then isolated and amplified, and DNA was isolated and subjected to Sanger sequencing.

Analysis of Viral and Host Protein Signal Cleavage

293T cells were transfected with pXJ constructs containing SPase processing sites of the dengue virus polyprotein, C_{14}-prM-E and 2K-NS4B-V5. C_{14} consists of the last 14 amino acids of the capsid, and 2K-NS4B-V5 represents the internal signal peptide C terminal of NS4A followed by NS4B with a V5 tag. At 6 hr post-transfection, the cells were treated with 1 μM cavinafungin, 1 μM cavinafungol, or DMSO. Cells were lysed 24 hr later using RIPA buffer, and the clarified cell lysates were analyzed by western blot using either polyclonal rabbit anti-prM (GeneTex), polyclonal rabbit anti-E (GeneTex), or monoclonal mouse anti-NS4B antibodies (in-house; Xie et al., 2014), respectively. The blots were also probed with polyclonal rabbit anti-GAPDH antibody (Sigma) as a loading control. The bands were detected with ECL Western Blotting System (GE Healthcare) and imaged with ChemiDoc western blot imaging system (Bio-Rad). Densitometry analysis was performed with Image Lab software (Bio-Rad), curves were fitted using the logistic regression curve fit function in GraphPad Prism version 7.00 (GraphPad Software), and weighted mean IC_{50} values and weighted errors were calculated as described previously (Jones et al., 2010).

For metabolic labeling, HeLa α cells were transiently transfected with C_{14}-prM-E or 2K-NS4B-V5 in pXJ or with the cDNA of human pre-provasopressin in the expression plasmid pRc/RSV using Fugene HD (Promega) and used after 2 days. INS-1E insulinoma cells (Janjic et al., 1999) were grown in RPMI medium and in addition 10 mM HEPES, 1 mM sodium pyruvate, and 20 μM β-mercaptoethanol. Cells were starved for 30 min in methionine/cysteine-free medium and labeled for 40 min at 37°C in starvation medium with 100 μCi/ml [^{35}S]methionine/cysteine (Hartmann Analytics). Cavinafungin or the reduced alcohol form was added. Cells were washed at 4°C, lysed, and subjected to immunoprecipitation using a rabbit anti-provasopressin (Friberg et al., 2004), a monoclonal mouse anti-proinsulin (CCI-17; HyTest), a rabbit anti-prM (GeneTex), and an anti-V5 antibody (R960-25, Invitrogen), respectively, and protein A-Sepharose. For 2K-NS4B-V5, the immunoprecipitates were sometimes digested with endoglycosidase H before SDS-gel electrophoresis and autoradiography.

SEC11/SEC11A Complementation

The heterozygous *SEC11/sec11Δ::kanMX4* BY4743 strain from the official yeast deletion collection was transformed with an integrative plasmid containing a *HIS3* marker and the human SEC11A cDNA under control of the *S.c. ADH1* promoter. Successful transformants were subjected to sporulation and tetrad dissection. Spores were analyzed for growth on media containing G418 or lacking histidine. All G418-resistant spores also were prototroph for histidine. Two independent clones (1 Mat a and 1 Mat alpha) were picked and correct integration and genotype confirmed by analytical PCR.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, one table, and one dataset and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2017.03.071.

AUTHOR CONTRIBUTIONS

Conceptualization, D.H., M.S., and G.M.C.B.; Methodology, D.E., T.A., I.F., W.K.F., O.G., M.J., C.M.L, B.H.L., P.M., T.S., and R.R.; Formal Analysis, G.M., and C.R.; Writing – Original Draft, D.H., M.S. and G.M.C.B.; Supervision, M.S. and T.B.

ACKNOWLEDGMENTS

We thank the Novartis Natural Products unit for support in compound supply, Nicole Hartmann and Juerg Eichenberger for processing of the HIP HOP microarrays, Alicia Lindeman for sequencing support, and Zinger Yang for data analysis support. We thank Malini Varadarajan for generation of the HCT116-Cas9 line and Christian Studer for generation of the SEC11 replacement strains. We are grateful to Drs. Julia Birk (University of Basel), P. Maechler, and C. Wollheim (University of Geneva) for INS-1E cells. We thank C. Johns and N. Jarouse (NIBR, Emeryville) for the testing of cavinafungin on HSV, BKV, CMV, and HRV, as well as Pei-Yong Shi and his team (UTMB) for testing this compound on Zika virus. We are grateful to Siew Pheng Lim, Cheah Chen Seh, Yi Hua Chan, and Hao Ying Xu for performing the dengue virus assay on primary cells. We also would like to thank Xuping Xie and Jing Zou for the dengue pXJ constructs. Authors with affiliation Novartis Institutes for Biomedical Resarch or Novartis Institutes for Tropical Diseases are employees of Novartis and may own stock.

Received: September 22, 2016
Revised: March 6, 2017
Accepted: March 24, 2017
Published: April 18, 2017

REFERENCES

Beuret, N., Rutishauser, J., Bider, M.D., and Spiess, M. (1999). Mechanism of endoplasmic reticulum retention of mutant vasopressin precursor caused by a signal peptide truncation associated with diabetes insipidus. J. Biol. Chem. *274*, 18965–18972.

Bhatt, S., Gething, P.W., Brady, O.J., Messina, J.P., Farlow, A.W., Moyes, C.L., Drake, J.M., Brownstein, J.S., Hoen, A.G., Sankoh, O., et al. (2013). The global distribution and burden of dengue. Nature *496*, 504–507.

Böhni, P.C., Deshaies, R.J., and Schekman, R.W. (1988). SEC11 is required for signal peptide processing and yeast cell growth. J. Cell Biol. *106*, 1035–1042.

Brady, O.J., Gething, P.W., Bhatt, S., Messina, J.P., Brownstein, J.S., Hoen, A.G., Moyes, C.L., Farlow, A.W., Scott, T.W., and Hay, S.I. (2012). Refining the global spatial limits of dengue virus transmission by evidence-based consensus. PLoS Negl. Trop. Dis. *6*, e1760.

Craney, A., and Romesberg, F.E. (2015). The inhibition of type I bacterial signal peptidase: Biological consequences and therapeutic potential. Bioorg. Med. Chem. Lett. *25*, 4761–4766.

Cugola, F.R., Fernandes, I.R., Russo, F.B., Freitas, B.C., Dias, J.L., Guimarães, K.P., Benazzato, C., Almeida, N., Pignatari, G.C., Romero, S., et al. (2016). The Brazilian Zika virus strain causes birth defects in experimental models. Nature *534*, 267–271.

Cui, J., Chen, W., Sun, J., Guo, H., Madley, R., Xiong, Y., Pan, X., Wang, H., Tai, A.W., Weiss, M.A., et al. (2015). Competitive inhibition of the endoplasmic reticulum signal peptidase by non-cleavable mutant preprotein cargos. J. Biol. Chem. *290*, 28131–28140.

Estoppey, D., Hewett, J.W., Guy, C.T., Harrington, E., Thomas, J.R., Schirle, M., Cuttat, R., Waldt, A., Gerrits, B., Yang, Z., et al. (2017). Identification of a novel NAMPT inhibitor by CRISPR/Cas9 chemogenomic profiling in mammalian cells. Sci. Rep. *7*, 42728.

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.

Friberg, M.A., Spiess, M., and Rutishauser, J. (2004). Degradation of wild-type vasopressin precursor and pathogenic mutants by the proteasome. J. Biol. Chem. *279*, 19441–19447.

Gaynor, E.C., and Emr, S.D. (1997). COPI-independent anterograde transport: cargo-selective ER to Golgi protein transport in yeast COPI mutants. J. Cell Biol. *136*, 789–802.

Giaever, G., Shoemaker, D.D., Jones, T.W., Liang, H., Winzeler, E.A., Astromoff, A., and Davis, R.W. (1999). Genomic profiling of drug sensitivities via induced haploinsufficiency. Nat. Genet. *21*, 278–283.

Hedstrom, L. (2002). Serine protease mechanism and specificity. Chem. Rev. *102*, 4501–4524.

Hipfner, D.R., Deeley, R.G., and Cole, S.P. (1999). Structural, mechanistic and clinical aspects of MRP1. Biochim. Biophys. Acta *1461*, 359–376.

Hoepfner, D., McNamara, C.W., Lim, C.S., Studer, C., Riedl, R., Aust, T., McCormack, S.L., Plouffe, D.M., Meister, S., Schuierer, S., et al. (2012). Selective and specific inhibition of the plasmodium falciparum lysyl-tRNA synthetase by the fungal secondary metabolite cladosporin. Cell Host Microbe *11*, 654–663.

Hoepfner, D., Helliwell, S.B., Sadlish, H., Schuierer, S., Filipuzzi, I., Brachat, S., Bhullar, B., Plikat, U., Abraham, Y., Altorfer, M., et al. (2014). High-resolution chemical dissection of a model eukaryote reveals targets, pathways and gene functions. Microbiol. Res. *169*, 107–120.

Hoon, S., Smith, A.M., Wallace, I.M., Suresh, S., Miranda, M., Fung, E., Proctor, M., Shokat, K.M., Zhang, C., Davis, R.W., et al. (2008). An integrated platform of genomic assays reveals small-molecule bioactivities. Nat. Chem. Biol. *4*, 498–506.

Huang, Z., Chen, K., Zhang, J., Li, Y., Wang, H., Cui, D., Tang, J., Liu, Y., Shi, X., Li, W., et al. (2013). A functional variomics tool for discovering drug-resistance genes and drug targets. Cell Rep. *3*, 577–585.

Janjic, D., Maechler, P., Sekine, N., Bartley, C., Annen, A.S., and Wolheim, C.B. (1999). Free radical modulation of insulin release in INS-1 cells exposed to alloxan. Biochem. Pharmacol. *57*, 639–648.

Jessie, K., Fong, M.Y., Devi, S., Lam, S.K., and Wong, K.T. (2004). Localization of dengue virus in naturally infected human tissues, by immunohistochemistry and in situ hybridization. J. Infect. Dis. *189*, 1411–1418.

Jones, D.C., Hallyburton, I., Stojanovski, L., Read, K.D., Frearson, J.A., and Fairlamb, A.H. (2010). Identification of a κ-opioid agonist as a potent and selective lead for drug development against human African trypanosomiasis. Biochem. Pharmacol. *80*, 1478–1486.

Junne, T., Wong, J., Studer, C., Aust, T., Bauer, B.W., Beibel, M., Bhullar, B., Bruccoleri, R., Eichenberger, J., Estoppey, D., et al. (2015). Decatransin, a new natural product inhibiting protein translocation at the Sec61/SecYEG translocon. J. Cell Sci. *128*, 1217–1229.

König, R., Chiang, C.Y., Tu, B.P., Yan, S.F., DeJesus, P.D., Romero, A., Bergauer, T., Orth, A., Krueger, U., Zhou, Y., and Chanda, S.K. (2007). A probability-based approach for the analysis of large-scale RNAi screens. Nat. Methods *4*, 847–849.

Lee, A.Y., St Onge, R.P., Proctor, M.J., Wallace, I.M., Nile, A.H., Spagnuolo, P.A., Jitkova, Y., Gronda, M., Wu, Y., Kim, M.K., et al. (2014). Mapping the cellular response to small molecules using chemogenomic fitness signatures. Science *344*, 208–211.

Lindenbach, B.D., Thiel, H.J., and Rice, C.M. (2007). Flaviviridae: the viruses and their replication. In Lippincott, Williams & Wilkins, Fields Virology, D.M. Knipe, and P.M. Howley, eds., pp. 1101–1152.

Ma, H., Dang, Y., Wu, Y., Jia, G., Anaya, E., Zhang, J., Abraham, S., Choi, J.G., Shi, G., Qi, L., et al. (2015). A CRISPR-based screen identifies genes essential for West-Nile-virus-induced cell death. Cell Rep. *12*, 673–683.

Marceau, C.D., Puschnik, A.S., Majzoub, K., Ooi, Y.S., Brewer, S.M., Fuchs, G., Swaminathan, K., Mata, M.A., Elias, J.E., Sarnow, P., and Carette, J.E. (2016). Genetic dissection of Flaviviridae host factors through genome-scale CRISPR screens. Nature *535*, 159–163.

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.

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.

Ortíz-López, F.J., Monteiro, M.C., González-Menéndez, V., Tormo, J.R., Genilloud, O., Bills, G.F., Vicente, F., Zhang, C., Roemer, T., Singh, S.B., and Reyes, F. (2015). Cyclic colisporifungin and linear cavinafungins, antifungal lipopeptides isolated from Colispora cavincola. J. Nat. Prod. *78*, 468–475.

Roby, J.A., Setoh, Y.X., Hall, R.A., and Khromykh, A.A. (2015). Post-translational regulation and modifications of flavivirus structural proteins. J. Gen. Virol. *96*, 1551–1569.

Savidis, G., McDougall, W.M., Meraner, P., Perreira, J.M., Portmann, J.M., Trincucci, G., John, S.P., Aker, A.M., Renzette, N., Robbins, D.R., et al. (2016). Identification of Zika virus and dengue virus dependency factors using functional genomics. Cell Rep. *16*, 232–246.

Semenza, J.C., Hardwick, K.G., Dean, N., and Pelham, H.R. (1990). ERD2, a yeast gene required for the receptor-mediated retrieval of luminal ER proteins from the secretory pathway. Cell *61*, 1349–1357.

Shan, C., Xie, X., Muruato, A.E., Rossi, S.L., Roundy, C.M., Azar, S.R., Yang, Y., Tesh, R.B., Bourne, N., Barrett, A.D., et al. (2016). An infectious cDNA clone of Zika virus to study viral virulence, mosquito transmission, and antiviral inhibitors. Cell Host Microbe *19*, 891–900.

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

Tsetsarkin, K., Higgs, S., McGee, C.E., De Lamballerie, X., Charrel, R.N., and Vanlandingham, D.L. (2006). Infectious clones of Chikungunya virus (La Réunion isolate) for vector competence studies. Vector Borne Zoonotic Dis. *6*, 325–337.

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

Xie, X., Zou, J., Wang, Q.Y., Noble, C.G., Lescar, J., and Shi, P.Y. (2014). Generation and characterization of mouse monoclonal antibodies against NS4B protein of dengue virus. Virology *450-451*, 250–257.

Yang, Z., Zhuang, L., Szatmary, P., Wen, L., Sun, H., Lu, Y., Xu, Q., and Chen, X. (2015). Upregulation of heat shock proteins (HSPA12A, HSP90B1, HSPA4, HSPA5 and HSPA6) in tumour tissues is associated with poor outcomes from HBV-related early-stage hepatocellular carcinoma. Int. J. Med. Sci. *12*, 256–263.

Yin, Z., Chen, Y.L., Schul, W., Wang, Q.Y., Gu, F., Duraiswamy, J., Kondreddi, R.R., Niyomrattanakit, P., Lakshminarayana, S.B., Goh, A., et al. (2009). An adenosine nucleoside inhibitor of dengue virus. Proc. Natl. Acad. Sci. USA *106*, 20435–20439.

Young, P.G., Proft, T., Harris, P.W., Brimble, M.A., and Baker, E.N. (2014). Structure and activity of Streptococcus pyogenes SipA: a signal peptidase-like protein essential for pilus polymerisation. PLoS ONE *9*, e99135.

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.

7. Material and Methods

7.1. Chemicals

Table 4: Chemicals used in this study.

Name	Company	Solution/Concentration
BSA	Sigma-Aldrich	100 mg/ml in H_2O
LiAc	Sigma-Aldrich	Powder
NaN3	Applichem	ddH_2O (1M)
PEG	Sigma-Aldrich	Powder
PMSF	Applichem	Isopropanol (500 mM)
Salmon Sperm DNA	Sigma-Aldrich	10 mg/ml in H_2O
Triton	Sigma-Aldrich	100% solution
Carbenicillin	Applichem	50 mg/ml
[^{35}S]Methionine	Hartmann Analytics	Storage solution: 50 mM Tricine (pH 7.4), stabilized Specific activity: >37TBq (1000Ci)/mmol Concentration: 185Mbq(5mCi)/ml
Cycloheximide	Sigma-Aldrich	Powder

7.2. Antibodies

Table 5: Antibodies used in this publication

Name	**Supplier**	**Host**
α-HA	Biozentrum Basel Spiess Group	Mouse (IgG2b) monoclonal
α-V5	Thermo Fisher Scientific	Mouse (IgG2a) monoclonal
α-NPII	Biozentrum Basel Spiess Group	Rabbit polyclonal

α-prM	Genetex	Rabbit polyclonal

7.3. Strains

Table 6: Bacterial strains and yeast strains.

Name	Genotype
E.Coli Top 10	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(StrR) endA1 λ-
S. cerevisiae RSY1293	MATα, ura3-1, leu2-3,-112, his3-11,15, trp1-1, ade2-1, can1-100, sec61::HIS3

7.4. Restriciton Enzymes

Table 7: Restriction enzymes used in this publication.

Restriction Enzyme	Supplier	Conditions	Overhang
AflII	New England Biolabs	37 °C 1x Cutsmart Buffer	5′…C`TTAAG…3′ 3′…GAATT`C…5′
AgeI	New England Biolabs	37 °C 1x Cutsmart Buffer	5′…A’CCGGT…3′ 3′…TGGCC’A…5′
BamH1-HF	New England Biolabs	37 °C 1x Cutsmart Buffer	5’…G`GATCC…3’ 3’…CCTAG`G…5’
BglII	New England Biolabs	37 °C 1x Cutsmart Buffer	5′…A’GATCT…3′ 3′…TCTAG’A…5′
HindIII-HF	New England Biolabs	37 °C 1x Cutsmart Buffer	5’…A`AGCTT…3’ 3’…TTCGA`A…5’
HpaI	New England Biolabs	37 °C 1x Cutsmart Buffer	5′…GTT’AAC…3′ 3′…CAA’TTG…5′
RsrII	New England Biolabs	37 °C 1x Cutsmart Buffer	5′…CG’GWCCG…3′ 3′…GCCWG’GC…5′
SalI-HF	New England Biolabs	37 °C 1x Cutsmart Buffer	5’…G`TCGAG…3’ 3’…CAGCT`G…5’
SpeI	New England Biolabs	37 °C 1x Cutsmart Buffer	5′…A’CTAGT…3′ 3′…TGATC’A…5′
XhoI	New England Biolabs	37 °C 1x Cutsmart Buffer	5’…C`TCGAG…3’ 3’…GAGCT`C…5’
XmaI	New England Biolabs	37 °C 1x Cutsmart Buffer	5′…C’CCGGG…3′ 3′…GGGCC’C…5′

7.5. Buffers

Table 8: Buffers used in this study.

Buffer	Component	Concentration
EndoH Buffer	Sodium citrate SDS in PBS	50 mM 1%
Fixation-buffer	Methanol Acetic acid	45% 10%
LiACTE (10x)	Lithium acetate pH 7.5 Tris/HCl pH 7.5 EDTA pH 8.0	1 M 100 mM 10 mM
Lysis-buffer	Tris/HCl pH 7.5 NaCl EDTA Triton X-100 PMSF PiC	30 mM 120 mM 5 mM 1% 1 mM 1 mM
PBS (10x, pH 7.5)	NaCl $NaPO_4$	125 mM 12.5 mM
PEG 4000 (40%)-Solution	PEG 4000 (40%) 10x LiAcTE	40 % 1x
PiC (500x)	Benzamidine Pepstatin Leupeptin Antipain Chymostatin DMSO EtOH	5 mg/ml 1 mg/ml 1 mg/ml 1 mg/ml 1 mg/ml 40 % 60 %
Running buffer (10x)	Glycine Tris base SDS ddH_2O	144 g 30 g 10 g 100 ml
Sample buffer (2x)	Tris/HCl pH 6.8 SDS Bromphenol blue Gylcerol β-mercaptoethanol	100 mM 4 % 0.2 % 20 % 1.4 %
Sorbitol (1 M, pH 7.5)	Sorbitol Tris/HCl pH 7.5	1 M 100 mM

TEP-buffer	Tris/HCl pH 7.5 EDTA PMSF	50 mM 5 mM 1 mM
TNET	Tris/HCl pH 7.5 NaCl EDTA Triton X-100	30 mM 120 mM 5 mM 1 %
IP mix	TNET NaN_3 BSA PMSF α-HA antibody	800 µl/sample 5 µl/sample 5 µl/sample 2 µl/sample 10 µl/sample
DMEM		
Lysis-Buffer for cell line labeling		
IP-Buffer for cell line labeleling		

7.6. Cloning Strategies

The constructs used in this work were created by using conventional cloning techniques, like polymerase chain reaction, restriction digestion and ligation.
The plasmid used for cloning and expression in yeast is pRS426 (from N. Kralli, Scripps Institute, San Diego, CA).

The construct RI-Sp125 was created by cloning the spectrin sequence out of the plasmid pRSETA-Sp16, a plasmid which we received from Prof. Jane Clarke (University of Cambridge).

The constructs C14-prM-E and 2K-NS4B-V5 were received from Dominic Höpfner (Novartis, Basel)

7.7. Yeast Transformation using the LiAc-Method

Yeast cells were grown in YPD media at 30°C until an OD_{600} of 0.6 – 0.8. The cells were harvested, washed with 5 ml 1x LiAc and resuspended in 2 ml 1x LiAc. 50 µl of the prepared yeast suspension was used for a single transformation. The cells were mixed with 500 µg of DNA and 300 µl transformation solution, containing 100 µg SS DNA. The samples were heat shocked for 30 min at 42°C, followed by washing and resuspension in 1 M sorbitol solution, followed by plating on SD_{-URA} plates. Transformants were obtained after incubation at 30°C for 3 – 4 days.

7.8. Radiolabeling with [^{35}S]Methionine/Cysteine in yeast

Yeast transformants were inoculated in 3 ml YPD media overnight at 30°C. The next day, the OD_{600} of the overnight culture was measured via a photometer, adjusted to 0.6 and regrown for 2 h in SD_{-URA} media at 30°C. 1.5 ml of the cell suspension was harvested, resuspended in 150 µl SD_{-URA} and incubated at 30°C for 15 min. The cells were in vivo pulse-labeled for 5 – 10 min with 150 µCi/mL [^{35}S] methionine/cysteine (Perkin Elmer) at 30°C. Translation was stopped by adding 2.5 µl of a 1 M NaN_3 solution and putting the samples on ice. Cells were washed with 150 µl PBS and resuspended in 150 µl TEP, followed by cell lysis by adding 100 µl glass beads and vortexing for 5 – 10 min at 4°C. 150 µl TEP + 2 % SDS were added to the cells, heated at 95°C for 10 min, cleared by centrifugation and subjected to immunoprecipitation overnight in IP Buffer. The following day, 20 µl Sepharose A beads were added to the samples, incubated for 1 h at 4°C, washed three times with TNET buffer and resuspended in 50 µl 2x sample buffer. The probes were heated to 95°C for 10 minutes, loaded on an SDS-gel and run overnight. The gels were fixed and destained for 45 min by using the fixation buffer, followed by excessive washing with H_2O. The gels were put onto whatman paper and dryed for 2 – 3 h. Afterwards, the dryed gels were put into a phosphorimager plate, and analyzed after 3 – 14 days.

8. References

ANFINSEN, C.B., HABER, E., SELA, M., and WHITE, F.H. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc. Natl. Acad. Sci. U.S.a. *47*, 1309–1314.

Araki, K., and Nagata, K. (2011). Protein folding and quality control in the ER. Cold Spring Harb Perspect Biol *3*, a007526–a007526.

Ast, T., Michaelis, S., and Schuldiner, M. (2016). The Protease Ste24 Clears Clogged Translocons. Cell *164*, 103–114.

Auclair, S.M., Bhanu, M.K., and Kendall, D.A. (2012). Signal peptidase I: cleaving the way to mature proteins. Protein Sci. *21*, 13–25.

Aviram, N., Ast, T., Costa, E.A., Arakel, E.C., Chuartzman, S.G., Jan, C.H., Haßdenteufel, S., Dudek, J., Jung, M., Schorr, S., et al. (2016). The SND proteins constitute an alternative targeting route to the endoplasmic reticulum. Nature *540*, 134–138.

Bacher, G., Lütcke, H., Jungnickel, B., Rapoport, T.A., and Dobberstein, B. (1996). Regulation by the ribosome of the GTPase of the signal-recognition particle during protein targeting. Nature *381*, 248–251.

Barkocy-Gallagher, G.A., and Bassford, P.J. (1992). Synthesis of precursor maltose-binding protein with proline in the +1 position of the cleavage site interferes with the activity of Escherichia coli signal peptidase I in vivo. J. Biol. Chem. *267*, 1231–1238.

Becker, J., Walter, W., Yan, W., and Craig, E.A. (1996). Functional interaction of cytosolic hsp70 and a DnaJ-related protein, Ydj1p, in protein translocation in vivo. Mol. Cell. Biol. *16*, 4378–4386.

Behnke, J., Feige, M.J., and Hendershot, L.M. (2015). BiP and its nucleotide exchange factors Grp170 and Sil1: mechanisms of action and biological functions. J. Mol. Biol. *427*, 1589–1608.

Beltzer, J.P., Fiedler, K., Fuhrer, C., Geffen, I., Handschin, C., Wessels, H.P., and Spiess, M. (1991). Charged residues are major determinants of the transmembrane orientation of a signal-anchor sequence. J. Biol. Chem. *266*, 973–978.

Bhatt, S., Gething, P.W., Brady, O.J., Messina, J.P., Farlow, A.W., Moyes, C.L., Drake, J.M., Brownstein, J.S., Hoen, A.G., Sankoh, O., et al. (2013). The global distribution and burden of dengue. Nature *496*, 504–507.

Blond-Elguindi, S., Cwirla, S.E., Dower, W.J., Lipshutz, R.J., Sprang, S.R., Sambrook, J.F., and Gething, M.J. (1993). Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP. Cell *75*, 717–728.

Brylinski, M., Konieczny, L., and Roterman, I. (2006). Hydrophobic collapse in (in silico) protein folding. Comput Biol Chem *30*, 255–267.

Bukau, B., and Horwich, A.L. (1998). The Hsp70 and Hsp60 chaperone machines. Cell *92*, 351–366.

Bukau, B., Deuerling, E., Pfund, C., and Craig, E.A. (2000). Getting newly synthesized proteins into shape. Cell *101*, 119–122.

Carvalho, P., Goder, V., and Rapoport, T.A. (2006). Distinct ubiquitin-ligase complexes define convergent pathways for the degradation of ER proteins. Cell *126*, 361–373.

Chakrabarti, A., Chen, A.W., and Varner, J.D. (2011). A review of the mammalian unfolded protein response. Biotechnol. Bioeng. *108*, 2777–2793.

Chartron, J.W., Suloway, C.J.M., Zaslaver, M., and Clemons, W.M. (2010). Structural characterization of the Get4/Get5 complex and its interaction with Get3. Proc. Natl. Acad. Sci. U.S.a. *107*, 12127–12132.

Chen, X., Shen, J., and Prywes, R. (2002). The luminal domain of ATF6 senses endoplasmic reticulum (ER) stress and causes translocation of ATF6 from the ER to the Golgi. J. Biol. Chem. *277*, 13045–13052.

Clerc, S., Hirsch, C., Oggier, D.M., Deprez, P., Jakob, C., Sommer, T., and Aebi, M. (2009). Htm1 protein generates the N-glycan signal for glycoprotein degradation in the endoplasmic reticulum. J. Cell Biol. *184*, 159–172.

Clerico, E.M., Tilitsky, J.M., Meng, W., and Gierasch, L.M. (2015). How hsp70 molecular machines interact with their substrates to mediate diverse physiological functions. J. Mol. Biol. *427*, 1575–1588.

Cymer, F., Heijne, von, G., and White, S.H. (2015). Mechanisms of integral membrane protein insertion and folding. J. Mol. Biol. *427*, 999–1022.

Dalbey, R.E. (2013). Chapter 774 Signal Peptidase I. In Handbook of Proteolytic Enzymes, (Elsevier), pp. 3493–3497.

De Marothy, M.T., and Elofsson, A. (2015). Marginally hydrophobic transmembrane α-helices shaping membrane protein folding. Protein Sci. *24*, 1057–1074.

Demirci, E., Junne, T., Baday, S., Bernèche, S., and Spiess, M. (2013). Functional asymmetry within the Sec61p translocon. Proc. Natl. Acad. Sci. U.S.a. *110*, 18856–18861.

Denic, V. (2012). A portrait of the GET pathway as a surprisingly complicated young man. Trends Biochem. Sci. *37*, 411–417.

Denic, V., Dötsch, V., and Sinning, I. (2013). Endoplasmic reticulum targeting and insertion of tail-anchored membrane proteins by the GET pathway. Cold Spring Harb Perspect Biol *5*, a013334–a013334.

Denzer, A.J., Nabholz, C.E., and Spiess, M. (1995). Transmembrane orientation of signal-anchor proteins is affected by the folding state but not the size of the N-terminal domain. Embo J. *14*, 6311–6317.

Deshaies, R.J., Sanders, S.L., Feldheim, D.A., and Schekman, R. (1991). Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex. Nature *349*, 806–808.

Devaraneni, P.K., Conti, B., Matsumura, Y., Yang, Z., Johnson, A.E., and Skach, W.R. (2011). Stepwise insertion and inversion of a type II signal anchor sequence in the ribosome-Sec61 translocon complex. Cell *146*, 134–147.

Do, H., Falcone, D., Lin, J., Andrews, D.W., and Johnson, A.E. (1996). The cotranslational integration of membrane proteins into the phospholipid bilayer is a multistep process. Cell *85*, 369–378.

Eisenberg, D. (2003). The discovery of the alpha-helix and beta-sheet, the principal structural features of proteins. Proc. Natl. Acad. Sci. U.S.a. *100*, 11207–11210.

Ellgaard, L., and Helenius, A. (2003). Quality control in the endoplasmic reticulum. Nat. Rev. Mol. Cell Biol. *4*, 181–191.

Ellis, R.J., and Minton, A.P. (2006). Protein aggregation in crowded environments. Biol. Chem. *387*, 485–497.

Fedorov, A.N., and Baldwin, T.O. (1997). Cotranslational protein folding. J. Biol. Chem. *272*, 32715–32718.

Fersht, A.R. (1995). Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. Proc. Natl. Acad. Sci. U.S.a. *92*, 10869–10873.

Gasper, R., Meyer, S., Gotthardt, K., Sirajuddin, M., and Wittinghofer, A. (2009). It takes two to tango: regulation of G proteins by dimerization. Nat. Rev. Mol. Cell Biol. *10*, 423–429.

Goder, V., and Spiess, M. (2003). Molecular mechanism of signal sequence orientation in the endoplasmic reticulum. Embo J. *22*, 3645–3653.

Görlich, D., and Rapoport, T.A. (1993). Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. Cell *75*, 615–630.

Görlich, D., Hartmann, E., Prehn, S., and Rapoport, T.A. (1992). A protein of the endoplasmic reticulum involved early in polypeptide translocation. Nature *357*, 47–52.

Green, N., Fang, H., and Walter, P. (1992). Mutants in three novel complementation groups inhibit membrane protein insertion into and soluble protein translocation across

the endoplasmic reticulum membrane of Saccharomyces cerevisiae. J. Cell Biol. *116*, 597–604.

Harada, Y., Li, H., Wall, J.S., Li, H., and Lennarz, W.J. (2011). Structural studies and the assembly of the heptameric post-translational translocon complex. J. Biol. Chem. *286*, 2956–2965.

Harayama, T., and Riezman, H. (2018). Understanding the diversity of membrane lipid composition. Nat. Rev. Mol. Cell Biol. *19*, 281–296.

Hartmann, E., Rapoport, T.A., and Lodish, H.F. (1989). Predicting the orientation of eukaryotic membrane-spanning proteins. Proc. Natl. Acad. Sci. U.S.a. *86*, 5786–5790.

Hedin, L.E., Ojemalm, K., Bernsel, A., Hennerdal, A., Illergård, K., Enquist, K., Kauko, A., Cristobal, S., Heijne, von, G., Lerch-Bader, M., et al. (2010). Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. J. Mol. Biol. *396*, 221–229.

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

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

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

Helmers, J., Schmidt, D., Glavy, J.S., Blobel, G., and Schwartz, T. (2003). The beta-subunit of the protein-conducting channel of the endoplasmic reticulum functions as the guanine nucleotide exchange factor for the beta-subunit of the signal recognition particle receptor. J. Biol. Chem. *278*, 23686–23690.

Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S.H., and Heijne, von, G. (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature *433*, 377–381.

Hetz, C., and Glimcher, L.H. (2009). Fine-tuning of the unfolded protein response: Assembling the IRE1alpha interactome. Mol. Cell *35*, 551–561.

Higy, M., Junne, T., and Spiess, M. (2004). Topogenesis of membrane proteins at the endoplasmic reticulum. Biochemistry *43*, 12716–12722.

Ito, M., Oiso, Y., Murase, T., Kondo, K., Saito, H., Chinzei, T., Racchi, M., and Lively, M.O. (1993). Possible involvement of inefficient cleavage of preprovasopressin by signal peptidase as a cause for familial central diabetes insipidus. J. Clin. Invest. *91*, 2565–2571.

Jiang, Y., Cheng, Z., Mandon, E.C., and Gilmore, R. (2008). An interaction between the SRP receptor and the translocon is critical during cotranslational protein translocation. J. Cell Biol. *180*, 1149–1161.

Jung, S.-J., Kim, J.E.H., Reithinger, J.H., and Kim, H. (2014). The Sec62-Sec63 translocon facilitates translocation of the C-terminus of membrane proteins. J. Cell. Sci. *127*, 4270–4278.

Junne, T., and Spiess, M. (2017). Integration of transmembrane domains is regulated by their downstream sequences. J. Cell. Sci. *130*, 372–381.

Junne, T., Kocik, L., and Spiess, M. (2010). The hydrophobic core of the Sec61 translocon defines the hydrophobicity threshold for membrane integration. Mol. Biol. Cell *21*, 1662–1670.

Karaoglu, D., Kelleher, D.J., and Gilmore, R. (1997). The highly conserved Stt3 protein is a subunit of the yeast oligosaccharyltransferase and forms a subcomplex with Ost3p and Ost4p. J. Biol. Chem. *272*, 32513–32520.

Karpinets, T.V., Greenwood, D.J., Sams, C.E., and Ammons, J.T. (2006). RNA:protein ratio of the unicellular organism as a characteristic of phosphorous and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis. BMC Biol. *4*, 30.

Kebache, S., Cardin, E., Nguyên, D.T., Chevet, E., and Larose, L. (2004). Nck-1 antagonizes the endoplasmic reticulum stress-induced inhibition of translation. J. Biol. Chem. *279*, 9662–9671.

Keenan, R.J., Freymann, D.M., Walter, P., and Stroud, R.M. (1998). Crystal structure of the signal sequence binding subunit of the signal recognition particle. Cell *94*, 181–191.

Kim, Y.E., Hipp, M.S., Bracher, A., Hayer-Hartl, M., and Hartl, F.U. (2013a). Molecular chaperone functions in protein folding and proteostasis. Annu. Rev. Biochem. *82*, 323–355.

Kim, Y.E., Hipp, M.S., Bracher, A., Hayer-Hartl, M., and Hartl, F.U. (2013b). Molecular chaperone functions in protein folding and proteostasis. Annu. Rev. Biochem. *82*, 323–355.

Kocik, L., Junne, T., and Spiess, M. (2012). Orientation of internal signal-anchor sequences at the Sec61 translocon. J. Mol. Biol. *424*, 368–378.

Kornfeld, R., and Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. Annu. Rev. Biochem. *54*, 631–664.

Kozlov, G., Bastos-Aristizabal, S., Määttänen, P., Rosenauer, A., Zheng, F., Killikelly, A., Trempe, J.-F., Thomas, D.Y., and Gehring, K. (2010). Structural basis of cyclophilin B binding by the calnexin/calreticulin P-domain. J. Biol. Chem. *285*, 35551–35557.

Krogh, A., Larsson, B., Heijne, von, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. *305*, 567–580.

Kubota, K., Yamagata, A., Sato, Y., Goto-Ito, S., and Fukai, S. (2012). Get1 stabilizes an open dimer conformation of get3 ATPase by binding two distinct interfaces. J. Mol. Biol. *422*, 366–375.

Lakkaraju, A.K.K., Thankappan, R., Mary, C., Garrison, J.L., Taunton, J., and Strub, K. (2012). Efficient secretion of small proteins in mammalian cells relies on Sec62-dependent posttranslational translocation. Mol. Biol. Cell *23*, 2712–2722.

Lerch-Bader, M., Lundin, C., Kim, H., Nilsson, I., and Heijne, von, G. (2008). Contribution of positively charged flanking residues to the insertion of transmembrane helices into the endoplasmic reticulum. Proc. Natl. Acad. Sci. U.S.a. *105*, 4127–4132.

Li, L., Vorobyov, I., and Allen, T.W. (2013). The different interactions of lysine and arginine side chains with lipid membranes. J Phys Chem B *117*, 11906–11920.

Lundin, C., Kim, H., Nilsson, I., White, S.H., and Heijne, von, G. (2008). Molecular code for protein insertion in the endoplasmic reticulum membrane is similar for N(in)-C(out) and N(out)-C(in) transmembrane helices. Proc. Natl. Acad. Sci. U.S.a. *105*, 15702–15707.

MacCallum, J.L., and Tieleman, D.P. (2011). Hydrophobicity scales: a thermodynamic looking glass into lipid-protein interactions. Trends Biochem. Sci. *36*, 653–662.

Malkin, L.I., and Rich, A. (1967). Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. J. Mol. Biol. *26*, 329–346.

Mariappan, M., Mateja, A., Dobosz, M., Bove, E., Hegde, R.S., and Keenan, R.J. (2011). The mechanism of membrane-associated steps in tail-anchored protein insertion. Nature *477*, 61–66.

Marino, J., Heijne, von, G., and Beckmann, R. (2016). Small protein domains fold inside the ribosome exit tunnel. FEBS Lett. *590*, 655–660.

Mason, N., Ciufo, L.F., and Brown, J.D. (2000). Elongation arrest is a physiologically important function of signal recognition particle. Embo J. *19*, 4164–4174.

Mateja, A., Szlachcic, A., Downing, M.E., Dobosz, M., Mariappan, M., Hegde, R.S., and Keenan, R.J. (2009). The structural basis of tail-anchored membrane protein recognition by Get3. Nature *461*, 361–366.

Matlack, K.E., Misselwitz, B., Plath, K., and Rapoport, T.A. (1999). BiP acts as a molecular ratchet during posttranslational transport of prepro-alpha factor across the ER membrane. Cell *97*, 553–564.

Michalak, M., Groenendyk, J., Szabo, E., Gold, L.I., and Opas, M. (2009). Calreticulin, a multi-process calcium-buffering chaperone of the endoplasmic reticulum. Biochem. J. *417*, 651–666.

Miller, J.D., Tajima, S., Lauffer, L., and Walter, P. (1995). The beta subunit of the signal recognition particle receptor is a transmembrane GTPase that anchors the alpha subunit, a peripheral membrane GTPase, to the endoplasmic reticulum membrane. J. Cell Biol. *128*, 273–282.

Misselwitz, B., Staeck, O., Matlack, K.E., and Rapoport, T.A. (1999). Interaction of BiP with the J-domain of the Sec63p component of the endoplasmic reticulum protein translocation complex. J. Biol. Chem. *274*, 20110–20115.

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

Ngosuwan, J., Wang, N.M., Fung, K.L., and Chirico, W.J. (2003). Roles of cytosolic Hsp70 and Hsp40 molecular chaperones in post-translational translocation of presecretory proteins into the endoplasmic reticulum. J. Biol. Chem. *278*, 7034–7042.

Nilsson, I., Lara, P., Hessa, T., Johnson, A.E., Heijne, von, G., and Karamyshev, A.L. (2015a). The code for directing proteins for translocation across ER membrane: SRP cotranslationally recognizes specific features of a signal sequence. J. Mol. Biol. *427*, 1191–1201.

Nilsson, O.B., Hedman, R., Marino, J., Wickles, S., Bischoff, L., Johansson, M., Müller-Lucks, A., Trovato, F., Puglisi, J.D., O'Brien, E.P., et al. (2015b). Cotranslational Protein Folding inside the Ribosome Exit Tunnel. Cell Rep *12*, 1533–1540.

Nilsson, O.B., Nickson, A.A., Hollins, J.J., Wickles, S., Steward, A., Beckmann, R., Heijne, von, G., and Clarke, J. (2017). Cotranslational folding of spectrin domains via partially structured states. Nat. Struct. Mol. Biol. *24*, 221–225.

Norazharuddin, H., and Lai, N.S. (2018). Roles and Prospects of Dengue Virus Non-structural Proteins as Antiviral Targets: An Easy Digest. Malays J Med Sci *25*, 6–15.

Nyathi, Y., Wilkinson, B.M., and Pool, M.R. (2013). Co-translational targeting and translocation of proteins to the endoplasmic reticulum. Biochim. Biophys. Acta *1833*, 2392–2402.

Ohgushi, M., and Wada, A. (1983). “Molten-globule state”: a compact form of globular proteins with mobile side-chains. FEBS Lett. *164*, 21–24.

Oliver, J.D., van der Wal, F.J., Bulleid, N.J., and High, S. (1997). Interaction of the thiol-dependent reductase ERp57 with nascent glycoproteins. Science *275*, 86–88.

Onuchic, J.N., Luthey-Schulten, Z., and Wolynes, P.G. (1997). Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem *48*, 545–600.

Panzner, S., Dreier, L., Hartmann, E., Kostka, S., and Rapoport, T.A. (1995). Posttranslational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p. Cell *81*, 561–570.

Park, E., and Rapoport, T.A. (2012). Mechanisms of Sec61/SecY-mediated protein translocation across membranes. Annu Rev Biophys *41*, 21–40.

Peleg, O., and Lim, R.Y.H. (2010). Converging on the function of intrinsically disordered nucleoporins in the nuclear pore complex. Biol. Chem. *391*, 719–730.

Peluso, P., Herschlag, D., Nock, S., Freymann, D.M., Johnson, A.E., and Walter, P. (2000). Role of 4.5S RNA in assembly of the bacterial signal recognition particle with its receptor. Science *288*, 1640–1643.

Peluso, P., Shan, S.O., Nock, S., Herschlag, D., and Walter, P. (2001). Role of SRP RNA in the GTPase cycles of Ffh and FtsY. Biochemistry *40*, 15224–15233.

Perera, R., and Kuhn, R.J. (2008). Structural proteomics of dengue virus. Curr. Opin. Microbiol. *11*, 369–377.

Plemper, R.K., Böhmler, S., Bordallo, J., Sommer, T., and Wolf, D.H. (1997). Mutant analysis links the translocon and BiP to retrograde protein transport for ER degradation. Nature *388*, 891–895.

Pool, M.R. (2005). Signal recognition particles in chloroplasts, bacteria, yeast and mammals (review). Mol. Membr. Biol. *22*, 3–15.

Preissler, S., and Deuerling, E. (2012). Ribosome-associated chaperones as key players in proteostasis. Trends Biochem. Sci. *37*, 274–283.

Ptitsyn, O.B. (1973). [Stages in the mechanism of self-organization of protein molecules]. Dokl. Akad. Nauk SSSR *210*, 1213–1215.

Quan, E.M., Kamiya, Y., Kamiya, D., Denic, V., Weibezahn, J., Kato, K., and Weissman, J.S. (2008). Defining the glycan destruction signal for endoplasmic reticulum-associated degradation. Mol. Cell *32*, 870–877.

Raden, D., Song, W., and Gilmore, R. (2000). Role of the cytoplasmic segments of Sec61alpha in the ribosome-binding and translocation-promoting activities of the Sec61 complex. J. Cell Biol. *150*, 53–64.

Rapoport, T.A., Li, L., and Park, E. (2017). Structural and Mechanistic Insights into Protein Translocation. Annu. Rev. Cell Dev. Biol. *33*, annurev–cellbio–100616–060439.

Raven, J.F., Baltzis, D., Wang, S., Mounir, Z., Papadakis, A.I., Gao, H.Q., and Koromilas, A.E. (2008). PKR and PKR-like endoplasmic reticulum kinase induce the proteasome-dependent degradation of cyclin D1 via a mechanism requiring eukaryotic initiation factor 2alpha phosphorylation. J. Biol. Chem. *283*, 3097–3108.

Römisch, K. (2017). A Case for Sec61 Channel Involvement in ERAD. Trends Biochem. Sci. *42*, 171–179.

Ruggiano, A., Foresti, O., and Carvalho, P. (2014). Quality control: ER-associated degradation: protein quality control and beyond. J. Cell Biol. *204*, 869–879.

Sato, B.K., Schulz, D., Do, P.H., and Hampton, R.Y. (2009). Misfolded membrane proteins are specifically recognized by the transmembrane domain of the Hrd1p ubiquitin ligase. Mol. Cell *34*, 212–222.

Schatz, G., and Dobberstein, B. (1996). Common principles of protein translocation across membranes. Science *271*, 1519–1526.

Schrag, J.D., Bergeron, J.J., Li, Y., Borisova, S., Hahn, M., Thomas, D.Y., and Cygler, M. (2001). The Structure of calnexin, an ER chaperone involved in quality control of protein folding. Mol. Cell *8*, 633–644.

Schuldiner, M., Metz, J., Schmid, V., Denic, V., Rakwalska, M., Schmitt, H.D., Schwappach, B., and Weissman, J.S. (2008). The GET complex mediates insertion of tail-anchored proteins into the ER membrane. Cell *134*, 634–645.

Shen, J., Chen, X., Hendershot, L., and Prywes, R. (2002). ER stress regulation of ATF6 localization by dissociation of BiP/GRP78 binding and unmasking of Golgi localization signals. Dev. Cell *3*, 99–111.

Shore, G.C., Papa, F.R., and Oakes, S.A. (2011). Signaling cell death from the endoplasmic reticulum stress response. Curr. Opin. Cell Biol. *23*, 143–149.

Siegel, V., and Walter, P. (1986). Removal of the Alu structural domain from signal recognition particle leaves its protein translocation activity intact. Nature *320*, 81–84.

Sommer, N., Junne, T., Kalies, K.-U., Spiess, M., and Hartmann, E. (2013). TRAP assists membrane protein topogenesis at the mammalian ER membrane. Biochim. Biophys. Acta *1833*, 3104–3111.

Stefanovic, S., and Hegde, R.S. (2007). Identification of a targeting factor for posttranslational membrane protein insertion into the ER. Cell *128*, 1147–1159.

Stefer, S., Reitz, S., Wang, F., Wild, K., Pang, Y.-Y., Schwarz, D., Bomke, J., Hein, C., Löhr, F., Bernhard, F., et al. (2011). Structural basis for tail-anchored membrane protein biogenesis by the Get3-receptor complex. Science *333*, 758–762.

Stigliano, I.D., Alculumbre, S.G., Labriola, C.A., Parodi, A.J., and D'Alessio, C. (2011). Glucosidase II and N-glycan mannose content regulate the half-lives of monoglucosylated species in vivo. Mol. Biol. Cell *22*, 1810–1823.

Tai, V.W., and Imperiali, B. (2001). Substrate specificity of the glycosyl donor for oligosaccharyl transferase. J. Org. Chem. *66*, 6217–6228.

Tanford, C., and Reynolds, J. (2003). Nature's Robots (OUP Oxford).

Taylor, S.C., Ferguson, A.D., Bergeron, J.J.M., and Thomas, D.Y. (2004). The ER protein folding sensor UDP-glucose glycoprotein-glucosyltransferase modifies substrates distant to local changes in glycoprotein conformation. Nat. Struct. Mol. Biol. *11*, 128–134.

Therien, A.G., Huber, J.L., Wilson, K.E., Beaulieu, P., Caron, A., Claveau, D., Deschamps, K., Donald, R.G.K., Galgoci, A.M., Gallant, M., et al. (2012). Broadening the spectrum of β-lactam antibiotics through inhibition of signal peptidase type I. Antimicrob. Agents Chemother. *56*, 4662–4670.

Tripathi, A., Mandon, E.C., Gilmore, R., and Rapoport, T.A. (2017). Two alternative binding mechanisms connect the protein translocation Sec71-Sec72 complex with heat shock proteins. J. Biol. Chem. *292*, 8007–8018.

Trombetta, E.S., Simons, J.F., and Helenius, A. (1996). Endoplasmic reticulum glucosidase II is composed of a catalytic subunit, conserved from yeast to mammals, and a tightly bound noncatalytic HDEL-containing subunit. J. Biol. Chem. *271*, 27509–27516.

Tyson, J.R., and Stirling, C.J. (2000). LHS1 and SIL1 provide a lumenal function that is essential for protein translocation into the endoplasmic reticulum. Embo J. *19*, 6440–6452.

Udgaonkar, J.B. (2013). Polypeptide chain collapse and protein folding. Arch. Biochem. Biophys. *531*, 24–33.

Van den Berg, B., Clemons, W.M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C., and Rapoport, T.A. (2004). X-ray structure of a protein-conducting channel. Nature *427*, 36–44.

Wahlberg, J.M., and Spiess, M. (1997). Multiple determinants direct the orientation of signal-anchor proteins: the topogenic role of the hydrophobic signal domain. J. Cell Biol. *137*, 555–562.

Walker, S.J., and Lively, M.O. (2013). Chapter 778 Signal Peptidase (Eukaryote). In Handbook of Proteolytic Enzymes, (Elsevier), pp. 3512–3517.

Wang, F., Brown, E.C., Mak, G., Zhuang, J., and Denic, V. (2010). A chaperone cascade sorts proteins for posttranslational membrane insertion into the endoplasmic reticulum. Mol. Cell *40*, 159–171.

Watson, H. (2015). Biological membranes. Essays Biochem. *59*, 43–69.

Wereszczynski, J., and McCammon, J.A. (2012). Nucleotide-dependent mechanism of Get3 as elucidated from free energy calculations. Proc. Natl. Acad. Sci. U.S.a. *109*, 7759–7764.

Wetlaufer, D.B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. Proc. Natl. Acad. Sci. U.S.a. *70*, 697–701.

Wilkinson, B.M., Brownsword, J.K., Mousley, C.J., and Stirling, C.J. (2010). Sss1p is required to complete protein translocon activation. J. Biol. Chem. *285*, 32671–32677.

Wimley, W.C., and White, S.H. (1996). Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nat. Struct. Biol. *3*, 842–848.

Wolfenden, R. (2007). Experimental measures of amino acid hydrophobicity and the topology of transmembrane and globular proteins. J. Gen. Physiol. *129*, 357–362.

Xie, W., Kanehara, K., Sayeed, A., and Ng, D.T.W. (2009). Intrinsic conformational determinants signal protein misfolding to the Hrd1/Htm1 endoplasmic reticulum-associated degradation system. Mol. Biol. Cell *20*, 3317–3329.

Yoshida, H., Matsui, T., Yamamoto, A., Okada, T., and Mori, K. (2001). XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. Cell *107*, 881–891.

Zhang, X., Schaffitzel, C., Ban, N., and Shan, S.-O. (2009). Multiple conformational switches in a GTPase complex control co-translational protein targeting. Proc. Natl. Acad. Sci. U.S.a. *106*, 1754–1759.

Zheng, N., and Gierasch, L.M. (1996). Signal sequences: the same yet different. Cell *86*, 849–852.

Appendix

A. Internal initiation of protein synhesis

A phenomenon observed in several construct series throughout this work was the appearance of a lower band that was shown to be sensitive to deglycosylation with EndoH, indicating that it is translocated into the lumen of the endoplasmic reticulum. C-terminal cleavage of the proteins could be excluded, since the immunoprecipitation was done with an antibody against the very C-terminal end of the protein. Another explanation for this lower band is, that protein synthesis is initiated at an internal methionine codon, resulting in a truncated version of the protein. For example, the full-size protein of RI-DP124 is 92 kDa in size. A possible truncated version, with the methionine codon at position 149 acting as an internal start codon, would have a size of only 75 kDa. This size is in accordance with the bands height on the autoradiograph. Intriguingly, the truncated version of the model protein would act as a stop-transfer model protein, showing the characteristic pattern of stop-transfer glycosylation (**Figure 25 A,** compare with **Figure 10 C)**. The fully glycosylated form of the truncated version has a size of 96 kDa, thus appears in the autoradiographs slightly above the bands, representing the unglycosylated form of the full-size construct (92 kDa). This is also what we find on the autoradiograph (**Figure 25 B)**.

To further analyze, whether the methionine codon at position 149 is responsible for the lower band we mutated the methionine to an alanine in some of the constructs where the problem of internal initiation occurred. Upon expression of RI-DP124 and the methionine-less mutant RI-DP124 M149A of the constructs, the lower band vanished, proofing that the lower band is caused by internal initiation (**Figure 25 C**)

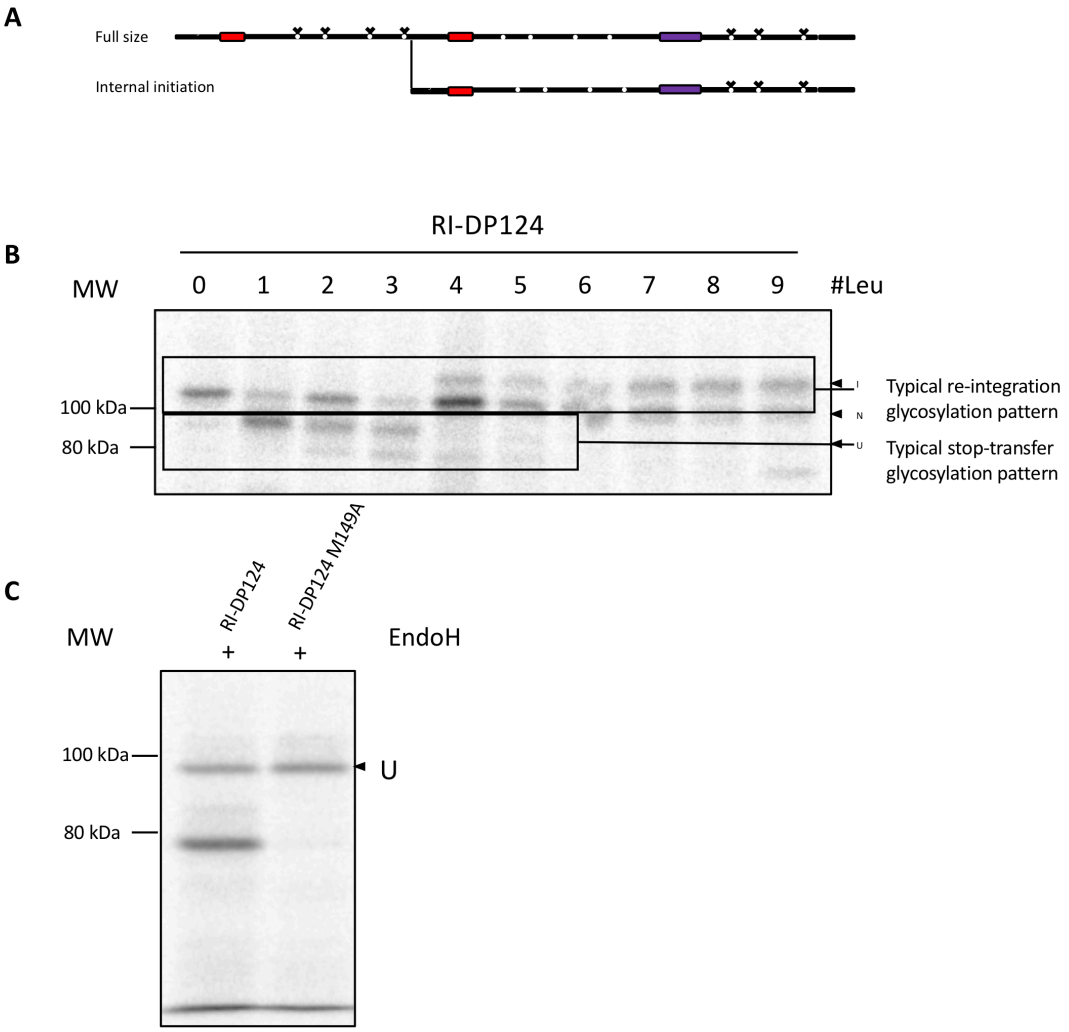


Figure 25: A Schematic representation of the model protein (above) and the construct that is produced due to internal initiation of a start codon at position 149. **B** Autoradiograph of RI-DP124. The lower bands represent a pattern that is common for stop-transfer integration (compare to **Figure 10**), while the upper bands are representing re-integration. **C** Autoradiograph of RI-DP124 and RI-DP124 M149A. The proteins were deglycosylated by endoglycosidase H after immunoprecipitation. The lower band below 80 kDa is does not appear for the mutant construct, proofing that this band appears due to internal initiation.

B. Protein sequences of the re-integration constructs

The H-segment consists of 19 host alanines that are replaced with leucines to modulate the hydrophobicity of the domain. The H-segments used in this work are shown in the following list

```
L00        GGPGAAAAAAAAAAAAAAAAAGPGG
L01        GGPGAAAAAAAAALAAAAAAAAAGPGG
L02        GGPGAAAALAAAALAAAAAAAAAGPGG
L03        GGPGAAAALAAAALAAAALAAAAGPGG
L04        GGPGAAAALALAALAAAALAAAAGPGG
L05        GGPGAAAALALAALAALALAAAAGPGG
L06        GGPGAAAALALALALALALAAAAGPGG
L07        GGPGAAAALALALALALALALAAGPGG
L08        GGPGAALALALALALALALALAAGPGG
L09        GGPGAALALALALALALALALALGPGG
L11        GGPGLLALALALALALALALLGPGG
L15        GGPGLLLALLLALLLALLLALLLGPGG
L19        GGPGLLLLLLLLLLLLLLLLLLLGPGG
```

The following list contains all the construct series used in this work. The signal anchor and the stop-transfer domain are highlighted in red and orange, the H-segment is marked in purple. The cytosolic loop is colored in green.

```
>RI-DP124
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKYHELQWISDNKIESNDLGLYVTFMN
DSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITASPDLKRLLIRTNSVQNWRHSTFGGPGAAAAAAAAAAAAAAAAAAGPGG
WWEITHNTLFIPANETFDRPHNGYVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFISTRKSSTERHVY
YIDLRSPNEIIEVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAV
PRKSFRELNLGKDEFGKDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVASQLNAIVVVVDGRGTGFKG
QDFRSLVRDRLGDYEARDQISAASLYGSLTFVDPQKISLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRFYDSVYTERYMH
TPQENFDGYVESSVHNVTALAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIRYHNANVIVFDKLLDWA
KRAFDGQFVKACYPYDVPDYAGYPYDVPDYAYPYDVPDYA

>RI-DP96
EGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKYH
ELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLLD
KLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKYHELQWISDNKIESNDLGLYVTFMND
SYVVKSVYDDSLKRLLIRTNSVQNWRHSTFGGPGAAAAAAAAAAAAAAAAAAAGPGGWWEITHNTLFIPANETFDRPHNGYVDILP
IGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFISTRKSSTERHVYYIDLRSPNEIIEVTDTSEDGVYDVSFSSG
RRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDEFGKDILVNSYEILPN
DFDETLSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVASQLNAIVVVVDGRGTGFKGQDFRSLVRDRLGDYEARDQISAASLYGSL
TFVDPQKISLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRFYDSVYTERYMHTPQENFDGYVESSVHNVTALAQANRFLLM
HGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIRYHNANVIVFDKLLDWAKRAFDGQFVKACYPYDVPDYAGYPYDVPD
YAYPYDVPDYA
```

>RI-DP71
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKYHELQWISDNKILKRLLIRTNSVQN
WRHSTFGGPGAAAAAAAAAAAAAAAAAAAGPGGWWEITHNTLFIPANETFDRPHNGYVDILPIGGYNHLAYFENSNSSHYKTLTEG
KWEVVNGPLAFDSMENRLYFISTRKSSTERHVYYIDLRSPNEIIEVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSR
KAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDEFGKDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQ
VVKTFSVGFNEVVASQLNAIVVVVDGRGTGFKGQDFRSLVRDRLGDYEARDQISAASLYGSLTFVDPQKISLFGWSYGGYLTLKTL
EKDGGRHFKYGMSVAPVTDWRFYDSVYTERYMHTPQENFDGYVESSVHNVTALAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNG
VENYDVHVFPDSDHSIRYHNANVIVFDKLLDWAKRAFDGQFVKACYPYDVPDYAGYPYDVPDYAYPYDVPDYA

>RI-DP45
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVLKRLLIRTNSVQNWRHSTFGGPGAAAAAAAAAAAAAAA
AAAAGPGGWWEITHNTLFIPANETFDRPHNGYVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFISTRK
SSTERHVYYIDLRSPNEIIEVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHLEKNEVLT
KILEDYAVPRKSFRELNLGKDEFGKDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVASQLNAIVVVVD
GRGTGFKGQDFRSLVRDRLGDYEARDQISAASLYGSLTFVDPQKISLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRFYDS
VYTERYMHTPQENFDGYVESSVHNVTALAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIRYHNANVIV
FDKLLDWAKRAFDGQFVKACYPYDVPDYAGYPYDVPDYAYPYDVPDYA

>RI-DP17
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLLKRLLIRTNSVQNWRHSTFGGPGAAAAAAAAAAAAAAAAAAAGPGGWWEITHNTLFIPANETFDR
PHNGYVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFISTRKSSTERHVYYIDLRSPNEIIEVTDTSED
GVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDEFGKDI
LVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVASQLNAIVVVVDGRGTGFKGQDFRSLVRDRLGDYEARDQ
ISAASLYGSLTFVDPQKISLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRFYDSVYTERYMHTPQENFDGYVESSVHNVTA
LAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIRYHNANVIVFDKLLDWAKRAFDGQFVKACYPYDVPD
YAGYPYDVPDYAYPYDVPDYA

>RI-GS122
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLLKGASSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG
SGSGSGSGSSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSTSGGPGAAAAAAAAAAAAAAAAAAAGPGGWW
EITHNTLFIPANETFDRPHNGYVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFISTRKSSTERHVYYI
DLRSPNEIIEVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVPR
KSFRELNLGKDEFGKDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVASQLNAIVVVVDGRGTGFKGQD
FRSLVRDRLGDYEARDQISAASLYGSLTFVDPQKISLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRFYDSVYTERYMHTP
QENFDGYVESSVHNVTALAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIRYHNANVIVFDKLLDWAKR
AFDGQFVKACYPYDVPDYAGYPYDVPDYAYPYDVPDYA

>RI-GS99
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLLKGASSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG
SGSGSGSGSSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSTSGGPGAAAAAAAAAAAAAAAAAAAGPGGWWEITHNTLFIPANETFDRPHNGYV
DILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFISTRKSSTERHVYYIDLRSPNEIIEVTDTSEDGVYDVS
FSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDEFGKDILVNSYE
ILPNDFDETLSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVASQLNAIVVVVDGRGTGFKGQDFRSLVRDRLGDYEARDQISAASL
YGSLTFVDPQKISLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRFYDSVYTERYMHTPQENFDGYVESSVHNVTALAQANR
FLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIRYHNANVIVFDKLLDWAKRAFDGQFVKACYPYDVPDYAGYPY
DVPDYAYPYDVPDYA

>RI-GS76
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLLKGASSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG
SGSGSGSGSTSGGPGAAAAAAAAAAAAAAAAAAAAGPGGWWEITHNTLFIPANETFDRPHNGYVDILPIGGYNHLAYFENSNSSHYK
TLTEGKWEVVNGPLAFDSMENRLYFISTRKSSTERHVYYIDLRSPNEIIEVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIV
DFHSRKAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDEFGKDILVNSYEILPNDFDETLSDHYPVFFFAYGG
PNSQQVVKTFSVGFNEVVASQLNAIVVVVDGRGTGFKGQDFRSLVRDRLGDYEARDQISAASLYGSLTFVDPQKISLFGWSYGGYL
TLKTLEKDGGRHFKYGMSVAPVTDWRFYDSVYTERYMHTPQENFDGYVESSVHNVTALAQANRFLLMHGTGDDNVHFQNSLKFLDL
LDLNGVENYDVHVFPDSDHSIRYHNANVIVFDKLLDWAKRAFDGQFVKACYPYDVPDYAGYPYDVPDYAYPYDVPDYA

>RI-GS53
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLLKGASSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSTSGGPGAAAAAAAA
AAAAAAAAAAAAGPGGWWEITHNTLFIPANETFDRPHNGYVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRL
YFISTRKSSTERHVYYIDLRSPNEIIEVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHL
EKNEVLTKILEDYAVPRKSFRELNLGKDEFGKDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVASQLN
AIVVVVDGRGTGFKGQDFRSLVRDRLGDYEARDQISAASLYGSLTFVDPQKISLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVT

DWRFYDSVYTERYMHTPQENFDGYVESSVHNVTALAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIRY
HNANVIVFDKLLDWAKRAFDGQFVKACYPYDVPDYAGYPYDVPDYAYPYDVPDYA

>RI-GS31
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLLKASSGSGSGSGSGSGSGSGSGSGSGSTSGGPGAAAAAAAAAAAAAAAAAAGPGGWWEITHN
TLFIPANETFDRPHNGYVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFISTRKSSTERHVYYIDLRSP
NEIIEVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSFRE
LNLGKDEFGKDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVASQLNAIVVVVDGRGTGFKGQDFRSLV
RDRLGDYEARDQISAASLYGSLTFVDPQKISLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRFYDSVYTERYMHTPQENFD
GYVESSVHNVTALAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIRYHNANVIVFDKLLDWAKRAFDGQ
FVKACYPYDVPDYAGYPYDVPDYAYPYDVPDYA

>RI-Nup109
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLLKASSGSGSGSGSGSGSGSGSGSGSTSGGPGAAAAAAAAAAAAAAAAAAGPGGWWEITHN
TLFIPANETFDRPHNGYVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFISTRKSSTERHVYYIDLRSP
NEIIEVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSFRE
LNLGKDEFGKDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVASQLNAIVVVVDGRGTGFKGQDFRSLV
RDRLGDYEARDQISAASLYGSLTFVDPQKISLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRFYDSVYTERYMHTPQENFD
GYVESSVHNVTALAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIRYHNANVIVFDKLLDWAKRAFDGQ
FVKACYPYDVPDYAGYPYDVPDYAYPYDVPDYA

>RI-Sp125
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLLKGASSSSKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAA
HEPAIQGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLEEHRLGSTSGGPGAAAAAAAAAAAAAAAAAAGPG
GWWEITHNTLFIPANETFDRPHNGYVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFISTRKSSTERHV
YYIDLRSPNEIIEVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHLEKNEVLTKILEDYA
VPRKSFRELNLGKDEFGKDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVASQLNAIVVVVDGRGTGFK
GQDFRSLVRDRLGDYEARDQISAASLYGSLTFVDPQKISLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRFYDSVYTERYM
HTPQENFDGYVESSVHNVTALAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIRYHNANVIVFDKLLDW
AKRAFDGQFVKACYPYDVPDYAGYPYDVPDYAYPYDVPDYA

>RI-ZG61
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLLKGASSKPYPCGLCNRCFTRRDLLIRHAQKIHSGNGSSGSGSGSGSGSGSGSGSGSGSTSGGG
PGAAAAAAAAAAAAAAAAAAAAGPGGWWEITHNTLFIPANETFDRPHNGYVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGP
LAFDSMENRLYFISTRKSSTERHVYYIDLRSPNEIIEVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKG
NVLGKSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDEFGKDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQVVKTFSVG
FNEVVASQLNAIVVVVDGRGTGFKGQDFRSLVRDRLGDYEARDQISAASLYGSLTFVDPQKISLFGWSYGGYLTLKTLEKDGGRHF
KYGMSVAPVTDWRFYDSVYTERYMHTPQENFDGYVESSVHNVTALAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHV
FPDSDHSIRYHNANVIVFDKLLDWAKRAFDGQFVKACYPYDVPDYAGYPYDVPDYAYPYDVPDYA

>RI-PG61
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLLKGASSKPYPCGLCNRCFTRRDLLIRPAQKIHSGNGSSGSGSGSGSGSGSGSGSGSGSTSGGG
PGAAAAAAAAAAAAAAAAAAAAGPGGWWEITHNTLFIPANETFDRPHNGYVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGP
LAFDSMENRLYFISTRKSSTERHVYYIDLRSPNEIIEVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKG
NVLGKSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDEFGKDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQVVKTFSVG
FNEVVASQLNAIVVVVDGRGTGFKGQDFRSLVRDRLGDYEARDQISAASLYGSLTFVDPQKISLFGWSYGGYLTLKTLEKDGGRHF
KYGMSVAPVTDWRFYDSVYTERYMHTPQENFDGYVESSVHNVTALAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHV
FPDSDHSIRYHNANVIVFDKLLDWAKRAFDGQFVKACYPYDVPDYAGYPYDVPDYAYPYDVPDYA

>RI-4xZf103
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLLKGASSKPYPCGLCNRCFTRRDLLIRHAQKIHSGNGSSKPYPCGLCNRCFTRRDLLIRHAQKIHS
GNGSSKPYPCGLCNRCFTRRDLLIRHAQKIHSGNGSSKPYPCGLCNRCFTRRDLLIRHAQKIHSGNGSTSGGPGAAAAAAAAAAAA
AAAAAAAGPGGWWEITHNTLFIPANETFDRPHNGYVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFIS
TRKSSTERHVYYIDLRSPNEIIEVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHLEKNE
VLTKILEDYAVPRKSFRELNLGKDEFGKDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVASQLNAIVV
VVDGRGTGFKGQDFRSLVRDRLGDYEARDQISAASLYGSLTFVDPQKISLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRF
YDSVYTERYMHTPQENFDGYVESSVHNVTALAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIRYHNAN
VIVFDKLLDWAKRAFDGQFVKACYPYDVPDYAGYPYDVPDYAYPYDVPDYA

>RI-4xPf103
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL

DKLIRVGIILVLLIWGTVLLLLKGASSKPYPCGLCHRCFTRRDLLIRPAQKIHSGNGSSKPYPCGLCHRCFTRRDLLIRPAQKIHS
GNGSSKPYPCGLCHRCFTRRDLLIRPAQKIHSGNGSSKPYPCGLCHRCFTRRDLLIRPAQKIHSGNGSTSGGPGAAAAAAAAAAAA
AAAAAAAGPGGWWEITHNTLFIPANETFDRPHNGYVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFIS
TRKSSTERHVYYIDLRSPNEIIEVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHLEKNE
VLTKILEDYAVPRKSFRELNLGKDEFGKDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVASQLNAIVV
VVDGRGTGFKGQDFRSLVRDRLGDYEARDQISAASLYGSLTFVDPQKISLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRF
YDSVYTERYMHTPQENFDGYVESSVHNVTALAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIRYHNAN
VIVFDKLLDWAKRAFDGQFVKACYPYDVPDYAGYPYDVPDYAYPYDVPDYA

>RI-Chp109
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLLKGASTQESGKSAHWDFAWPWASRSGDSAFWGLWPWEAQESGKSAHWDFAWPWASRSGDSAFWGL
WPWEAQESGKSAHWDFAWPWASRSGDSAFWGLWPWEAQESGSTSGGPGAAAAAAAAAAAAAAAAAAAGPGGWWEITHNTLFIPANE
TFDRPHNGYVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFISTRKSSTERHVYYIDLRSPNEIIEVTD
TSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDEF
GKDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVASQLNAIVVVVDGRGTGFKGQDFRSLVRDRLGDYE
ARDQISAASLYGSLTFVDPQKISLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRFYDSVYTERYMHTPQENFDGYVESSVH
NVTALAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIRYHNANVIVFDKLLDWAKRAFDGQFVKACYPY
DVPDYAGYPYDVPDYAYPYDVPDYA

>RI-Scr109
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLLKGASTQESGWKWSAHSWRFAPALSWSDGAFEWAPDWGQESGWKWSAHSWRFAPALSWSDGAFEW
APDWGQESGWKWSAHSWRFAPALSWSDGAFEWAPDWGQESGSTSGGPGAAAAAAAAAAAAAAAAAAAGPGGWWEITHNTLFIPANE
TFDRPHNGYVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFISTRKSSTERHVYYIDLRSPNEIIEVTD
TSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDEF
GKDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVASQLNAIVVVVDGRGTGFKGQDFRSLVRDRLGDYE
ARDQISAASLYGSLTFVDPQKISLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRFYDSVYTERYMHTPQENFDGYVESSVH
NVTALAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIRYHNANVIVFDKLLDWAKRAFDGQFVKACYPY
DVPDYAGYPYDVPDYAYPYDVPDYA

>RI-nChp109
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLLKGASTQESGKSAAGEYYAALASRSGDSAAGEYYAALAQESGKSAAGEYYAALASRSGDSAAGEY
YAALAQESGKSAAGEYYAALASRSGDSAAGEYYAALAQESGSTSGGPGAAAAAAAAAAAAAAAAAAAGPGGWWEITHNTLFIPANE
TFDRPHNGYVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFISTRKSSTERHVYYIDLRSPNEIIEVTD
TSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDEF
GKDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVASQLNAIVVVVDGRGTGFKGQDFRSLVRDRLGDYE
ARDQISAASLYGSLTFVDPQKISLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRFYDSVYTERYMHTPQENFDGYVESSVH
NVTALAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIRYHNANVIVFDKLLDWAKRAFDGQFVKACYPY
DVPDYAGYPYDVPDYAYPYDVPDYA

>RI-PMA133
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLLKQEFQAGSIVDELKKTLANTAVVIRDGQLVEIPANEVVPGDILQLEDGTVIPTDGRIVTEDCFL
QIDQSAITGESLAVDKHYGDQTFSSSTVKRGEGFMVVTATGDNTFVGRAAALVNKAAGGQGHFTEVLNGGPGAALALALALALALA
LALALGPGGWWEITHNTLFIPANETFDRPHNGYVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFISTR
KSSTERHVYYIDLRSPNEIIEVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHLEKNEVL
TKILEDYAVPRKSFRELNLGKDEFGKDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVASQLNAIVVVV
DGRGTGFKGQDFRSLVRDRLGDYEARDQISAASLYGSLTFVDPQKISLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRFYD
SVYTERYMHTPQENFDGYVESSVHNVTALAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIRYHNANVI
VFDKLLDWAKRAFDGQFVKACYPYDVPDYAGYPYDVPDYAYPYDVPDYA

>RI-EPT109
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLLKGASSDPNLNTDTPRWTYFSYALGVFLYQTFDGCDGVHARRINQSGPLGELFDHSIDAINSTLS
IFIFASETGMGFSYNLMLSQFAMLTNFYLSTWEEYHTHTLYLSEGGPGAALALALALALALALALGPGGWWEITHNTLFIPANE
TFDRPHNGYVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFISTRKSSTERHVYYIDLRSPNEIIEVTD
TSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDEF
GKDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVASQLNAIVVVVDGRGTGFKGQDFRSLVRDRLGDYE
ARDQISAASLYGSLTFVDPQKISLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRFYDSVYTERYMHTPQENFDGYVESSVH
NVTALAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIRYHNANVIVFDKLLDWAKRAFDGQFVKACYPY
DVPDYAGYPYDVPDYAYPYDVPDYA

>RI-GPR348
MEGGEEEVERIPDELFDTKKKHLLDKLIRVGIILVLLIWGTVLLLKSIPHHSNTPDYQEPNSNYTNDGKLKVSFSVVRNNTFQPKY
HELQWISDNKIESNDLGLYVTFMNDSYVVKSVYDDSYNSVLLEGKTFIHNGQNLTVESITRSMEGGEEEVERIPDELFDTKKKHLL
DKLIRVGIILVLLIWGTVLLLKSESKRIKAQIGDFNHNVLEEEKEKKKLFGLGHWGKAKWYFRSYFKLPLLHLLRNLKNFFTISF
IDPNEETDDSGSSNGTFNFGESSNEIPTLFRKTNTGSDENVSASGGVRLLDYNSAKPLDMSKYAMSEQPDLERNNPFDCENDITLN
PSELVSKQKEHKVTFSVENEGLDTRKSSMLGHQTFSCQNSLESPLAMYDNKNDNSDITSNIKEKGGIINNNSNNDDDDNNNNNDND
NDNNNSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNSNNIKNNVDNNNTNPADNIPTLSNEAFTPSQQFSQERVNNNADRCENSSF
TNVQQHFQAQTYKQMKKRRAQIQKNGGPGAALALALALALALALALGPGGWWEITHNTLFIPANETFDRPHNGYVDILPIGGYN
HLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFISTRKSSTERHVYYIDLRSPNEIIEVTDTSEDGVYDVSFSSGRRFGL
LTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDEFGKDILVNSYEILPNDFDET
LSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVASQLNAIVVVVDGRGTGFKGQDFRSLVRDRLGDYEARDQISAASLYGSLTFVDP
QKISLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRFYDSVYTERYMHTPQENFDGYVESSVHNVTALAQANRFLLMHGTGD
DNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIRYHNANVIVFDKLLDWAKRAFDGQFVKACYPYDVPDYAGYPYDVPDYAYPY
DVPDYA

The following list contains all the constructs, used for Part II. The signal sequence is marked in red. The prM domain is highlighted in blue, the E domain in green. Concerning the construct 2K-NS4B-V5, the NS4B domain is marked in violet, the V5-Tag in yellow.

>Insulin
MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSL
QKRGIVEQCCTSICSLYQLENYCN

>Vasopressin
MPDTMLPACFLGLLAFSSACYFQNCPRGGKRAMSDLELRQCLPCGPGGKGRCFGPSICCADELGCFVGTAEALRCQEENYLPSPCQ
SGQKACGSGGRCAAFGVCCNDESCVTEPECREGFHRRARASDRSNATQLDGPAGALLLRLVQLAGAPEPFEPAQPDAY

>C14-prM-E
MTAGMIIMLIPTVMAFHLTTRNGEPHMIVSRQEKGKSLLFKTEDGVNMCTLMAMDLGELCEDTITYKCPFLRQNEPEDIDCWCNST
STWVTYGTCTTTGEHRREKRSVALVPHVGMGLETRTETWMSSEGAWKHAQRIETWILRHPGFTIMAAILAYTIGTTHFQRALIFIL
LTAVAPSMTMRCIGISNRDFVEGVSGGSWVDIVLEHGSCVTTMAKNKPTLDFELIKTEAKQPATLRKYCIEAKLTNTTTDSRCPTQ
GEPSLNEEQDKRFVCKHSMVDRGWGNGCGLFGKGGIVTCAMFTCKKNMKGKVVQPENLEYTIVITPHSGEEHAVGNDTGKHGKEIK
ITPQSSITEAELTGYGTVTMECSPRTGLDFNEMVLLQMENKAWLVHRQWFLDLPLPWLPGADTQGSNWIQKETLVTFKNPHAKKQD
VVVLGSQEGAMHTALTGATEIQMSSGNLLFTGHLKCRLRMDKLQLKGMSYSMCTGKFKVVKEIAETQHGTIVIRVQYEGDGSPCKI
PFEIMDLEKRHVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKLNWFKKGSSIGQMIETTMRGAKRMAILGDTAWD
FGSLGGVFTSIGKALHQVFGAIYGAAFSGVSWTMKILIGVIITWIGMNSRSTSLSVSLVLVGVVTLYLGVMVQA

>2K-NS4B-V5
MANEMGFLEKTKKDLGLGSITTQQPESNILDIDLRPASAWTLYAVATTFVTPMLRHSIENSSVNVSLTAIANQATVLMGLGKGWPL
SKMDIGVPLLAIGCYSQVNPITLTAALFLLVAHYAIIGPGLQAKATREAQKRAAAGIMKNPTVDGITVIDLDPIPYDPKFEKQLGQ
VMLLVLCVTQVLMMRTTWALCEALTLATGPISTLWEGNPGRFWNTTIAVSMANIFRGSYLAGAGLLFSIMKNTTNTRRGKPIPNPL
LGLDST

C. Curriculum Vitae

Name:	Marco Janoschke
Date of birth:	08.11.1986
Place of origin:	Schwäbisch-Hall (Deutschland)
Nationality:	German

Education

2014 – 2018	PhD in biochemistry under the guidance of Prof. Martin Spiess; Biozentrum, University of Basel. Thesis title: The biogenesis of Membrane proteins at the Sec61 translocon and their integration into the lipid bilayer of the endoplasmic reticulum
2011 – 2014	Master of Science in Chemistry at the University of Freiburg (Germany)
2007 – 2011	Bachelor of Science in Chemistry at the University of Freiburg (Germany)
2006 – 2007	Military Service in the German Bundeswehr in Mittenwald
2003 – 2006	Wirtschaftsgymnasium (Öhringen), Abitur in 2006
1997 – 2003	Secondary school and Mittlere Reife (Realschule Öhringen)
1993 – 1997	Primary school (Hungerfeldschule Öhringen)

Languages

German	mother tongue
Englisch	fluent

Publications

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.

Steimle, S., Willistein, M., Hegger, P., **Janoschke, M**., Erhardt, H., and Friedrich, T. (2012). Asp563 of the horizontal helix of subunit NuoL is involved in proton translocation by the respiratory complex I. FEBS Lett. *586*, 699–704.