Doctoral Thesis in Biochemistry, Stockholm University, Sweden

Membrane protein topogenesis

Karin Öjemalm

Stockholm University

Membrane protein topogenesis

Karin Öjemalm

©Karin Öjemalm, Stockholm University 2015

ISBN 978-91-7649-089-1, pages 1-67

Printed in Sweden by Publit, Stockholm 2015
Distributor: Department of Biochemistry and Biophysics, Stockholm University

To Johan, Erik and Birk

List of publications

I. Hedin LE*, **Öjemalm K***, Bernsel A, Hennerdal A, Illergård K, Enquist K, Kauko A, Cristobal S, von Heijne G, Lerch-Bader M, Nilsson I, and Elofsson A (2010) Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context *J Mol Biol 396(1): 221-229*

II. **Öjemalm K***, Halling KK*, Nilsson I, von Heijne G (2012) Orientational preferences of neighboring helices can drive ER insertion of a marginally hydrophobic transmembrane helix *Mol Cell 45(4): 529-540*

III. **Öjemalm K***, Botelho S*, Stüdle C, von Heijne G (2013) Quantitative analysis of SecYEG-mediated insertion of transmembrane alpha-helices into the bacterial inner membrane *J Mol Biol 425(15): 2813-2822*

IV. **Öjemalm K**, Higuchi T, Jiang Y, Langel Ü, Nilsson I, White SH, Suga H, and von Heijne G (2011) Apolar surface area determines the efficiency of translocon-mediated membrane-protein integration into the endoplasmic reticulum *Proc Natl Acad Sci U S A 108(31): E359-364*

* Authors contributed equally

Additional publications

I. Enquist K, Fransson M, Boekel C, Bengtsson I, **Geiger K**, Lang L, Pettersson A, Johansson S, von Heijne G, and Nilsson I (2009) Membrane-integration characteristics of two ABC transporters, CFTR and P-glycoprotein *J Mol Biol 387(5): 1153-1164*

II. **Öjemalm K**, Watson HR, Roboti P, Cross BCS, Warwicker J, von Heijne G, and High S (2013) Positional editing of transmembrane domains during ion channel assembly *J Cell Sci 126(Pt 2): 464-472*

III. Gadalla SE, **Öjemalm K**, Lara P, Nilsson I, Ericsson C, Zhao J, and Nistér M (2013) EpCAM associates with endoplasmic reticulum aminopeptidase 2 (ERAP2) in breast cancer cells *Biochem Biophys Res Commun 439(2): 203-208*

IV. Wanngren J, Lara P, **Öjemalm K**, Maioli S, Moradi N, Chen L, Tjernberg, LO, Lundkvist J, Nilsson I, and Karlström H (2014) Changed membrane integration and catalytic site conformation are two mechanisms behind the increased Abeta42/Abeta40 ratio by presenilin 1 familial Alzheimer-linked mutations *FEBS Open Bio 4: 393-406*

V. Vargas-Uribe M, Rodnin MV, **Öjemalm K**, Holgado A, Kyrychenko A, Nilsson I, Posokhov YO, Makhatadze G, von Heijne G, and Ladokhin AS (2014) Thermodynamics of Membrane Insertion and Refolding of the Diphtheria Toxin T-Domain *J Membr Biol, in press*

VI. Goel, S, Palmkvist M, Moll K, Joannin N, Lara P, Akhouri R, Moradi N, **Öjemalm K**, Westman M, Angeletti D, Kjellin H, Lehtiö J, Hult AK, Olsson ML, von Heijne G, Nilsson I, and Wahlgren M RIFINs are Adhesins Implicated in Severe *Plasmodium falciparum* Malaria *Nature Medicine, in press*

Abstract

The membranes of cells are highly complex and heterogeneous structures that fulfill multiple vital tasks. They form thin barriers that seal out the environment, thus defining the cell's boundaries. They mediate the selective exchange of information and substances between the inside and outside of cells, thus making cellular life and survival possible and allowing fast adaptation to changing conditions. Not least importantly, they harbor key components of many essential processes such as the photosynthesis and respiration. Membranes are composed of a largely apolar lipid matrix densely punctuated with a large number of different proteins. These so-called membrane proteins usually span the lipid matrix and protrude out into the space on either side of the membrane.

Over millions of years of evolution, cells have developed an incredible machinery to facilitate the insertion of membrane proteins into the membrane. Our understanding of these machines and the insertion processes they mediate has reached a point where we have a very good picture of membrane protein biogenesis in various types of cells. However, more data still needs to be collected to completely comprehend the complex molecular mechanisms and the physical chemistry that underlies the different insertion processes.

The work presented in this thesis contributes to that understanding. More precisely, we have studied how weakly hydrophobic transmembrane elements of membrane proteins, which cannot spontaneously enter the largely apolar membrane matrices, are efficiently incorporated. Indeed, such elements are quite common in membrane proteins and our work has lead to the formulation of a novel mechanism for how they can be integrated into biological membranes.

We have also added to the understanding of the physical chemistry that underlies the membrane insertion process by systematically introducing non-proteinogenic amino acids into a membrane-spanning segment of a membrane protein and studying its membrane insertion capability.

Contents

Abbreviations

aaRS	Amino acid tRNA synthetase
°C	Degree Celsius
C-terminus	Carboxy terminus
ΔG_{app}	Apparent change in free energy in kilo calories / mol
ΔG_{pred}	Predicted change in free energy in kilo calories / mol
EM	Electron microscopy
ER	Endoplasmic reticulum
kDa	Kilo Dalton
Lep	Leader peptidase
mRNA	Messenger RNA
N-terminus	Amino terminus
PE	Posphatidylethanolamine
SA	Signal anchor
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SRP	Signal recognition particle
TM	Transmembrane
tRNA	Transfer RNA
Å	Ångström

Amino acids

Alanine	Ala, A		
Arginine	Arg, R	Methionine	Met, M
Asparagine	Asn, N	Phenylalanine	Phe, F
Aspartic acid	Asp, D	Proline	Pro, P
Cysteine	Cys, C	Serine	Ser, S
Glutamic acid	Gln, Q	Threonine	Thr, T
Glutamine	Glu, E	Tryptophan	Trp, R
Glycine	Gly, G	Tyrosine	Tyr, Y
Histidine	His, H	Valine	Val, V
Isoleucine	Ile, I		
Leucine	Leu, L		
Lysine	Lys, K		

Introduction

The lipid matrix of the membranes

The beginning of life is marked by evolutionary breakthroughs such as catalytic and autocatalytic reactions that lead to nucleotides, RNA and DNA, amino acids and proteins. One of those important breakthroughs was the emergence of a liposome system that eventually evolved to the membranes of the cells. What probably started as a simple double layer of lipids, forming vesicles that allowed the physical separation of a small space from the external environment, has become a highly complex, heterogeneous and functionally very important part of every cell. The membrane forms not only a barrier that seal out the cell's external environment and so defines its boundary, but also mediates the selective exchange of information and substances. Furthermore, membranes are the sites where key steps of many vital processes take place.

The membrane matrix is a double layer of amphipathic lipids that most solutes and water cannot pass. Entropically driven by the relaxation of the "water structure", known as the hydrophobic effect (Figure 1B), the lipid molecules arrange themselves with their long aliphatic fatty acyl chains facing each other, thus forming a bilayer. The chains make up the non-polar core of the membrane matrix that is at room temperature flexible and disordered (Figure 1C). The disorder originates from the high degree of rotational freedom of each bond within the saturated fatty acyl chains, and the order-disturbing, *cis*-unsaturated bonds of many of them (Figure 1D)[1]. The core spans the central half of the membrane[2] and is lined by two interfaces. They are composed of the backbones and the "headgroups" of the lipids (Figure 1D). The interfaces account for the other half of the membrane diameter[2] (Figure 1A). A large hydrogen-bond network strengthens the rather organized interfaces and stabilizes the lipid matrix but anyhow allows fast lateral diffusion of membrane components[1]. Both interfaces are hydrated and posses steep polarity-gradients along the membrane normal. They change from high-polarity and high-hydration at the outer borders, to low-polarity and dehydration at the inner border where they merge with the membrane core (Figure 1A)[3].

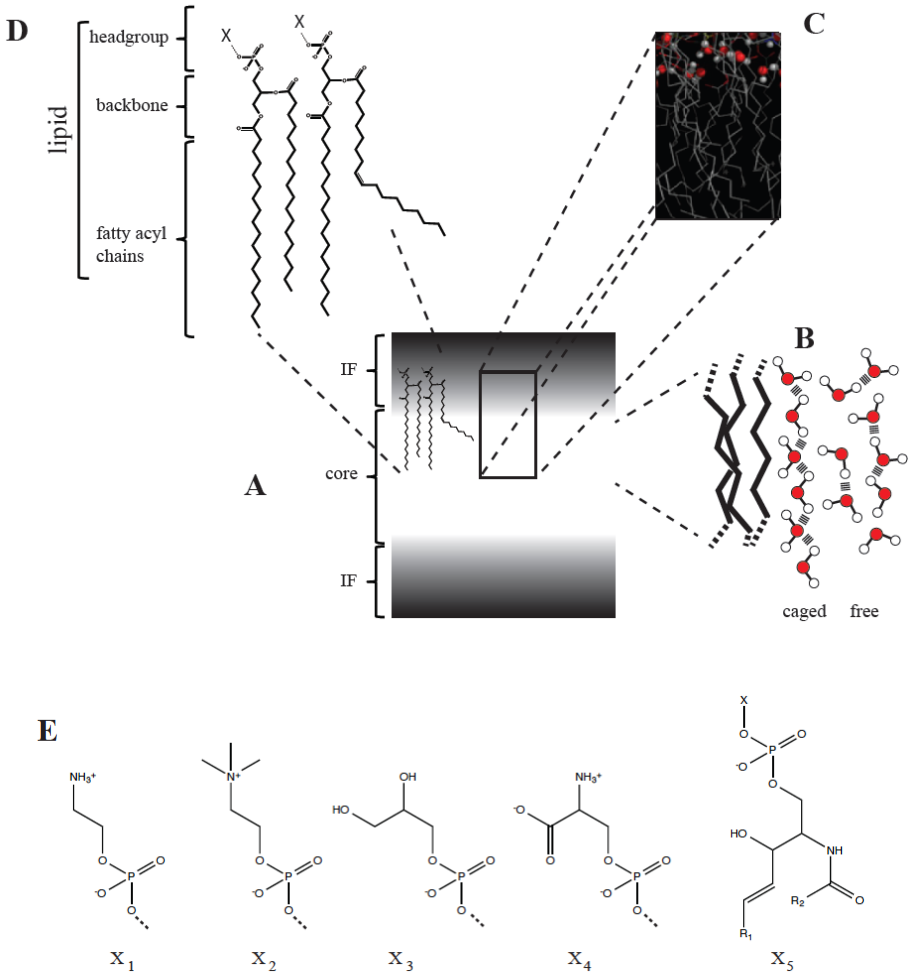

Figure 1: The membrane matrix. A) Schematic illustration of the lipid matrix of a membrane. The color gradient represents the polarity- and hydration gradients (black = polar, hydrated; white = apolar, dehydrated). The interfaces (IF) and the core are marked with brackets. B) Hypothetical exposure of fatty acid chains (black zigzag lines) to water. Water molecules cannot form hydrogen bonds to pure hydrocarbon chains. They need to form hydrogen-bond cages around the apolar moiety instead, which are less dynamic than free hydrogen bonds in bulk water. Thus there is an overall entropy loss and, depending on the size of the water-exposed apolar surface area, an additional change in enthalpy. The latter is a result of the reduced number of hydrogen bonds. To decrease the loss of entropy, biological systems minimize their apolar surface area exposed to water. That drives the formation of vesicles or bilayers. C) Section of a model membrane bilayer from a snapshot of a simulation. It shows the highly flexible fatty acid chains of fluid membranes (adapted from Johansson *et al*[4].) D) Single lipids drawn in their most extended conformation. The fatty acid acylated to the second carbon of the glycerol backbone is often *cis*-unsaturated causing a kink in the hydrocarbon chain. E) Common lipid headgroups including the phosphate group of the backbone: X_1 = ethanolamine, X_2 = choline, X_3 = glycerol, X_4 = serine, X_5 = the sphingosine backbone.

A large number of different lipid species can be found in the various membranes of prokaryotic and eukaryotic cells. Even a particular membrane is composed of several types of lipids. This lipid polymorphism is important for the physico-chemical properties of membranes and for the formation and maintenance of their particular shapes[5].
Typical bacterial and eukaryotic membrane lipids are either composed of a lipid headgroup linked by a glycerol backbone to two fatty acyl chains of 12-20 hydrocarbon groups (Figure 1D, E) or are composed of a lipid headgroup and a sphingosine-derived lipid backbone acylated with only one fatty acid (Figure 1E, X_5). The lipid headgroups are chemically diverse and may include dipoles, charges or sugar groups. They largely determine the lipid's geometrical and chemical properties[1].

The proteins of the membranes

Signal conduction as well as substance exchange through membranes, involving active and passive transport of ions, macromolecules and metabolites, are predominantly mediated by proteins. They are either an integral part of the membrane bilayer or are peripherally associated with it. Proteins involved in many important processes such as the photosynthesis and respiration, as well as proteins conducting diverse enzymatic reactions are located within membranes. Other membrane proteins are key players during organelle biogenesis and intracellular trafficking. The protein content in membranes is therefore usually high, being about equal the mass of the lipids[6]. It is estimated that the human cell contains about 5500 different types of membrane proteins[7], while *Escherichia coli* has about 900[8].
Because membrane proteins carry out various important functions and because of their location at the interface between the inside and outside of a cell, they are major drug targets. It is estimated that more than 50% of all human drugs affect membrane proteins[9]. In accordance with their importance for the cell as well as from a pharmaceutical perspective, it is crucial to understand membrane proteins, their structure, biogenesis and functions in eukaryotic as well as in prokaryotic (pathogenic) cells. However, their inherent property of being at least partially hydrophobic, to match the apolar membrane cores, makes membrane proteins more difficult to study. If isolated from their membranous environment they tend to aggregate, thereby losing structure and function.
During the last few decades the scientific community has overcome many of these difficulties by developing methods and assays adjusted to their special needs. This has led to a tremendously improved understanding of membrane protein biogenesis, of their structures and functions. Still, many unanswered questions remain.

Membranes

Bacterial membranes

A typical bacterial cell possesses one or two principal membranes, a cell membrane and - if present - an outer membrane, which together with an extracellular peptidoglycaneous cell wall, define the cell envelope. The membranes are of species-specific composition[10] and the peptidoglycan layer that gives the cell strength, rigidity and protection is of species-specific thickness and composition[11]. Besides their high protein content, the inner and outer membranes of the widely used bacterial model organism, the gram-negative *E. coli*, are composed mainly of the neutral phosphatidylethanolamine (PE, 70-80%, Figure 1E) and the anionic lipid species phosphatidylglycerol (PG, 20-25%, Figure 1E) and cardiolipin (CL, 5-10%)[12,13].
Anchored in the inner membrane of gram-negative bacteria such as *E. coli*, proton-transporting complexes of the electron transport chain accumulate protons in the periplasm. This transmembrane (TM) proton gradient, the proton motive force, drives various energetically unfavorable processes such as the ATP synthesis and the uptake of solutes against their concentration gradient via proton-coupled secondary transporters[14].

Eukaryotic membranes

The more complex eukaryotic cell is surrounded by a plasma membrane that is very rigid, densely packed and strengthened by an extracellular matrix composed of proteins, proteoglycans, glycoproteins and glycolipids (and the cell wall of plant cells). The plasma membrane accommodates a large number of membrane protein species and an array of different types of lipids asymmetrically distributed between the two leaflets of the bilayer[15]. The main lipid components of the plasma membrane are phosphatidylcholine (PC, Figure 1E), with one unsaturated and one saturated fatty acyl chain, PE, a high percentage of the sphingosine derived sphingomyelin (SM) and cholesterol[6,16]. The amphipathic and bulky structure of cholesterol makes the membrane less permeable to small molecules by supporting membrane packing of lipids with unsaturated fatty acid chains[17]. Cholesterol molecules also loosen the packing of saturated lipid patches[1]. SM, having two saturated hydrocarbon chains which allow tighter packing of the membrane core, is enriched in the outer leaflet of the plasma membrane[18]. SM and cholesterol are the main lipid components of “lipid rafts”, membrane patches found in the outer leaflet of the bilayer. They are thicker and more stable than the rest of the plasma membrane and harbor a distinct subset of integral membrane proteins[19]. Lipid rafts can sequester membrane proteins to activate or inhibit

them. A demonstrative example is the P2X family of ATP-gated ion channels. Members of the P2X family are localized to lipid rafts and for $P2X_1$ this association is absolutely essential for its activation[20] while for $P2X_7$ it has a strong inhibitory effect[21].
The complexity of eukaryotic cells arises from the internal, membrane bound compartments including organelles such as mitochondria and chloroplasts, and the large network of interconnected tubules and sheets of the endoplasmic reticulum (ER). The ER extends through the whole cell in higher eukaryotes and is continuous with the nuclear envelope[22]. It is very dynamic, tubules form and disappear and the sheets reorganize constantly[23]. The rough ER forms sheet-like structures with assotiated ribosomes and is the place where membrane and secretory proteins are synthesized[22].
The lipid matrix of the ER membrane differs substantially from that of the plasma membrane in its composition and the symmetric distribution of the lipids between the two leaflets of the bilayer[24]. The ER membrane is significantly thinner than the plasma membrane[25]. The lipid matrix consists of more than 50% phosphatidylcholine (PC), a high fraction of PE, phosphatidylinositol (PI) and some percentage phosphatidylserine (PS, Figure 1E)[6]. It contains essentially no cholesterol, even though it is synthesized in the ER. The lack of cholesterol increases the fluidity and loosens the packing of the membrane. This makes the ER membrane more flexible and allows hydrophobic matching of its apolar core to the hydrophobic surface of the TM domains of membrane proteins[26]. This adaptive property of the ER membrane makes it well suited to accept the different TM elements, the TM α-helices, of freshly synthesized membrane proteins and provides any type of membrane protein an adequate environment for folding to its native structure. Eukaryotic membrane proteins are exclusively synthesized at the rough ER with the exception of mitochondrial and chloroplast membrane proteins[27].
Membrane proteins undergo co- or posttranslational modifications in the ER before they are distributed to their final destinations by vesicular transport via the Golgi apparatus.

Membrane protein topology and structures

Approximately 20-30 % of all genes in eukaryotic and prokaryotic cells encode proteins with one or more TM domains[8]. These domains generally adopt one of two structural elements: hydrophobic α-helices or β-barrels with hydrophobic outer surfaces[28,29] (Figure 3), both stabilized by hydrogen-bonding of the peptide backbone. The secondary structure formation partially dehydrates the polar peptide backbone and thereby lowers the energetic cost for its partitioning into a membrane[30-32]. Combined with the free energy gain of the solvation of apolar amino acid side-chains, located within TM

stretches, in the hydrophobic core of the membrane[33], secondary structure formation enables polypeptide chains to stably span membranes.
There is a large topological variety within the group of "α-helical" membrane proteins. They range from single- to multi-spanning membrane proteins with more than 30 TM α-helices[7]. The latter usually assemble into helix-bundles. α-helical membrane proteins can be further grouped according to the location of their N-termini relative to the membrane and whether or not they have a cleavable signal sequence[34] (Figure 2, for signal sequence see next paragraph).

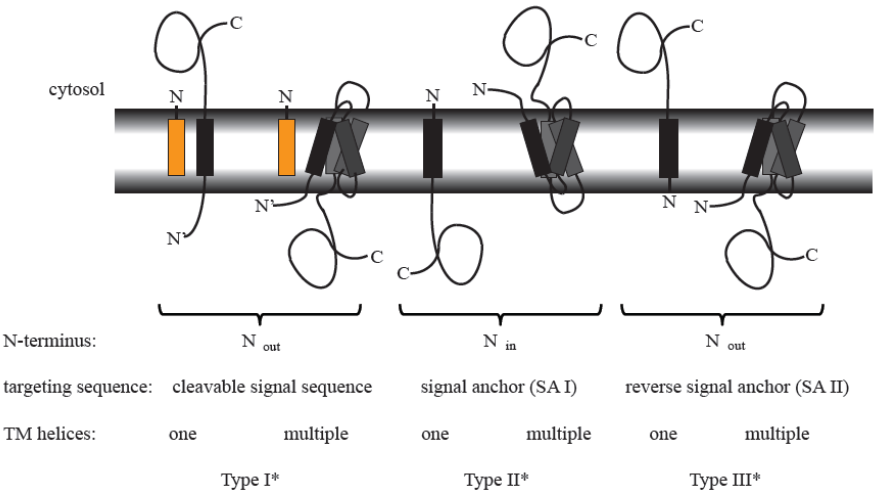

Figure 2: Grouping of membrane proteins according to the location of their N-termini, the possession of a signal sequence and their type (single-spanning or multi-spanning). Note that the designation "signal anchor" and "reverse signal anchor" is also used for TM helices following the targeting sequences in multi-spanning membrane proteins.
* denotes the types of TM helices according to Higy *et al*[34]

The structures of α-helical membrane proteins are as diverse as their functions despite their common structural element, the TM α-helix. They range from small, predominantly α-helical proteins such as GlpG, (Figure 3A) which is mostly buried in the membrane, to proteins with large cytosolic or non-cytosolic domains. The functional unit of α-helical membrane proteins is often homo- or hetero-oligomeric, adding to the structural as well as the functional complexity. These proteins are present in all types of membranes across all kingdoms of life[35].

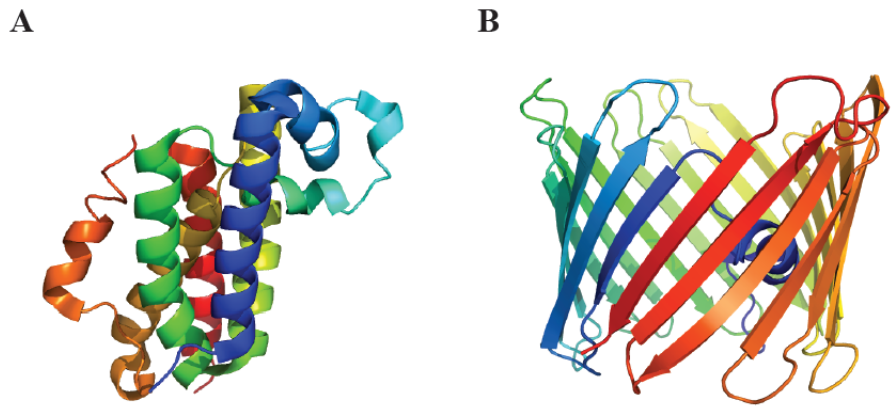

Figure 3: Representative examples of and an α-helical and a β-barrel membrane protein. A) The intramembrane protease GlpG from the inner membrane of *E. coli* is a small, predominantly α-helical membrane protein. B) The voltage dependent anion channel VDAC is a mitochondrial β-barrel membrane protein with a 19-stranded β-barrel. Note that it features two parallel β-strands (N-terminal strand labeled in blue, C-terminal strand labeled in red) which is uncommon.

Proteins with a membrane spanning β-barrel domain, which is an assembly of multiple antiparallel β-strands, belong to the second large group of membrane proteins (Figure 3B). They are found in the outer membranes of mitochondria, chloroplasts and gram-negative bacteria[36]. These membranes have no proton motive force and β-barrel membrane proteins form mostly passive pores. However, some are receptors or enzymes. Others are selective transporters that move the substrates against their concentration gradient, energized by inner membrane proteins. β-barrel proteins are also part of the secretion systems of gram-negative bacteria[29].

Membrane protein topogenesis and topogenic determinants

The biogenesis of α-helical membrane proteins can be divided into different stages and for the purpose of this thesis the two-stage model[37] seems the most adequate. During the initial stage, membrane proteins are synthesized by the ribosome, their TM domains are integrated into the target membrane and the proteins adopt their topologies (topogenesis). As has been stated already in the early 1980's "the topology is a result of topogenic sequences, permanent or transient features within the polypeptide chain that are shared

by many, structually different, proteins with a common topogenesis"[38]. During the second stage, which may or may not be temporally separated from topogenesis, membrane proteins fold and adopt their native structures. They may also aquire co-factors or prostetic groups, get enzymatically covalently modified or undergo side-chain reactions such as disulfid-bridge formation. The following paragraphs are dedicated to the first stage, the topogenesis of membrane proteins, that is the focus of this thesis.

Targeting to the membrane

Mature mRNAs encoding membrane proteins or soluble proteins destined for non-cytosolic locations engage the ribosome just like mRNAs of cytosolic proteins. Unlike them however, relatively short, cleavable **signal sequences** of 7-15 residues[39] or somewhat longer, non-cleavable, **signal anchors** of 19-27 residues[39] are usually encoded in the 5' ends of mRNAs of membrane or soluble non-cytosolic proteins[40,41]. Signal sequences are generally composed of three structural regions: a central stretch of hydrophobic residues, the h-region that is flanked by a n-region with a high density of positively charged residues and a polar c-region that includes a signal peptidase cleavage site[42] (Figure 5). When translated by the ribosomes and emerging from their exit tunnels, signal sequences and signal anchors target the ribosomes to the ER membrane in eukaryotes and to the plasma (inner) membrane in bacteria and archaea[41,43,44] via the co-translational SRP-dependent (**S**ignal **R**ecognition **P**article) pathway[45,46] or the, predominantly prokaryotic, SRP-independent, post-translational pathway[47] (Figure 4). During co-translational targeting, the ribosome is transferred from SRP to a highly conserved, membrane anchored protein-conducting channel, the Sec translocon[48] that is ubiquitous in all three kingdom of life[49].

The most important determinant of which route the ribosome takes to the membrane in bacteria is the hydrophobicity of the signal sequence or the signal anchor of the protein it just synthesizes[50]. According to this idea, a hydrophobic non-cleavable signal anchor, thus the most N-terminal TM α-helix of an inner membrane protein, strongly interacts with SRP. The ribosome is therefore funneled into the SRP-dependent pathway. Less hydrophobic signal sequences of secreted proteins do not efficiently interact with SRP and therefore take an alternative pathway. Besides the hydrophobicity, the helix-propensity of signal sequences[51], positively charged residues in the n-region[52] and even the length of the polypeptide chain[53] have been proposed to influence signal sequence recognition by SRP. An estimated 25% of all signal sequences of proteins in *E. coli* interact with SRP[54].

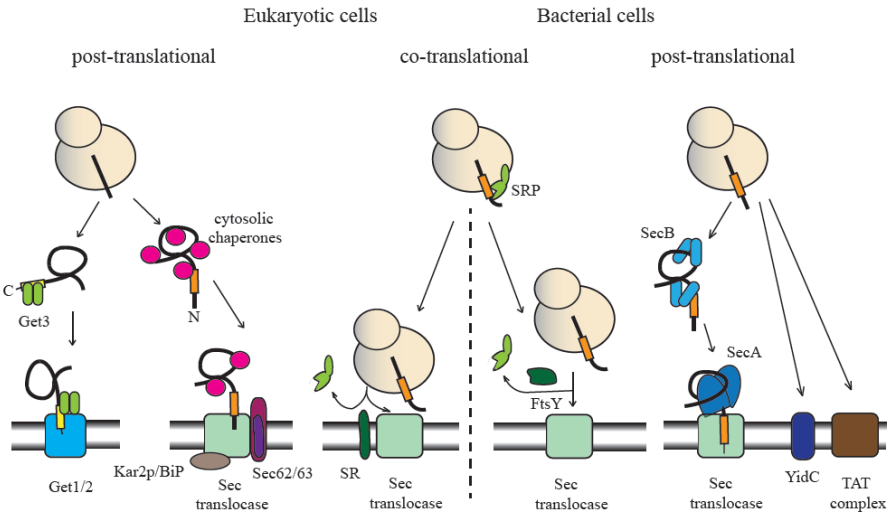

Figure 4: Different routes to the ER membrane and to the inner membrane of *E. coli*. Middle: SRP binds to the translating ribosome and targets it to the ER membrane in eukaryotic cells or to the inner membrane in *E. coli*. The ribosome is transferred to the Sec61 translocon upon SRP-SR interaction (**S**RP **R**eceptor) at the ER membrane and subsequent GTP hydrolysis. The membrane-targeted ribosome in *E. coli* is transferred to the SecYEG translocon upon SRP (Ffh)-FtsY interaction and GTP hydrolysis. Right: if a signal sequence does not efficiently interact with SRP, protein synthesis is completed in the cytosol. The nascent polypeptide chain is kept unfolded by SecB dimers and is targeted to the inner membrane via SecA. SecA helps to post-translationally translocate the polypeptide chain through SecYEG, energized by ATP hydrolysis. Alternatively membrane proteins can be inserted into the membrane via the insertase YidC. Already folded proteins can be translocated through the inner membrane via the TAT translocase. This pathway requires the proton motive force. Left: the polypeptide chain of a secretory protein in yeast, completely synthesized in the cytosol, is kept unfolded by cytosolic chaperones. It is post-translationally targeted to the Sec61-Sec62/63 complex and pulled through it by Kar2p (BiP) energized by ATP hydrolysis. Likewise, tail-anchored proteins, being proteins with a single TM sequence at their very C-termini, are synthesized completely in the cytosol. "Pretargeting factors" receive the C-terminal TM sequences emerging at the ribosomal exit tunnels and guide the proteins to a Get3 dimer. Upon ATP hydrolysis, the TM sequences are transferred to the membrane bound Get1/2 complex and are inserted into the ER membrane.

Because higher eukaryotes mainly use the SRP-dependent pathway, there is no need for signal sequences/anchor discrimination. However, they are generally variable within the frame of the three structural regions[42,55]. Their affinity for SRP correlates with the length of their h-regions while being fine-tuned by positively charged residues flanking it[56].

The bacterial and yeast post-translational pathways are mostly used for secretory proteins and are not further discussed here, and neither are the Sec-translocase independent Tat- (Twin Arginine Translocation pathway)[57] and YidC pathways[58]. Likewise, archaeal co- and post-translational translocation and membrane insertion pathways are excluded from the discussion below because they are still partially unknown[44].

Signal sequences

An evolutionally conserved, heterotrimeric protein complex, called Sec61αβγ in eukaryotes, SecYEG in bacteria and SecYEβ in archaea is the functional core of the Sec translocation machinery. It interacts with a number of accessory proteins that modulate translocation, aid membrane integration of nascent polypeptide chains or covalently modify them (see also "the Sec-translocon machinery"). The pore of the Sec61α/SecY cannel subunit opens not only towards the cytosolic and non-cytosolic compartments but can also open laterally to release TM α-helices into the lipid bilayer via a lateral gate[48] (see also "structures of the bacterial and eukaryotic translocons").

Once targeted to the membrane, signal sequences or anchors of nascent polypeptide chains are inserted into the pore of Sec61α/SecY. They adopt one of two orientations relative to the membrane: their N-termini can be oriented towards the cytosolic side (N_{in}) or towards the non-cytosolic side of the membrane (N_{out}) (Figure 2). Signal sequences always adopt a N_{in} orientation. Their non-cytosolic C-terminus, which contains a signal peptidase cleavage site[59], is recognized by the signal peptidase (complex) and the signal sequences are subsequently cleaved off[41]. Non-cleavable, targeting signal anchors can adopt either orientation, forming signal anchors with an N_{in} (SA-I, signal anchor) or an N_{out} orientation (SA-II, inverted signal anchor)[60].

Topogenic determinants of signal sequences and anchors

Several features near the beginning of a nascent polypeptide chain determine the orientation of the signal sequence or anchor relative to the membrane. The most important is the "positive-inside rule", positively charged residues that are enriched in the n-region and depleted in the c-region of signal sequences[61]. Their dominance at one end of a signal anchor strongly favors its cytosolic location[61,62].

The orientation of membrane-targeting sequences depends further on their helix length and hydrophobicity. Re-orientation of long or very hydrophobic signal anchors within the translocon is impaired and they therefore insert with their N-termini to the non-cytosolic side[63-65]. Further, the N-terminus that precedes the first TM sequence has to be unfolded to become translocated[66]. A long N-terminal domain may fold prior to the synthesis of a signal anchor and so prevent N-terminus translocation through the membrane. Finally, in eukaryotes, covalent modification of the N-termini e.g. by addition of oligosaccharides (see also "the Sec-translocon machinery") effectively inhibits backsliding or re-orientation, thereby locking the N-termini of these

proteins in a specific orientation[67].

Multi-spanning membrane proteins

The signal sequence or the orientation of the first TM helix released through the lateral gate of the Sec translocon has a predominant decisive role for the topology of a membrane protein. That is because TM helices within the same polypeptide chain must adopt alternating orientations relative to the plane of the membrane. This is known as the sequential mode of topogenesis[38,68]. There is however evidence that at least a subset of membrane proteins follows a non-sequential mode. Removal of the targeting TM helix of the *E. coli* sugar transporter MalF for example does not change the topology of the whole protein within the bacterial inner membrane[69].

The topogenesis of multi-spanning membrane proteins involves an orchestrated recognition of TM helices, their release into the membrane via the Sec61α/SecY lateral gate either sequentially[70], in pairs[71], or even in bundles[72-75] as well as the retention of cytosolic and the translocation of extracytosolic domains.

Sequence intrinsic features of TM α-helices

α-helix formation of TM sequences prior to their membrane insertion is a prerequisite[32,76]. It can take place already in the ribosomal tunnel[76-79] or at the latest in the lateral gate of the translocon. The predisposition of a sequence within a polypeptide chain to adopt a helical secondary structure depends in general on its amino-acid composition and more precisely it depends to a substantial part on the chemical nature of the side-chains. Amino acids with a high tendency to form α-helices are e.g. alanine, leucine, methionine or lysine. The bulky and on the first side chain hydrocarbon group branched valine, threonine and isoleucine, as well as glycine that has no side chain, have a low tendency to form helices[77,80]. Proline is a well known "helix breaker" because of its cyclic side chain that involves the amino group of the peptide backbone. This cyclisation prohibits peptide backbone hydrogen-bonding and thus local α-helix formation. It is the reason why proline is energetically costly within TM helices even though its side-chain is purely aliphatic[81,82].

The insertion efficiency of TM helices into membranes critically and predictably depends on their hydrophobicity and their length[82,83]. The more hydrophobic residues (leucine, isoleucine, valine, phenylalanine) a potential TM helix has and the longer the helix is, the higher its membrane insertion probability[33,83]. Thus the contribution of residues within the TM helix to its hydrophobicity and therefore to its membrane insertion probability is in general additive. An exception to this rule is a cluster of large and bulky leucines within a TM helix that seems to influence its membrane-insertion

probability in a non-additive way[84]. Charged or polar residues, especially when located within the central half of a TM helix, increase the energetic cost for the membrane insertion of that helix and thus decrease its insertion probability. Such "costly" residues have to be counterbalanced by apolar residues within the same helix in order to enable its membrane integration[31,82].

The hydrophobic moment of TM helices, a measure of the amphiphaticity of a helix, has been shown to inversely correlate with the insertion probability: the lower the moment, the higher the insertion probability[82].

Tryptophan and tyrosine, when located at the extremes of TM helices, can form favorable hydrogen bonds with the charged and polar backbone and headgroups of the lipids of the membrane while intercalating with their hydrocarbon chains[4,32,85]. These interactions lower the energetic cost for insertion of such helices into membranes[82].

While it is generally accepted that hydrophobicity, and thus the hydrophobic effect, is the main driving force for the integration of TM helices into membranes, its quantitation is still a matter of debate. Several experimentally as well as computationally derived "hydrophobicity scales" have been established[86]. These scales assign a free energy contribution for TM helix insertion into a membrane core to any proteinogenic amino acid when located in the center of a TM helix. Hydrophobicity scales correlate in general well but experimentally and computationally derived scales can differ in magnitude. An extensively discussed example is arginine for which the experimentally derived "biological hydrophobicity scale"[82] assigns about 2.5 kcal/mol. Energetic cost calculations, derived from simulations of the insertion of arginine analogues into a membrane core, are in the range of 15-20 kcal/mol[87,88]. This discrepancy results at least partially from the possibility of arginine in the context of a TM helix located in a natural membrane to snorkel with its terminal side-chain guanidinium group towards the more polar interfaces[89], the simplification of the model bilayer used for simulations compared to actual membranes of cells[88] and the exclusion of the translocon[89].

An extended hydrophobicity scale, the "position-specific biological hydrophobicity scale"[83], mirrors very accurately the statistical distribution of the different types of amino acid within TM helices of membrane proteins which have been crystalized and structure-solved[83,90].

Sequence extrinsic features of TM α-helices

In general, there is a significant positive charge bias between the cytosolic and external parts of membrane proteins that correlates with the membrane protein topology[61,62,91].

Positively charged residues can locally affect the co-translational insertion of TM helices by either supporting it[92,93] when located in their cytosolic flanking polypeptide regions, or by prohibiting it when located in their non-cytosolic flanking polypeptide regions. The latter yields in "topological frus-

trated helices"[94]. Positively charged residues may also have a global effect on membrane protein topogenesis exemplified by the small multidrug transporter EmrE from the inner membrane of *E. coli* that spans the membrane four times. It belongs to the group of dual-topology membrane proteins that have an amazing polypeptide-inherent property that allows their single gene product to adopt two different orientations within the membrane[95,96]. These, often small, multi-spanning membrane proteins have in general very few positively charged residues and a small charge bias between the external and cytosolic loops[95,97]. Manipulating the charge bias of EmrE by strategically placing a single positively charged residue into one of the short loops connecting the TM helices results in EmrE protein species with a single, fixed topology (either C_{in} or C_{out}) even if the added positively charged residue is located at its very C-terminus[97]. Finally, positively charged residues within the polypeptide chain slow down the rate of the nascent chain movement through the ribosome and, to some extent, through the translocon[98].

A TM helix may undergo helix-helix interaction with a subsequent TM helix during the membrane integration step. These interactions include hydrogen bonds[99,100] and electrostatic[101] interactions[102] between side-chains of the two partner helices. This is an extrinsic property of a particular TM helix in the sense that an interaction-partner helix has to be encoded within the same polypeptide chain. Such helix-helix interactions can lower the energetic cost of the insertion of TM helices into membranes and may be the first step of protein folding.

In the most extreme case only a few residues separate the two TM helices. It has been shown that a single helix breaking residue, e.g. proline, can convert an α-helix composed of 30 leucines that spans the membrane once into a "helical hairpin" that spans the membrane twice if introduced into the center of the poly-leucine stretch[103]. Likewise, a cluster of aspartates or lysines, located in the cytosolic C-terminal flank of a very long hydrophobic helix, can convert it into a helical hairpin with a lumenal turn. In contrast, such charge clusters impede helical hairpin formation of a long helix when present in a lumenal C-terminal flank[104]. Charged, polar or weakly hydrophobic residues within the very short loop connecting two TM segments facilitate helical hairpin formation[103,104]. The helical hairpin is a common structural element of membrane proteins[96] and can form already within the ribosomal exit tunnel[79].

In eukaryotes, co-translational addition of oligosaccharides to glycosylation acceptor sites, mediated by the oligosaccharyltransferase complex (see also "the Sec-translocon machinery"), can lock loops or domains within the ER lumen[34,67]. This polypeptide chain modification inhibits the re-orientation of TM helices within the membrane and the re-location of translocated polypeptide sequences to the ER membrane or cytosol. They may also induce forward movement during translocation[105].

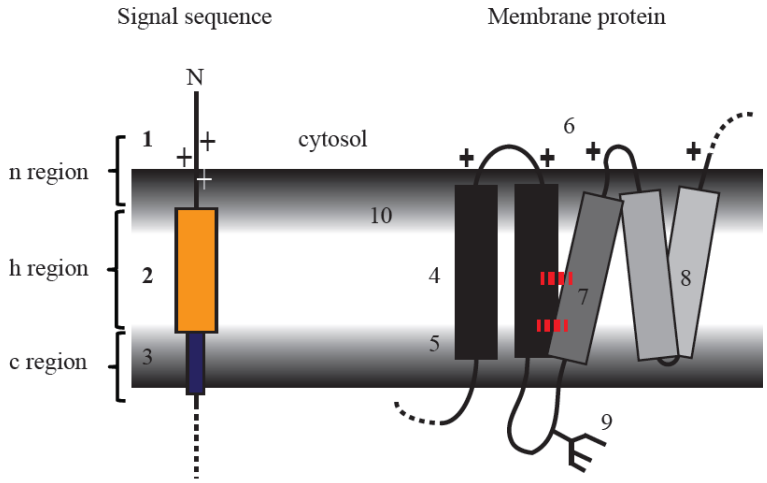

Figure 5: Summary of sequence features important for signal sequence and membrane protein topogenesis. 1) Positively charged residues within the n-region (the positive-inside rule); 2) hydrophobicity, helical propensity and length of the signal sequence; 3) the polar c-region including a signal peptidase cleavage site; 4) hydrophobicity, helical propensity and length of a TM sequence, hydrophobic moment and distribution of polar and charged residues within the TM sequence; 5) tyrosine and tryptophan residues in the TM helix extremes; 6) positive-inside bias and positively charged residues in cytosolic flanking regions of TM helices; 7) electrostatic and hydrophobic helix-helix interactions and inter-helical hydrogen bonds; 8) helical hairpins; 9) covalent modification of lumenal loops such as glycan attachment (eukaryotes only); 10) lipid composition of the membrane (in bacteria).

Marginally hydrophobic TM helices

While TM helices of single-spanning membrane proteins are predominantly "ideal", i.e. α-helical and clearly hydrophobic[83], 20-30% of TM helices from multi-spanning membrane proteins are "marginally hydrophobic"[28,33,96]. They lack the most important feature for efficient membrane insertion, namely sufficient hydrophobicity, are often short, kinked or broken α-helices or have in general low α-helical propensities[28,33]. These helices are often functionally relevant[106] and/or involved in inter- and intra-molecular assembly steps within membranes[107]. For such helices, sequence-extrinsic features play an important role for their Sec translocon-mediated recognition and membrane insertion.

Positively charged residues in cytosolic flanking loops of marginally hydrophobic TM helices increase the insertion efficiency in a distance- and density-specific manner[92]. They can also act over long distances: they may induce the retrograde movement and subsequent membrane anchoring of already translocated marginally hydrophobic helices even when placed up to 60 residues C-terminal to that helix[108].

Electrostatic interactions between TM helices of the same polypeptide

chain[101,109], specific hydrogen bond formation between two TM helices[99] as well as hydrophobic packing of neighboring TM helices[102] have been shown to reduce the energetic cost of inserting marginally hydrophobic TM helices into membranes. However, these helix-helix interactions might be distance-dependent as has been shown for the weakly hydrophobic TM helix 2 of the human erythrocyte anion exchanger (band 3 protein). This TM sequence only inserts efficiently into the ER membrane if the loop connecting it with its preceding TM helix is short i.e. 11 residues but not when it is as long as 200 residues[102]. Helix hairpin formation can lower the energetic cost of membrane insertion of the helices involved, especially if they are marginally hydrophobic[110]. A strong preference of two adjacent TM helices to adopt the same orientation within the ER membrane can force a hydrophilic segment, a part of the loop connecting the two TM helices, to become membrane-spanning[111,112]. Examples of the membrane insertion of helices constrained by neighboring TM helices with strong orientational preferences in a natural context are the marginally hydrophobic TM helix 9 of the human drug efflux pump p-glycoprotein that requires the presence of TM helices 8 and 10 to stably anchoring the membrane[93] or TM helix 3 of the human erythrocyte anion exchanger that first translocates into the ER lumen but then is pulled back into the membrane by the subsequent, hydrophobic TM helix[112].
Last but not least, marginally hydrophobic TM helices have a membrane insertion potential that is less, or even far less than 100%. They can be responsible for the generation of multiple topologies of a single gene product[96].

Post-translational topogenesis of membrane proteins

Protein topogenesis is not restricted to the co-translational insertion of TM helices into the membrane. On the contrary, more and more examples of post-translational rearrangements of protein parts relative to the membrane during folding and oligomerization are being found. A pioneering study in the mid 2000's showed that the water channel aquaporin 1 inserts with a four TM helix topology into the ER membrane but undergoes a large post-translational topological rearrangement. The second TM helix inverts its orientation relative to the membrane and pulls two adjacent membrane-excluded sequences into the membrane. They become TM helices 2 and 4[70,113]. Likewise, TM helix 8 of the glutamate transporter homolog GlpT from the archaeon *Pyrococcus horikoshii* gets post-translationally pulled into the membrane upon folding or oligomerization[114]. Specific TM helices may shift position relative to the membrane during folding or oligomerization[115]. Finally, TM helices may post-translationally acquire a large tilt relative to the plane of the membrane[116].
An interesting example of a post-translational rearrangement is the TCRα

chain of the αβT cell receptor, a single-spanning membrane protein with a moderate hydrophobic signal anchor sequence. It translocates post-translationally from the ER membrane to the lumen if it fails to assemble within a certain timeframe with its co-receptor, the CDδε dimer, to form a stable αβT cell receptor[107].
The membrane itself may be able to modulate the final topology of a membrane protein as demonstrated for the *E. coli* inner membrane protein lactose permease (LacY) that spans the membrane 12 times. The orientation of the first half of the protein relative to the membrane can be reversibly inverted by drastic changes in the inner membrane lipid composition[12].

Structures of the bacterial and eukaryotic Sec translocons

The first crystal structure of a translocon, the heterotrimeric SecYEβ from the archaeon *Methanococcus jannashii*, published in 2004[48] (Figure 6), revealed a protein channel composed of one copy each of SecY, E and β. The largest subunit, SecY that spans the membrane with 10 TM helices, formed the actual channel. Its polypeptide chain folded into a N-terminal (TM helices 1-5) and a C-terminal half (TM helices 6-10), pseudo-symmetrically arranged and stabilized on three membrane-spanning sides by the two single-spanning TM proteins SecE and Secβ. The fourth side, the "protein front" composed of the TM helices 2b, 3, 7, and 8 was "open", thus not interacting with SecE and Secβ and seemed therefore to be the, in 2004 only postulated, **lateral gate**. The aqueous pore of SecY was constrained by a central hydrophobic ring made up of six apolar side chains, yielding an overall hourglass-shaped pore structure. One of the most surprising features of the 2004 Sec structure was the "plug domain" a short, distorted helical sequence that extended from the non-cytosolic side of SecY about halfway through the pore. It seemed to seal the channel from the external side.
Today, in the beginning of 2015, we have a much broader understanding of the structure and the structure-function relations of the Sec translocases thanks to several high-resolution crystal and well-resolved EM structures of prokaryotic and eukaryotic translocons. They provide snapshots of the idle, "closed" translocon[48], the "primed" translocon that interacts with SecA[117] or with a SecA mimic[118], with the ribosome[119-123] or with a nascent chain mimic[124]. Other structures show the translocon in action, thus visualizing the translocon-ribosome-nascent chain complex (translocon-RNC) at different stages during protein translocation. In those structures, the Sec translocases harbor either undefined polypeptides or arrested polypeptide chains that position specific hydrophobic or hydrophilic sequences within the pore. Their lateral gates are largely closed, partially open or completely open[121-

[123,125,126]. One recent cryo-EM structure of the ribosome-SecY complex, that carried a multi-spanning membrane protein intermediate of proteorhodopsin, captured the translocon in a conformation where the lateral gate was pre-open. What separates this structure from others are the first two TM helices of proteorhodopsin, which were already released through the lateral gate and located just adjacent to SecY[127].

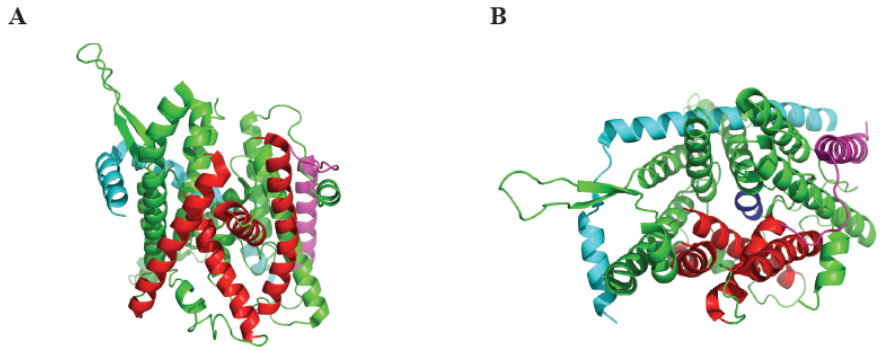

Figure 6: Structure of the closed archaeal Sec translocon. SecY is shown in green, SecE in light blue and Secβ in pink. A) Visualization of the ”protein front” with the lateral gate helices colored in red. B) View from the top (cytosolic side) with the plug domain colored in dark blue.

Even in the absence of a nascent chain, the ribosome binds Sec61α/SecY with high affinity[128], by interacting with **the cytosolic loops** of the Sec-translocase, namely the loops between TM helices 6 and 7 and between TM helices 8 and 9[119,120,123,126,129]. In eukaryotes, these loops as well as the N-terminal helix of Secγ specifically interact with the 28S rRNA of the large subunit of the ribosome, and with the ribosomal proteins uL23 and eL29[123].
Structure comparisons confirm that the translocon can undergo a large rigid body motion from a closed and sealed to a laterally open state during which the narrowest point of the hourglass-shaped pore, **the constriction ring**, changes its diameter from 4 Å up to 12-14 Å[122,124]. During translocation, the pore of the laterally closed or slightly open channel seems to be large enough to accommodate an unstructured polypeptide chain[123]. However, locking the lateral gate by disulfide crosslinking prevents post-translational translocation in *E. coli*[130].
Even with the lateral gate partially open, **the plug domain**, although displaced, still occludes the external funnel of Sec61α/SecY in the absence of a translocating polypeptide chain[117,124] or in the presence of a signal sequence that occupies the cytosolic vestibule[119,121]. A TM helix lining the open lateral gate also displaces the plug domain only within the translocon[122]. These

observations are experimentally supported by the immobilization of the plug domain within the translocon that does not affect translocation in bacteria[131]. Crosslinking the plug domain to SecE however results in a constitutive "open" conformation of the translocon and is lethal to *E. coli*[132]. It causes massive ion and water flux through the channel[133]. Interestingly, deletion of the plug domain does not inactivate the yeast[134] or the bacterial translocon, possibly because alternative plugs are formed[135]. Mutations in the plug domain or the constriction ring of the yeast Sec61α, known as prl mutations[136], lower the fidelity of signal sequence recognition[136,137]. This might be because of an impaired plug domain-induced stabilization of the translocon's closed conformation[135].

The Sec-translocon machinery

Sec61αγ and SecYE constitute the essential, functional and highly conserved core of the translocon. The Sec61β and SecG subunit, which have a structural role for their respective translocons, are not essential and are non-related proteins[47]. Reconstituted in proteoliposomes, the *E. coli* SecYEG is sufficient for post-translationally anchoring hydrophobic model TM helices in a liposome bilayer[138] and together with SecA is sufficient for post-translational protein translocation[139]. Likewise, reconstituted into proteoliposomes, Sec61αβγ is sufficient for the membrane integration of proteins[140] and together with the membrane protein TRAM (for TRAM see next paragraph) is the complex sufficient for translocation of secretory proteins into proteoliposomes[45].
Even though functional, the translocon core complex does not work alone. Several proteins and protein complexes have been identified that are in tight association, and act in concert with the eukaryotic and bacterial Sec-translocases.

The eukaryotic translocon

The Sec61 complex interacts with the multi-spanning membrane protein **TRAM** (**T**ranslocating chain **A**ssociated **M**embrane **P**rotein) that facilitates efficient translocation[141], aids the insertion of marginally hydrophobic TM helices[72,142] and supports membrane protein folding[74,143].
The ER-membrane anchored **oligosaccharyltransferase complex** (OST complex) is composed of 8 integral membrane proteins[144-146]. The OST complex scans polypeptide chains emerging at the lumenal pore exit of Sec61. It co-translationally transfers high-mannose oligosaccharides from a dolichol carrier to NX[S,T]z acceptor sites (X, z ≠ proline[147]) in newly synthesized

polypeptides of secretory and membrane proteins[146] if the sites are not located at their very C-termini[147]. The active site of the catalytic subunit Stt3 is located 30-40 Å away from the Sec61 lumenal pore exit, because the OST complex is located at the "back" of the translocon. It interacts with Sec61γ and the C-terminal part of Sec61α[144]. The OST complex can therefore transfer oligosaccharides only onto polypeptide chains if the asparagines of the glycosylation acceptor sites are at least 12-14 residues away from an N_{out} oriented and 10-12 residues away from an N_{in} oriented TM helix[39]. Recently, two sub-populations of the Sec61 translocon complex were identified: one that includes the large OST complex and one that does not[144].

Not so much is known about the hetero-tetrameric **TRAP** complex (**TR**anslocon **A**ssociated **P**rotein complex)[148] that simultaneously and stably interacts with the back of the translocon and with the OST complex[129,144]. It seems to affect translocation of positively charged residues within a polypeptide chain through the translocon and possibly modulates membrane protein topogenesis[149].

The **signal peptidase complex** (SPC) with a catalytic serine-protease subunit is the third complex associating with the Sec61 core complex. It cleaves off signal sequences of freshly synthesized pre-proteins[150,151].

Other integral membrane proteins found in close proximity to Sec61 in higher eukaryotes are **PAT-10** that may coordinate the exit of subdomains of multi-spanning membrane proteins into the bulk membrane[75,152] and **Ramp4** that is possibly involved in folding of membrane proteins[153]. In yeast, **Sec62** and **Sec63**, membrane proteins involved in post-translational protein translocation[154], as well as **ERj1** that contacts the chaperone **BiP** and the ribosome exit tunnel[155] are also associated with the translocon.

To avoid ion leakage from the ER lumen to the cytosol, and vice versa, during protein translocation through the open pore of Sec61α, **calmodulin** seals the channel from the cyotosolic side[156] and BiP from the lumenal side[157]. Because most TM helices probably contain Hsp70 (chaperone) binding sequences, BiP may interact with TM helices that missed membrane insertion and may funnel the proteins which contain such helices, and which are potentially misfolded, into the ER associated degradation pathway (ERAD)[107].

The bacterial translocon

It is estimated that about half of all SecYEG complexes in *E. coli* are associated with the chaperone and translocase **YidC**[47,158] and their cooperation seems to be essential for membrane protein insertion *in vivo*. YidC depletion causes inhibition of membrane protein insertion into the inner membrane[159]. For at least a subset of membrane proteins, YidC has been proposed to function as foldase or chaperone by receiving TM helices released from SecYEG

and allowing them to assemble into helix bundles[158,160]. YidC and the Sec translocase can act sequentially, which has been shown for CyoA, a subunit of the b_o oxidase. The N-terminal domain of CyoA gets inserted into the membrane with the help of YidC followed by SecYEG assisted membrane-insertion of the rest of the protein[161]. YidC has been shown to interact with all four helices of the lateral gate of SecY in absence of a RNC but not if SecY is engaged with a nascent chains[162].
SecYEG associates temporarily with the **SecDF-YajC** complex. This hetero-trimeric complex may be involved in stimulation of protein translocation using the proton motive force[163]. SecYEG, YidC and SecDF-YajC can form a stable complex, the holo-translocon, composed of one unit each. The holo-translocon is more efficient during co-translational insertion of membrane proteins and post-translational secretion[164].
Other SecYEG-interacting membrane anchored proteins are the **signal peptidase** (leader peptidase in *E. coli*) that cleaves off signal sequences from pre-proteins[165,166] and the protease **FtsH,** an AAA protease that degrades misfolded or damaged membrane proteins[167] as part of the membrane protein quality control system.
Finally, the membrane anchored chaperone **PpiD** that seems to be involved in post-translational translocation of outer membrane proteins via SecYEG[168] and the periplasmic proteins **Skp** and **YfgM**, both members of the periplasmic chaperone network, can associate with the translocon[169].
SecYEG may form homo-dimers via front-to-front[125] or back-to-back interactions[170]. The physiological role of a SecYEG dimer is not entirely clear. It has been suggested for a back-to-back dimer that one copy of SecYEG, the one not engaged in polypeptide translocation, provides a docking platform for SecA that might be essential for post-translational translocation[170].
During post-translational translocation of polypeptide chains and co-translational translocation of long periplasmic domains of membrane proteins, the ATPase **SecA** associates with SecYEG in a similar fashion as the ribosome[117] and energizes polypeptide translocation using ATP hydrolysis[171].

Molecular mechanisms of translocon opening, polypeptide translocation and membrane insertion

Translocon pore and lateral gate opening

The priming, high affinity interaction of the non-translating and translating ribosome with the cytosolic loops of SecY or Sec61α causes a slight destabi-

lization of the closed translocon and its lateral gate[128,172] while the plug domain seems not to be affected[121-123,126]. The insertion of a polypeptide chain into the translocon strengthens the ribosome-translocon interactions due to contacts between the freshly synthesized protein with the cytosolic loop between TM helices 6 and 7 of Sec61α [126] and probably likewise with the corresponding loop of SecY. These interactions may further destabilize the closed state of the translocon[128]. Signal sequences occupy a specific position within the cytosolic vestibule[60,173,174] and may open the lateral gate either partially or completely by inserting into a lateral cleft between TM helices 2b and 7[48,173]. The hydrophobic h-segment of signal sequences may so be able to form energy-lowering surface contacts with the apolar core of the lipid bilayer[175]. In general, the energy needed for the opening of the lateral gate seems to be small[124].
Signal sequence induced lateral gate opening may destabilize the plug domain. This, now more dynamic, plug domain may induce further destabilization of the translocon and especially of the lateral gate. It may become even more flexible[123]. Signal sequences may occupy the open lateral gate and become part of the channel wall, thereby increasing the size of the channel pore[121]. This is consistent with a crosslinking study proving the direct contact of a signal sequence, located most probably within the translocon, with lipids of the membrane[173].
During translocation, TM helix 10 of SecY, which is located at the back of the pore, has been found to shift by about 6 Å. This may allow the unhindered passage of any polypeptide chain through the translocon without the need of large-scale plug displacement or lateral gate opening[122]. In contrast to this model, it has been proposed that the lateral gate may continuously be open during polypeptide translocation and therefore all parts of membrane proteins may face temporary, at least partially, the lipid environment of the bilayer[124,176].

Signal sequence re-orientation

After the transfer of the RNC from SRP to the translocon, signal sequences can insert with their N-termini first and than reorient within the translocon[60,63] or may insert with their C-termini first to temporary form a hairpin-loop[177]. Even with the plug domain still in place and sealing the channel to the lumen or periplasm, the dimensions of the cytosolic vestibule of SecY/Sec61α is estimated to be large enough to allow the reorientation of signal sequences[123].
As discussed earlier, the main driving force for establishing the orientation of signaling sequences are positively charged residues enriched in their n-region[61,62,96]. Reasons for this may be i) specific negatively charged residues of Sec61α/SecY located within the cytosolic vestibule of the pore, in the

plug domain and at the C-terminal end of TM helix 8. They may act as electrostatic clamps by forming salt bridges to positively charged residues in signal sequences[63,136,175,178,179]. ii) TRAP has been shown to negatively affect the translocation of positively charged residues within a signal sequence through the translocon[149]. iii) The negative electrostatic potential created by the ribosomal RNA above the translocon may generally retain positively charged residues of polypeptide chains at the cytosolic side of the membrane. The ribosomal RNA may alternatively directly contact the positively charged residues at the ends of the targeting sequences. This has been shown for proteorhodopsin that interacts with its positively charged cytosolic loop between TM helix 1 and 2 with the rRNA loop 59[127]. iv) The lipid matrix may be involved at least in *E. coli*. The anionic cardiolipin has been found tightly bound to the translocon and may potentially be responsible for positive-charge retention on the cytosolic side[180]. PE has been shown to promote the effect of positively charged residues in polypeptide chains on membrane protein topogenesis, probably by neutralizing or diluting the topogenic effect of negatively charged residues within polypeptide chains[12].
The second main driving force for signal sequence topogenesis is the hydrophobicity of the signal sequence core. A computational study suggests that very hydrophobic targeting sequences may have already formed helices when entering the translocon. They can therefore rapidly undergo energetically favorable interactions with the membrane through the lateral gate. They may even open the lateral gate by promoting and stabilizing the open state. Less hydrophobic, potentially non-structured signal sequences may not be able to do that[178]. This might also be a reason for the observation that the more hydrophobic a signal anchor is, the higher the probability of its N_{out} orientation[63,64].

Molecular mechanisms of membrane protein topogenesis

The molecular mechanism underlying membrane proteins topogenesis via the Sec translocon is complex and probably modular because it must be applicable to thousands of kinds of membrane proteins with different amino acid compositions and structures. The following summary is not intended to be complete but is thought to be an introduction helping to understand the reasoning in the publications included in the thesis.

Release of membrane-excluded polypeptide sections

Cytosolic loops and domains may be released via a gap of about 10-20 Å on one side of the ribosome-translocon junction[119,121,181]. **Non-cytosolic loops and domains** are translocated through the pore of the Sec translocon and are released into the bacterial periplasm or the eukaryotic lumen of the ER. The ribosome-translocon interactions are weakened during translocation of long periplasmic domains (> 90 residues) in bacteria[127]. Further, SecA,

usually involved in post-translational translocation of proteins, is essential for translocation of periplasmic loops and domains longer than 60 residues[182]. Finally, the frequency of positively charged residues in long periplasmic loops is similar to that of soluble periplasmic proteins[183]. Integrating all these findings, it seems probable that the mode of translocation of long periplasmic domains is alike the post-translational translocation of soluble periplasmic proteins.

Release of membrane-spanning polypeptide parts

The lateral gate is opened to 12-14 Å when a TM helix has entered the channel[122,124]. This opening allows the helix to temporary equilibrate between the polar and aqueous interior of the translocon and the apolar environment of the membrane bilayer[33,82,184]. Driven by the hydrophobic effect, hydrophobic TM helices partition into the lipid bilayer while hydrophilic sequences are retained within the translocon. This thermodynamic partitioning model[33,82] is supported by the strong correlation of hydrophobicity scales derived from a large-scale study of Sec61-mediated membrane insertion of model TM helices and classical partitioning experiments solely based on the equilibration of small peptides between water and octanol, a membrane mimic[32,82]. The model is strengthened by the good correlation between the apolar surface area of TM helices and their membrane insertion efficiency (paper IV).

What sets the hydrophobicity threshold of whether a potential TM helix is laterally released or retained within the translocon? Substance equilibration between two phases, at given temperature and pressure, depends on the physico-chemical properties of the two phases and the solute[185]. In analogy to this, TM helix equilibration between the translocon pore and the proximal membrane environment should therefore depend on the intrinsic (and extrinsic) features of the TM helix, the electrostatics and the polarity of the translocon pore surface and interior and the composition of the membrane environment proximal to the lateral gate. Indeed, changing solely the electrostatics or polarity of the channel surface by replacing the six hydrophobic residues of the restriction ring by six lysines, asparagines, serines or glycines lowered the hydrophobicity threshold for membrane insertion of model TM helices of a membrane protein substantially[186].

It has been proposed that TM helix insertion is a kinetically controlled process. The kinetic model postulates that the opening of the lateral gate, being the slowest step during the membrane-insertion of a TM helix via the translocon, depends on the hydrophobicity of potential TM helices arriving in the translocon. According to this model, once the lateral gate is open, the helices are irreversibly released into the lipid bilayer[186,187]. The outcome of this alternative interpretation of the correlation between the hydrophobicity of a TM helix and its membrane insertion efficiency is currently experimentally indistinguishable from that of a thermodynamically driven process and a matter of debate. Somewhat supporting the kinetic model, a structure of the translocon reveals a lateral gate that adopted a nearly closed state after the

release of a signal sequence[126]. This may also be the default mechanism after membrane-insertion of a TM helix from a multi-spanning membrane protein. It is exemplified by the nearly closed lateral gate found in the structure of the ribosome-translocon complex harboring a proteorhodopsin intermediate[127]. Because the opening energetics of the lateral gate of the active translocon seems to be small[124], the lateral gate may fluctuate between several conformations including one that is closed and one that is open[72,176]. This adds an uncertainty to the interpretations of the state of the lateral gate after release of a membrane spanning polypeptide sequence solely based on structures of the translocon. It has been shown that the lateral gate opening is temperature dependent with low temperatures promoting the closed conformation[128]. Thus, the preparation of the sample subjected to structure determination of a specific ribosome-translocon complex might be critical for the conformation of the lateral gate.

While both the thermodynamically and the kinetically based mechanisms are sufficient to explain how hydrophobic TM helices enter the membrane, they do not explain how marginally hydrophobic TM helices are guided into the membrane. For such helices, the contribution of helix-extrinsic features has to be integrated into the model, which is still an ongoing process. Today we know of five principal TM helix-extrinsic features of the polypeptide chain that aid the membrane insertion of marginally hydrophobic helices. These are positively charged residues in the flanking regions of the polypeptide chain near the cytosolic end of the TM sequence[92] (paper I), helix-helix interactions that lower the energetic cost of membrane insertion[99], reduction of the hydrophobicity threshold for membrane insertion by a subsequent TM helix with a strong orientational preference (paper II), the turn propensity of short loops connecting two TM helices that form a helical hairpin structure as well as their electrostatic helix-helix interactions[110] and post-translational repositioning of TM helices relative to the membrane[114]. It can be speculated that the molecular mechanisms responsible for Sec-translocon mediated signal sequence and anchor recognition and reorientation also apply for marginally hydrophobic TM helices flanked by positively charged residues.

The passage of TM helices into the bulk membrane

The passage of membrane proteins from the lateral gate into the bulk membrane is not completely unraveled yet. It is probable that the progression into the bulk membrane is individual for every TM helix and may depend on its hydrophobicity, orientation and topogenic properties, as shown for aquaporin 4[70] and opsin[75,188] but may follow underlying general rules[189].

Direct release into the bulk membrane

Single-spanning membrane proteins, which usually have clearly hydrophobic TM helices[83], may spontaneously partitioning from the translocon to the lipid bilayer as soon as they have access to it and may freely diffuse away from the translocon[72].

Indirect release into the bulk membrane

In mammalian cells, besides experiencing Sec61α, TM helices of model and real multi-spanning membrane proteins have been cross-linked repetitively to TRAM during their membrane insertion progression[60,72,74,127,142,189] or to PAT10[75,188]. This suggests that at least a subset of TM helices of multi-spanning membrane proteins is not released directly into the bulk membrane but is temporarily deposited at one or more specific sites in the close vicinity of the translocon. These sites are termed "secondary binding site(s)". Such sites may be located at Sec61α, TRAM and PAT10. The retention of TM helices at such sites may be distinct for less hydrophobic helices[72,190]. Unfortunately, systematic large-scale studies that correlate the TM sequences intrinsic features and their local sequence context with occupation of such sites are missing.

In the inner membrane of *E. coli*, YidC interacts with the lateral gate of the resting translocon and may still be in its close vicinity during co-translational integration of membrane proteins into the inner membrane[162]. With its chaperoning and SecYEG-assisting functionalities for at least some membrane proteins[158] YidC may influence TM helix insertion directly[191]. Supporting this, the "hydrophobicity scale" measured for SecYEG-mediated TM helix insertion into the inner membrane of *E. coli* (paper III) correlates very well with the corresponding scale for solely YidC-mediated TM helix insertion into the inner membrane[192] despite their different insertion mechanisms.

The timing of TM helix release into the bulk membrane

The timing of the TM helix release into the bulk membrane may be dependent on the hydrophobicity of the helix for at least single-spanning membrane proteins. TM helices of low hydrophobicity have been shown to associate much longer with the translocon than clearly hydrophobic ones[72]. TM helices may generally stay for an extended time very close to the translocon. A model TM helix in a model protein separated from its subsequent, also engineered TM helix by more than 150 residues can still undergo helix-helix interactions to lower the energetic cost of inserting the second of the two helices into the membrane[99]. TM helix retention at the translocon has also been observed in a natural context, e.g. for $P2X_2$. Here, the N-terminally located signal anchor (SA-I) is retained at Sec61α at least until TM helix 2, which is separated from TM helix 1 by more 200 residues, has entered the translocon and gets inserted into the ER membrane[190].

Genetic code expansion and reprogramming using non-canonical amino acids

The idea to site-specifically incorporate amino acid derivatives with novel physico-chemical properties such as the ability to undergo light-inducible crosslinking reactions into proteins is not new. The first, semi-enzymatic synthesis of amino acid charged tRNA from pre-aminoacylated adenosinedi-phosphate (ppA-amino acid) and purified C-terminal truncated $tRNA^{Phe}$ (-pC-C_{OH})[193] was published already in 1978 but only ten years later coupling of an *E. coli* tRNA with a photo-reactive phenylalanine derivative, that could stimulate protein synthesis in an *in vitro* translation system, was accomplished[194]. This led to the development of the first genetic code expansion method aimed for site-specific incorporation of non-proteinogenic amino acids into proteins. This method is based on the assignment of a non-proteinogenic amino acid to the "blank" amber (UAG) stop codon. First, the anticodon of a particular tRNA is mutated to an amber stop anticodon ($tRNA^{amb}$)[195]. This $tRNA^{amb}$ is subsequently chemically acylated with the selected non-proteinogenic amino acid. This method is still in use but newer, more easy-to-use methods have emerged over the last decade.

One of them is based on the *in vivo* directed evolution of orthogonal (derived from another organism) amino-acid RNA synthetases (aaRS) to exclusively acylate their natural tRNA substrates with a particular non-natural amino acid[196]. The tRNA is engineered to harbor an amber stop anticodon, and thus, the non-proteinogenic amino acid is assigned to the amber stop codon. The subsequent *in vivo* co-expression of the aaRS/$tRNA^{amb}$ pair and the target gene with a pre-terminal amber stop codon, in presence of the "additional" amino acid, gives good yields of the expressed protein containing usually one or several residues of a particular non-canonical amino acid[197]. This intracellular reprogramming method, developed for *E. coli*, *Saccharomyces cerevisiae*, mammalian cell cultures and even *Caneorhabditis elegans*, has a vide application potential including mapping of protein-protein and protein-nucleotide interactions, visualization of conformational changes in proteins and protein dynamics, analysis of ligand binding sites and identification of receptors[198]. However, depending on the similarity of the non-natural amino acid to a canonical amino acid, the level of mischarged $tRNA^{amb}$ can be high[199].

Another, very potent, method for extending the genetic code *in vitro* is the "flexizyme system" named after an engineered and highly optimized ribozyme – the flexizyme[200]. It catalyzes tRNA acylation with a vide spectrum of non-natural amino acids, over 300 different so far, that includes also backbone-modified amino-acid analogues[201]. Flexizyme interacts with any tRNA only through base-pairing with the CCA terminus of the acceptor stem and with the different amino acid only by base stacking with an aromatic group of the amino acids; the latter being an absolute prerequisite[202]. For non-

aromatic amino acids, the good leaving group bound to the carboxy-group of the amino-acid backbone, which is the second prerequisite for flexizyme to work, can harbor the aromatic moiety[203].
The amino acylated $tRNA^{amb}$ can be added subsequently to an *in vitro* "amber suppression reaction" where the target gene harbors a pre-terminal amber stop codon. Using this method, proteins with site-specifically introduced non-canonical amino-acid residues can be produced efficiently.
The extreme promiscuity of the flexizyme allows genetic code reprogramming beyond single amino-acid code extensions. Namely, virtually all codons of the standard genetic code can be reassigned to natural or non-natural amino acids simply by selecting and acylating the appropriate tRNAs. This allowes the generation of exotic peptides, polyester and peptioids[201].
Mischarging of $tRNA^{amb}$ and amber stop codon read-through are the two main sources of problems when using non-natural amino acids, independent of the system used. Additionally, in multicellular organisms such as *C. elegans* further challenges arise as how to efficiently deliver the non-natural amino acid and how to express the orthogonal, (bacterial) tRNA stably in the target cells[204].

Introduction to the methodology

The model membrane protein "Lep"

Leader peptidase or "Lep" (encoded by *lepB*) is the only signal peptidase in *E. coli* but also a widely used model protein for studies on membrane protein topogenesis *in vitro* and *in vivo*. It has two N-terminal TM helices whereof the first one, a signal anchor (SA-II), is located at the very beginning of the polypeptide chain. TM helix 2 is separated from TM helix 1 by about 50 residues that form the only cytosolic loop[205]. The large periplasmic P2 domain has been crystalized several times and is mainly composed of β-sheet folds. It harbors the catalytic active site[166].
LepB can be expressed *in vitro* from a plasmid e.g. the pGEM1 vector (Invitrogen) in a cytosol mimic such as reticulocyte lysate (Promega) and in the presence of a membrane source that is in many studies provided by canine pancreatic microsomes. Microsomes are vesicle-like structures derived from the rough ER. Their membranes are identical to the membrane of the ER and the "microsomal lumen" is ER-lumen alike[206] (Figure 7).
Wild type Lep inserts into microsomes with an $N_{lumen}C_{lumen}$ topology (corresponds to $N_{out}C_{out}$)[207]. *LepB* used for *in vitro* studies is usually mutated multiple times. These mutations include the introduction of sequences encoding glycosylation acceptor sites anywhere within the protein. They allow the monitoring of co-translational translocation of Lep protein parts into the microsomal lumen via Sec61 because the OST complex adds very efficiently co-translationally oligosaccharides to glycosylation acceptor sites entering the microsomal lumen. The Lep topology during membrane insertion is derived from the glycosylation pattern of radiolabelled Lep and can easily be analyzed and quantified by SDS-PAGE and phosphoimager analysis because each oligosaccharide adds about 2.5 kDa to the molecular weight of the protein. A protease protection assay, that is the digestion of cytosolic protein extensions from the microsomal membranes, can further be used for verification of the final topology.

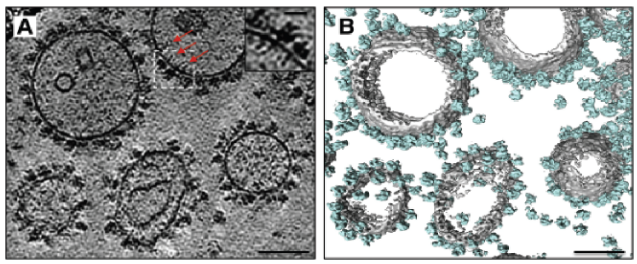

Figure 7: Cryo-EM tomogram of rough ER-derived microsomes. A) Slice through a representative tomogram showing, among others, membrane associated ribosomes (black dots on "circles"). B) Isosurface representation of the segmented volume; microsomal membranes are colored in gray, ribosomes in cyan (figure and text adapted from Pfeffer *et al*[206]).

The most common **Lep** variant (known simply as Lep or Lep-H3, Figure 8)[82] contains an additional, engineered model TM helix (H3, H-segment, test segment) in the P2 domain. Restriction sites for rapid sequence exchange, thus allowing large-scale studies, flank the H sequence in that *lepB* mutant. Glycosylation acceptor sites engineered into the *lepB* mutant up- and downstream of the H3 sequence allow the differentiation of radiolabelled Lep species with membrane spanning from those with membrane excluded H-segments. The extent of H3 integration into the membrane (1x glycosylated protein, f_{1x}) or translocation into the microsomal lumen (2x glycosylated protein, f_{2x}) can be analyzed and quantified by SDS-PAGE and phosphoimaging. The insertion efficiency of H3 into microsomal membranes, f_i, can subsequently be calculated as $f_i = f_{1x}/(f_{1x}+f_{2x})$. Assuming that the insertion of H3 segments into the ER membrane is an equilibrium process, the apparent change in free energy upon its membrane-insertion (ΔG_{app}) can be calculated using the relation $\Delta G_{app} = -RT\ln K$. R is the gas constant, T the absolute temperature in Kelvin and $K = f_{1x}/f_{2x}$.
Another *lepB* mutant used for large-scale studies is **Lep'** (Lep-H2, Figure 8)[208]. Restriction sites for rapid sequence exchange flank the DNA sequence of the natural TM helix 2 in Lep' and glycosylation acceptor sites are engineered into the *lepB* mutant gene such that the insertion efficiency of "H2" into the membrane can be monitored similarly to that of H3 in Lep.

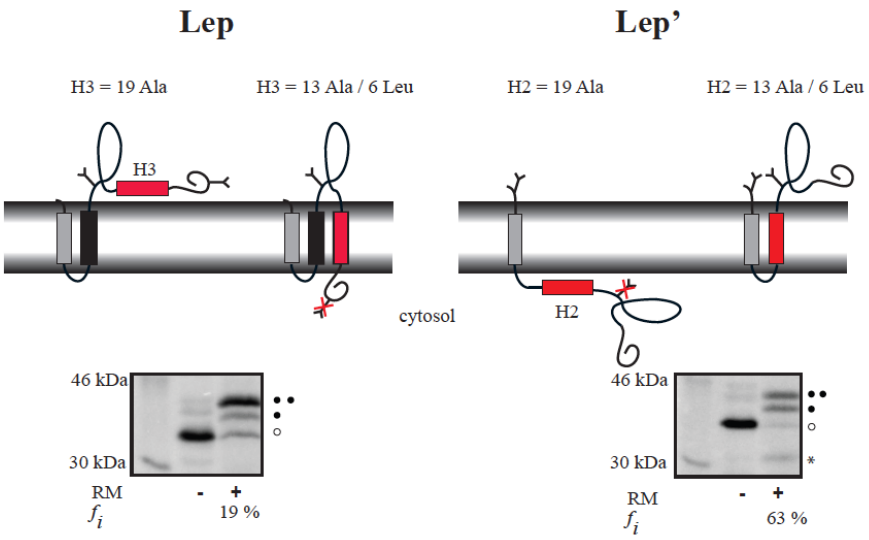

Figure 8: Lep and Lep'. Upper part: schematically drawn topology of the model proteins Lep and Lep' with either a membrane-excluded (H3 = 19Ala, red bar) or a hydrophobic, membrane-integrated H3-segment. Glycosylation acceptor sites within the protein are labeled with Y. Lower part: representative SDS-PAGE gel of a Lep and a Lep' construct expressed in the absence (RM -) or in the presence (RM +) of microsomal membranes. Non-glycosylated protein bands are labeled with open circles, singly glycosylated protein bands with one, and doubly glycosylated protein bands with two filled circles. Depending on the nature of H2 in Lep', it can be cleaved by the signal peptidase complex when spanning the membrane, yielding a truncated protein species (protein band marked with a star).

Lep^{IV} and **Lep^{V}** differ from each other by the presence of the original Lep TM helix 2 in Lep^{V} (Figure 9). These strongly modified *lepB* derived constructs contain sequences for two or three additional, exchangeable model TM helices within the P2 domain and one replacing TM helix 2 in Lep^{IV} (identically to Lep'). Expression of Lep^{IV} or Lep^{V} yields Lep species with model helix-triplets (H1, H2, H3) that are inversely oriented in microsomal membranes; given that all H-segments insert efficiently. The H-segments within both Lep variants are separated by loops of 48-50 residues and, besides from the N-terminus, all positively charged residues are replaced by alanines. Strategically placed glycosylation acceptor sites allow the identification of all possible topologies that radiolabeled Lep^{IV} and Lep^{V} can adopt within microsomal membranes and their subsequent quantification.

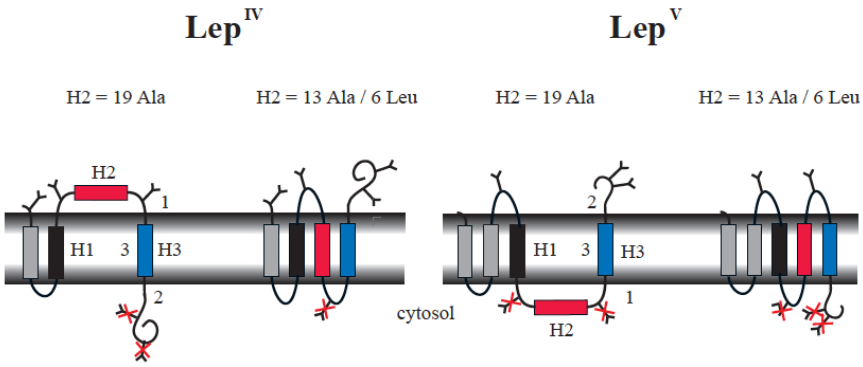

Figure 9: Lep^{IV} and Lep^{V}. Schematically drawn topology of Lep^{IV} and Lep^{V} with either a membrane-excluded or a membrane-inserted H2-segment (red bar). H1 (black bar) and H3 (blue bar) are engineered, hydrophobic helices that insert efficiently, co-translationally into membranes. An orientational preference of H3 can be "forced" either by introducing a cluster of positively charged residues into its N-terminal (1) or C-terminal (2) flanking region or by increasing its hydrophobicity (3).

Wild type or strongly mutated *lepB* can be expressed from a plasmid such as pING[151] *in vivo* in *E. coli* and these Lep species co-translationally insert into the inner membrane using the SRP-dependent pathway.

Lep^{LacY} the protein product of a *lepB* mutant in the pING vector is derived from Lep-H3 (Figure 10). Besides the exchangeable H3 sequence, a gene-fragment encoding the second TM helix of LacY and some of its flanking residues has been introduced C-terminally to the H3 sequence. LacY TM helix 2, named in Lep^{LacY} the "reporter segment" (R), is a good model substrate for the rhomboid protease GlpG of the *E. coli* inner membrane[209]. GlpG belongs to one of the most widespread families of membrane proteins present in most sequenced genomes[210], the intramembrane serine proteases. They cleave their TM peptide substrates within the plane of the membrane[211]. Analogously to the glycosylation assay *in vitro*, the reporter sequence can be used to distinguish Lep^{LacY} proteins with H3 segments inserted into the inner membrane (R is not cleaved) from those with H3 segments translocated into the periplasm (R is cleaved) and is an assay for assessing the final Lep topology. Full length and truncated Lep^{LacY} species can be separated by NuPAGE (Novex® NuPAGE® SDS-PAGE gel-system) and analyzed and quantified after Western blotting using Lep antisera.

Besides the applications described above, *lepB* has been used in a large number of studies not described in this thesis.

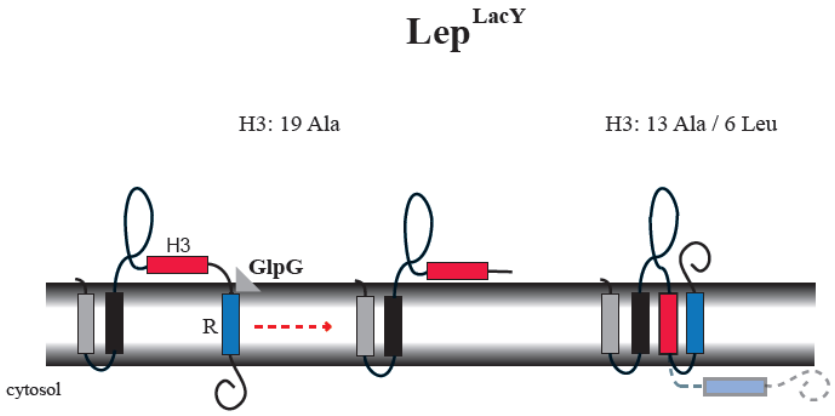

Figure 10: Lep^{LacY}. Schematically drawn topology of Lep^{LacY} with either a membrane-excluded or a membrane-inserted H3-segment (red bar). If H3 gets co-translationally integrated into the inner membrane of *E. coli*, the "reporter sequence" R (blue bar), being LacY TM helix 2, inserts with the wrong orientation or potentially does not insert into the membrane. It does not serve as a substrate for GlpG. If the H3 segment is co-translationally translocated into the periplasm, R inserts correctly oriented into the inner membrane and is cleaved by GlpG. This yields a truncated protein species.

In vitro expression of membrane proteins in presence of microsomal membranes

There are two very common methods for small-scale eukaryotic *in vitro* expression, radiolabelling and co-translational insertion of membrane proteins into canine pancreatic microsomes. **The single-tube assay**: plasmid DNA including the gene of the membrane protein, ^{35}S-methionine and microsomes are directly added to a salt-optimized reticulocyte lysate (TnT coupled transcription translation system, Promega) that is supplemented with the viral SP6 or T7 RNA polymerase. This allows mRNA synthesis and subsequent protein synthesis coupled to microsomal membrane insertion of the protein to take place in the same tube. **The two-step assay**: The mRNA of the membrane protein is first synthesized from the plasmid DNA using the SP6 or T7 RNA polymerase. The mRNA, ^{35}S-methionine and microsomes are directly added to reticulocyte lysate (rabbit reticulocyte lysate system, Promega) and the proteins are co-translationally inserted into the microsomal membranes. The separation of transcription and translation allows manual optimization of each synthesis step, which is not possible when using the

single-tube assay. In both cases, the samples are with minimal further manipulations subjected to SDS-PAGE.

Preparation of amber suppressor tRNA acylated with a non-natural amino acid using the flexizyme system

For usage in eukaryotic *in vitro* expression systems, the microbacteriophage L5-derived $tRNA^{Asn}$ ($ML\text{-}tRNA^{Asn}_{cta}$) with an engineered amber stop (UAG) anticodon has been found to be suitable. Mischarging of this $tRNA^{amb}$ by amino acid RNA synthetases in the reticulocyte lysate is minor, even negligible, compared to the amber suppression efficiency of added acylated $tRNA^{amb}$. To prepare $tRNA^{amb}$ that is acylated to a particular (non-natural) amino acid X, the flexizyme and the tRNA are separately transcribed repetitively from a DNA template and subsequently purified[212]. The ribozyme is then used to acylate the $tRNA^{amb}$ with the amino acid X and the $X\text{-}tRNA^{amb}$ then precipitated for storage. $X\text{-}tRNA^{amb}$ can be resuspended in a slightly acidic buffer and is added directly to the rabbit reticulocyte lysate of small-scaled *in vitro* translation reactions.

In vivo expression of the model membrane protein Lep^{LacY} in *E. coli*

The *E. coli* strain MC1061 is transformed with the expression vector pING, derived from pBR322, that carries the arabinose regulatory elements and the inducible *araB* promoter[151,213] followed by the mutant *lepB* gene encoding Lep^{LacY}. Cultures from single colonies of freshly transformed bacteria are grown over night at 37 °C, diluted and grown at 37 °C until they reach an optical density (OD) of about 0.35. L-arabinose is added to induce Lep^{LacY} expression and cells are harvested after 75 minutes. Cells are broken up, proteins are solubilized and denatured and subsequently separated using the NuPAGE system. After western blotting and selective protein band staining using a polyclonal Lep antibody the intensities of the different protein bands can be quantified.

Summary of publications

In 2007 it was predicted that a fraction as large as 30% of all TM helices of structure solved membrane proteins are marginally hydrophobic[83]. Our study on two members of the ABC transporter family (**A**denine **B**inding **C**assette transporter) the Cystic fibrosis TM conductance regulator (CFTR) and the drug efflux pump p-glycoprotein (p-gp) revealed that oddly only 3 TM helices of p-gp inserted well on their own into the ER membrane and all other TM helices needed support from other parts of the polypeptide chain[93]. The prediction and our study encouraged us to investigate how marginally hydrophobic helices are recognized and inserted into membranes.
We also started to systematically introduce non-natural amino acids into H3 of Lep-H3. Our goal was to extend the large-scale study that resulted in the biological hydrophobicity scales[82,83] to add to the understanding of the physical chemistry underlying the Sec61 mediated insertion of TM helix into the ER membrane.

Paper I

In this first study, we tried to identify features that support the membrane insertion of 16 marginally hydrophobic TM helices from membrane proteins of known structure. We chose only TM helices with a predicted insertion efficiency of ≤ 8 %, which corresponds to a predicted free energy of insertion (ΔG_{pred}) > 1.4 kcal/mol. Integrated into the H3 position of Lep-H3 we measured the insertion efficiency of those TM segments into microsomal membranes either as isolated TM helices, including their naturally flanking loop regions or including either the preceding, the subsequent or both neighboring TM helices.
We showed that 11 of the 16 marginally hydrophobic TM sequences identified by the ΔG predictor did not insert well into the ER membrane. We further showed that the local polypeptide sequence context provides features that help a marginally hydrophobic TM helix to insert properly into the ER membrane. These features included positively charged residues located in the cytosolic flanking loop regions of the marginally hydrophobic TM helix or resulted from a sequence-inherent property of adjacent TM helices.

Paper II

Based on the findings in paper I, we decided to systematically investigate the influence of the local sequence context on the insertion efficiency of marginally hydrophobic TM helices. Besides positively charged residues within flanking regions, a subsequent, clearly hydrophobic TM helix had the strongest influence on the membrane-insertion of a marginally hydrophobic helix and we therefore decided to focus on that. We created the two Lep variants Lep^{IV} and Lep^{V} with Lep-embedded model TM helix triplets and measured the hydrophobicity threshold (helix hydrophobicity needed for 50% membrane insertion) of the middle model helix into the microsomal membranes. We then introduced a strong orientational preference for membrane-insertion on the third model TM helix by adding a positively charged residue cluster to its N- or C-terminus and/or by changing its hydrophobicity.

We found that the orientational preference of a TM helix strongly influenced the membrane insertion efficiency of its preceding TM helix in a largely predictable way. Deducing from our results, we postulated that hydrophobic, orientationally constrained TM helices have the ability to post-translationally pull membrane-excluded marginally hydrophobic helices into the ER membrane or push already integrated, clearly hydrophobic TM helices out of the membrane, thereby possibly creating re-entrant loops.

We further verified our findings by a case study. TM helix 8 of the human apical sodium-dependent bile transporter (ASBT) adopts a broken helix structure within the membrane[214] and is further predicted not to be very hydrophobic. TM helix 9 has a strong orientational preference originating from a positive charge cluster at its C-terminal end and we could show that this cluster causes the post-translational integration of TM helix 8 into the ER membrane.

Paper III

The "ΔG predictor" algorithm is used to predict the free energy for membrane insertion of TM helices. It has also been used to estimate the fraction of marginally hydrophobic TM helices of structure solved membrane proteins[83]. It is based on experimental data derived from membrane-insertion studies of model TM helices into microsomal membranes, thus the ER membrane. Despite the strong conservation of the Sec-translocases, the question arises if the requirement of TM helices to partition into the inner membrane of *E. coli* is identical to the requirement for TM helix partitioning into the ER membrane. Therefore, do predicted marginally hydrophobic TM helices from *E. coli* inner membrane proteins really insert poorly into their

natural target membrane? No quantitative, systematic study for SecYEG mediated helix insertion into the inner membrane of *E. coli* existed at that time. We therefore carried out a large-scale study to identify the TM helix intrinsic sequence requirements that drive their SecYEG mediated, co-translational integration into the inner membrane. To do that, we developed and used the Lep^{LacY} based *in vivo* assay.
We found a somewhat lower hydrophobicity threshold for TM helix insertion into the inner membrane of *E. coli* compared to the corresponding threshold for insertion into the ER membrane. It was however similar to that of purely YidC-mediated TM helix insertion into the *E. coli* inner membrane. We further established a "biological hydrophobicity scale" for SecYEG mediated helix insertion into the inner membrane and measured the position-specific contribution of selected amino acids to the insertion efficiency of model TM helices into inner membrane. Both correlated very well with the corresponding scales for Sec61 mediated helix insertion into the ER membrane and confirm that the underlying physical chemistry is the same for both systems. Finally, we measured the insertion efficiency of 12 marginally hydrophobic TM helices from bacterial proteins (paper I) and compared their insertion efficiencies into the ER membrane and the inner membrane of *E. coli*.

Paper IV

The nature of the membrane-integration process of TM helices at the lateral gate of the Sec61 translocon is still under debate. In this study we tried to aid to the understanding of the physical chemistry underlying the TM helix membrane insertion process. We set out to answer the question of whether the non-polar surface area of TM helices correlates with their membrane insertion efficiencies or not. At the same time we wanted to measure how much the polar side-chain groups of tryptophan and tyrosine contribute to the membrane-insertion of helices. We site-specifically introduced a series of non-natural amino acids with purely aliphatic, non-branched or cyclic side chains, a series of amino acids with purely aromatic side chains as well as a series of tryptophan and tyrosine analogues into model helices in Lep-H3 using an amber suppression method.
We found that the apparent free energy of membrane insertion of such H3 segments varied in a regular fashion with the length of the side-chain of the non-natural, purely aliphatic residue. We noticed also a difference between the free energy for insertion of a helix-terminal, aliphatic non-branched residue to the same amino acid introduced at the central position. The difference increased with the side-chain length and was further asymmetric relative to the helix center. Amino acids with purely aliphatic but cyclic side chains and amino acids with purely aromatic side chains followed the same trends.

We further site-specifically introduced a single methylated tryptophan, methylated tyrosine, or benzothiophene-alanine into different helix-positions in H3 and measured their free energy of membrane insertion. Interestingly, they were very similar to the free energy of membrane insertion of H3 helices with a single phenylalanine at the same positions rather than a single tryptophan or tyrosine. Thus, the capability of tryptophan and tyrosine to form H-bonds seems to be the predominant cause of their positive influence for membrane insertion of TM helices.
Our results support the thermodynamic partitioning model of translocon-mediated helix insertion into the ER-membrane.

Conclusions and future perspective

At a first glance, "α-helical" membrane protein topogenesis seems not to be too complex. The relatively simple tertiary structure of the membrane spanning domains, generally hydrophobic TM helices with alternating orientation relative to the plane of the membrane, suggests topogenesis to be a straightforward process. In line with this, algorithms searching hydrophobic stretches in the sequence of multi-spanning membrane proteins are quite successful in predicting membrane protein topology and the number of membrane spanning TM helices[28].

The hydrophobicity of TM helices and positively charged residues located in the polypeptide sequences adjacent to TM helices play dominant roles in establishing the membrane protein topology and we largely understand how and why. But, we are still missing some details to completely understand the physical chemistry behind the process. Exemplified by using the previously discussed case of arginine residues within TM helices we still do not know e.g. how much the snorkeling of the side chain of arginine towards the interfaces contributes to the reduction of their free energy cost into membranes. In other words, how energetically "expensive" is a positively charged side-chain group within the center of the membrane? Questions like this may be answered by systematic studies using non-natural amino-acid residues within TM helices.

When it comes to marginally hydrophobic TM helices, we are still in the process of understanding the molecular mechanisms of membrane insertion. We know today of five basic principles that influence the membrane insertion of marginally hydrophobic helices. Are there more? How does the loop length connecting a marginally hydrophobic TM helix with its TM helix neighbors affect their membrane insertion efficiency? As has been proposed already over 10 years ago, TM helices accumulating at the translocon's lateral gate may create a local environment that is less hydrophobic than the bulk membrane[215]. If this is so, can marginally hydrophobic TM helices take advantage of this? Are marginally hydrophobic TM helices released directly into the membrane or do they regularly experience "secondary binding sites" in the vicinity of the large translocon machinery? These questions and many more are still in need of an answer makeing the study of membrane protein topogenesis an interesting topic to work on.

Populärvetenskaplig sammanfattning på svenska

Cellernas membraner är väldigt komplexa och heterogena strukturer som fyller många viktiga funktioner. De bildar tunna hinnor mot omvärlden och definierar på så sätt cellernas ytor. De tillåter utbyte av information och substanser mellan cellernas ut- och insida och möjliggör därför cellernas överlevnad och en snabb anpassning till växlande förhållanden. De innehåller nyckelkomponenter till många viktiga processer såsom fotosyntes och cellandning. Membraner består av en icke-polär lipidmatris fylld med en stor mängd olika membranproteiner. Dessa proteiner är i de flesta fall insatta tvärs igenom lipidmatrisen och sträcker sig in i utrymmena på båda sidor om membranet, d.v.s. de är trans-membrana.

Över miljontals år av evolution har celler utvecklat fantastiska maskiner som möjliggör integrering av membranproteiner i cellens membraner. Vår förståelse av dessa maskiner och integreringsprocesserna har kommit till en punkt där vi har en god överblick av membranproteinbiogenesen i olika typer av celler. Men när det kommer till detaljerad förståelse av de komplexa molekylära mekanismer och den fysikaliska kemin som ligger till grund för de olika integreringsprocesserna så samlar vi fortfarande den information som behövs för att förstå dem fullständigt.

Arbetet som presenteras i denna avhandling bidrar till förståelsen av den molekylära mekanism som medverkar till integrering av trans-membrana proteindelar i membranet. De proteindelar jag har studerat har den gemensamma egenskapen att de är ganska polära. Därför kan de inte spontant tränga in i den opolära membranmatrisen. Men trots detta bildar de ändå transmembrana element i många membranproteiner. Vårt arbete har lett till formuleringen av en ny mekanism för hur sådana relativt polära proteindelar integreras in i membranet.

Arbetet bidrar också till förståelsen av den fysikaliska kemin bakom den viktigaste integrationsprocessen för membranproteiner genom att systematiskt introducera noggrant utvalda onaturliga aminosyror in i trans-membrana delar av ett membranprotein och studera deras integrationsförmåga.

Acknowledgements

This is the final part of the thesis where I am allowed to thank all the people who, in any way, contributed to this thesis. Over the last seven years, I met many inspiring, motivating and nice people at DBB and was fortunate and happy to work with many of them on different projects. I was also lucky to be involved in several collaborations with people from around the world, leading to publications included or not included in this thesis. Thus, without you all, this thesis would not have become what it is or would not have been possible at all. Thank you!
There are some people I specially would like to thank:

My supervisor *Gunnar,* thank you for the great opportunity to work in your lab, for involving me in many interesting projects, for always having time to answer my questions and for your support. You are a very inspiring person and mentor and it has been an honor for me to be a part of your group. This PhD has been a great experience!
My second supervisor *IngMarie*, what would I have done without you! Thank you very much for all the discussions, the reminders of PhD program-related important time points and your help with the administrative and the experimental work.
Katrin, it was so much fun to work with you! Thank you for interesting conversations, for discussing results, work related topics and so much more. Thank you also for the great time in Japan!
Takashi, thank you for a great collaboration over several years, for teaching me how to make and use flexizyme and for showing me Tokyo. Thank you for all the scientific discussions; you are a fantastic chemist!
Patricia, you started as my student and I am happy that you stayed in IngMarie's lab as a PhD student. We worked together on smaller and bigger projects and I would like to thank you for all the time we spent together in the lab, for discussing experiments and results and for the great time in Japan.
Salome, thank you for continuing our project when I was at maternal leave and could not work. You put much effort in it and I appreciate it a lot.
Linnea and some other former and present members of the Elofsson group, thank you for a great collaboration leading to paper I of this thesis.

"My" students *Aron*, *Lisa*, *Nasim*, *Chiara, Hena* and *Aurora,* thank you for your work.
Stefan, thank you for your guidance during the seven years of my PhD.
Rob, thank you for many interesting discussions.

Thanks to all the present and former members of the *GvH group*, the *IMN group*, the *GvH community* and many thanks to *Joy's group* and to the *group of Prof. Hiroaki Suga*. It has been a great pleasure to meet you all and I wish you good luck for the future.

My friend and "neighbor" *Mirjam*, what a lucky coincidence that we were pregnant at the same time and worked both in the same corridor at DBB - because that's how we got to know each other. Thank you for the discussions about work, life and just everything, and many thanks for the time as neighbors in Bergshamra.

My parents *Theo and Barbara*, thank you for always believing in me even when I don't and for your support over all these years. I cannot express in words how much this means to me!
My sons *Erik* and *Birk*, you are the sunshine in my life. Thank you for you happiness, your love and making my life busy.
Johan, thank you for all the weekends we spent at DBB, for listening to my endless monologs about experiments, results and in general about biochemistry, your understanding when I was stressed due to upcoming presentations, exams or deadlines, your endless support and for just everything. You can always make me smile! I love you!

References

[1] Luckey, Mary, *Membrane structural biology*. (Cambridge university press, 2008).

[2] Wiener, M. C. and White, S. H., Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of x-ray and neutron diffraction data. III. Complete structure. *Biophys J* **61** (2), 434 (1992).

[3] White, S. H. and Wimley, W. C., Hydrophobic interactions of peptides with membrane interfaces. *Biochim Biophys Acta* **1376** (3), 339 (1998).

[4] Johansson, A. C. and Lindahl, E., Amino-acid solvation structure in transmembrane helices from molecular dynamics simulations. *Biophys J* **91** (12), 4450 (2006).

[5] Shibata, Y., Hu, J., Kozlov, M. M., and Rapoport, T. A., Mechanisms shaping the membranes of cellular organelles. *Annu Rev Cell Dev Biol* **25**, 329 (2009).

[6] van Meer, G., Voelker, D. R., and Feigenson, G. W., Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol* **9** (2), 112 (2008).

[7] Fagerberg, L. et al., Prediction of the human membrane proteome. *Proteomics* **10** (6), 1141 (2010).

[8] Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L., Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305** (3), 567 (2001).

[9] Bakheet, T. M. and Doig, A. J., Properties and identification of human protein drug targets. *Bioinformatics* **25** (4), 451 (2009).

[10] Bay, D. C. and Turner, R. J., Membrane composition influences the topology bias of bacterial integral membrane proteins. *Biochim Biophys Acta* **1828** (2), 260 (2013).

[11] Errington, J., L-form bacteria, cell walls and the origins of life. *Open Biol* **3** (1), 120143 (2013).

[12] Bogdanov, M., Dowhan, W., and Vitrac, H., Lipids and topological rules governing membrane protein assembly. *Biochim Biophys Acta* **1843** (8), 1475 (2014).

[13] Wikstrom, M. et al., Lipid-engineered Escherichia coli membranes reveal critical lipid headgroup size for protein function. *J Biol Chem* **284** (2), 954 (2009).

[14] Yan, N., Structural investigation of the proton-coupled secondary transporters. *Curr Opin Struct Biol* **23** (4), 483 (2013).

[15] Devaux, P. F. and Morris, R., Transmembrane asymmetry and lateral domains in biological membranes. *Traffic* **5** (4), 241 (2004).
[16] van Meer, G., Lipids of the Golgi membrane. *Trends Cell Biol* **8** (1), 29 (1998); van Meer, G., Lipid traffic in animal cells. *Annu Rev Cell Biol* **5**, 247 (1989).
[17] Bretscher, M. S. and Munro, S., Cholesterol and the Golgi apparatus. *Science* **261** (5126), 1280 (1993).
[18] Bretscher, M. S., Membrane structure: some general principles. *Science* **181** (4100), 622 (1973).
[19] Simons, K. and Ikonen, E., Functional rafts in cell membranes. *Nature* **387** (6633), 569 (1997).
[20] Allsopp, R. C., Lalo, U., and Evans, R. J., Lipid raft association and cholesterol sensitivity of P2X1-4 receptors for ATP: chimeras and point mutants identify intracellular amino-terminal residues involved in lipid regulation of P2X1 receptors. *J Biol Chem* **285** (43), 32770 (2010).
[21] Robinson, L. E., Shridar, M., Smith, P., and Murrell-Lagnado, R. D., Plasma membrane cholesterol as a regulator of human and rodent P2X7 receptor activation and sensitization. *J Biol Chem* **289** (46), 31983 (2014).
[22] Voeltz, G. K., Rolls, M. M., and Rapoport, T. A., Structural organization of the endoplasmic reticulum. *EMBO Rep* **3** (10), 944 (2002).
[23] Du, Y., Ferro-Novick, S., and Novick, P., Dynamics and inheritance of the endoplasmic reticulum. *J Cell Sci* **117** (Pt 14), 2871 (2004).
[24] Buton, X., Morrot, G., Fellmann, P., and Seigneuret, M., Ultrafast glycerophospholipid-selective transbilayer motion mediated by a protein in the endoplasmic reticulum membrane. *J Biol Chem* **271** (12), 6651 (1996).
[25] Sharpe, H. J., Stevens, T. J., and Munro, S., A comprehensive comparison of transmembrane domains reveals organelle-specific properties. *Cell* **142** (1), 158 (2010).
[26] Weiss, T. M. et al., Hydrophobic mismatch between helices and lipid bilayers. *Biophys J* **84** (1), 379 (2003).
[27] Shao, S. and Hegde, R. S., Membrane protein insertion at the endoplasmic reticulum. *Annu Rev Cell Dev Biol* **27**, 25 (2011).
[28] Elofsson, A. and von Heijne, G., Membrane protein structure: prediction versus reality. *Annu Rev Biochem* **76**, 125 (2007).
[29] Vinothkumar, K. R. and Henderson, R., Structures of membrane proteins. *Q Rev Biophys* **43** (1), 65 (2010).
[30] Wimley, W. C. et al., Folding of beta-sheet membrane proteins: a hydrophobic hexapeptide model. *J Mol Biol* **277** (5), 1091 (1998).
[31] White, S. H., Translocons, thermodynamics, and the folding of membrane proteins. *FEBS Lett* **555** (1), 116 (2003).
[32] Wimley, W. C., Creamer, T. P., and White, S. H., Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35** (16), 5109 (1996).

[33] White, S. H. and von Heijne, G., How translocons select transmembrane helices. *Annu Rev Biophys* **37**, 23 (2008).
[34] Higy, M., Junne, T., and Spiess, M., Topogenesis of membrane proteins at the endoplasmic reticulum. *Biochemistry* **43** (40), 12716 (2004).
[35] Wallin, E. and von Heijne, G., Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci* **7** (4), 1029 (1998).
[36] Fairman, J. W., Noinaj, N., and Buchanan, S. K., The structural biology of beta-barrel membrane proteins: a summary of recent reports. *Curr Opin Struct Biol* **21** (4), 523 (2011).
[37] Popot, J. L. and Engelman, D. M., Helical membrane protein folding, stability, and evolution. *Annu Rev Biochem* **69**, 881 (2000).
[38] Blobel, G., Intracellular protein topogenesis. *Proc Natl Acad Sci U S A* **77** (3), 1496 (1980).
[39] Nilsson, I., Whitley, P., and von Heijne, G., The COOH-terminal ends of internal signal and signal-anchor sequences are positioned differently in the ER translocase. *J Cell Biol* **126** (5), 1127 (1994).
[40] von Heijne, G., Transcending the impenetrable: how proteins come to terms with membranes. *Biochim Biophys Acta* **947** (2), 307 (1988).
[41] Blobel, G. and Dobberstein, B., Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J Cell Biol* **67** (3), 835 (1975).
[42] von Heijne, G., Signal sequences. The limits of variation. *J Mol Biol* **184** (1), 99 (1985).
[43] Nyathi, Y., Wilkinson, B. M., and Pool, M. R., Co-translational targeting and translocation of proteins to the endoplasmic reticulum. *Biochim Biophys Acta* **1833** (11), 2392 (2013).
[44] Calo, D. and Eichler, J., Crossing the membrane in Archaea, the third domain of life. *Biochim Biophys Acta* **1808** (3), 885 (2010).
[45] Gorlich, D. and Rapoport, T. A., Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. *Cell* **75** (4), 615 (1993).
[46] Halic, M. and Beckmann, R., The signal recognition particle and its interactions during protein targeting. *Curr Opin Struct Biol* **15** (1), 116 (2005).
[47] du Plessis, D. J., Nouwen, N., and Driessen, A. J., The Sec translocase. *Biochim Biophys Acta* **1808** (3), 851 (2010).
[48] Van den Berg, B. et al., X-ray structure of a protein-conducting channel. *Nature* **427** (6969), 36 (2004).
[49] Pohlschroder, M., Prinz, W. A., Hartmann, E., and Beckwith, J., Protein translocation in the three domains of life: variations on a theme. *Cell* **91** (5), 563 (1997).

[50] Ng, D. T., Brown, J. D., and Walter, P., Signal sequences specify the targeting route to the endoplasmic reticulum membrane. *J Cell Biol* **134** (2), 269 (1996).
[51] Gierasch, L. M., Signal sequences. *Biochemistry* **28** (3), 923 (1989).
[52] Peterson, J. H., Woolhead, C. A., and Bernstein, H. D., Basic amino acids in a distinct subset of signal peptides promote interaction with the signal recognition particle. *J Biol Chem* **278** (46), 46155 (2003).
[53] Lakkaraju, A. K. et al., Efficient secretion of small proteins in mammalian cells relies on Sec62-dependent posttranslational translocation. *Mol Biol Cell* **23** (14), 2712 (2012).
[54] Beckwith, J., The Sec-dependent pathway. *Res Microbiol* **164** (6), 497 (2013).
[55] von Heijne, G., Towards a comparative anatomy of N-terminal topogenic protein sequences. *J Mol Biol* **189** (1), 239 (1986).
[56] Nilsson, I. et al., The Code for Directing Proteins for Translocation across ER Membrane: SRP Cotranslationally Recognizes Specific Features of a Signal Sequence. *J Mol Biol*. (2014).
[57] Palmer, T. and Berks, B. C., The twin-arginine translocation (Tat) protein export pathway. *Nat Rev Microbiol* **10** (7), 483 (2012).
[58] Kumazaki, K. et al., Structural basis of Sec-independent membrane protein insertion by YidC. *Nature* **509** (7501), 516 (2014).
[59] von Heijne, G., Patterns of amino acids near signal-sequence cleavage sites. *Eur J Biochem* **133** (1), 17 (1983); von Heijne, G., How signal sequences maintain cleavage specificity. *J Mol Biol* **173** (2), 243 (1984).
[60] Devaraneni, P. K. et al., Stepwise insertion and inversion of a type II signal anchor sequence in the ribosome-Sec61 translocon complex. *Cell* **146** (1), 134 (2011).
[61] Heijne, G., The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J* **5** (11), 3021 (1986).
[62] von Heijne, G., Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* **341** (6241), 456 (1989).
[63] Goder, V. and Spiess, M., Molecular mechanism of signal sequence orientation in the endoplasmic reticulum. *EMBO J* **22** (14), 3645 (2003).
[64] Wahlberg, J. M. and Spiess, M., Multiple determinants direct the orientation of signal-anchor proteins: the topogenic role of the hydrophobic signal domain. *J Cell Biol* **137** (3), 555 (1997).
[65] Kocik, L., Junne, T., and Spiess, M., Orientation of internal signal-anchor sequences at the Sec61 translocon. *J Mol Biol* **424** (5), 368 (2012).
[66] Denzer, A. J., Nabholz, C. E., and Spiess, M., Transmembrane orientation of signal-anchor proteins is affected by the folding state but not the size of the N-terminal domain. *EMBO J* **14** (24), 6311 (1995).

[67] Goder, V., Bieri, C., and Spiess, M., Glycosylation can influence topogenesis of membrane proteins and reveals dynamic reorientation of nascent polypeptides within the translocon. *J Cell Biol* **147** (2), 257 (1999).
[68] Wickner, W. T. and Lodish, H. F., Multiple mechanisms of protein insertion into and across membranes. *Science* **230** (4724), 400 (1985).
[69] Ehrmann, M. and Beckwith, J., Proper insertion of a complex membrane protein in the absence of its amino-terminal export signal. *J Biol Chem* **266** (25), 16530 (1991).
[70] Sadlish, H., Pitonzo, D., Johnson, A. E., and Skach, W. R., Sequential triage of transmembrane segments by Sec61alpha during biogenesis of a native multispanning membrane protein. *Nat Struct Mol Biol* **12** (10), 870 (2005).
[71] Pitonzo, D. and Skach, W. R., Molecular mechanisms of aquaporin biogenesis by the endoplasmic reticulum Sec61 translocon. *Biochim Biophys Acta* **1758** (8), 976 (2006).
[72] Heinrich, S. U., Mothes, W., Brunner, J., and Rapoport, T. A., The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell* **102** (2), 233 (2000).
[73] Kanki, T. et al., The N-terminal region of the transmembrane domain of human erythrocyte band 3. Residues critical for membrane insertion and transport activity. *J Biol Chem* **278** (8), 5564 (2003).
[74] Do, H. et al., The cotranslational integration of membrane proteins into the phospholipid bilayer is a multistep process. *Cell* **85** (3), 369 (1996).
[75] Ismail, N. et al., Specific transmembrane segments are selectively delayed at the ER translocon during opsin biogenesis. *Biochem J* **411** (3), 495 (2008).
[76] Woolhead, C. A., McCormick, P. J., and Johnson, A. E., Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell* **116** (5), 725 (2004).
[77] Mingarro, I., Nilsson, I., Whitley, P., and von Heijne, G., Different conformations of nascent polypeptides during translocation across the ER membrane. *BMC Cell Biol* **1**, 3 (2000).
[78] Lu, J. and Deutsch, C., Folding zones inside the ribosomal exit tunnel. *Nat Struct Mol Biol* **12** (12), 1123 (2005).
[79] Tu, L., Khanna, P., and Deutsch, C., Transmembrane segments form tertiary hairpins in the folding vestibule of the ribosome. *J Mol Biol* **426** (1), 185 (2014).
[80] Pace, C. N. and Scholtz, J. M., A helix propensity scale based on experimental studies of peptides and proteins. *Biophys J* **75** (1), 422 (1998).

[81] Nilsson, I. et al., Proline-induced disruption of a transmembrane alpha-helix in its natural environment. *J Mol Biol* **284** (4), 1165 (1998).
[82] Hessa, T. et al., Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433** (7024), 377 (2005).
[83] Hessa, T. et al., Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450** (7172), 1026 (2007).
[84] Demirci, E. et al., Functional asymmetry within the Sec61p translocon. *Proc Natl Acad Sci U S A* **110** (47), 18856 (2013).
[85] Yau, W. M., Wimley, W. C., Gawrisch, K., and White, S. H., The preference of tryptophan for membrane interfaces. *Biochemistry* **37** (42), 14713 (1998).
[86] MacCallum, J. L. and Tieleman, D. P., Hydrophobicity scales: a thermodynamic looking glass into lipid-protein interactions. *Trends Biochem Sci* **36** (12), 653 (2008).
[87] MacCallum, J. L., Bennett, W. F., and Tieleman, D. P., Partitioning of amino acid side chains into lipid bilayers: results from computer simulations and comparison to experiment. *J Gen Physiol* **129** (5), 371 (2007); Dorairaj, S. and Allen, T. W., On the thermodynamic stability of a charged arginine side chain in a transmembrane helix. *Proc Natl Acad Sci U S A* **104** (12), 4943 (2007); Johansson, A. C. and Lindahl, E., Position-resolved free energy of solvation for amino acids in lipid membranes from molecular dynamics simulations. *Proteins* **70** (4), 1332 (2008).
[88] Johansson, A. C. and Lindahl, E., Protein contents in biological membranes can explain abnormal solvation of charged and polar residues. *Proc Natl Acad Sci U S A* **106** (37), 15684 (2009).
[89] Schow, E. V. et al., Arginine in membranes: the connection between molecular dynamics simulations and translocon-mediated insertion experiments. *J Membr Biol* **239** (1-2), 35 (2010).
[90] Ulmschneider, M. B., Ulmschneider, J. P., Sansom, M. S., and Di Nola, A., A generalized born implicit-membrane representation compared to experimental insertion free energies. *Biophys J* **92** (7), 2338 (2007).
[91] Nilsson, I. and von Heijne, G., Fine-tuning the topology of a polytopic membrane protein: role of positively and negatively charged amino acids. *Cell* **62** (6), 1135 (1990).
[92] Lerch-Bader, M. et al., Contribution of positively charged flanking residues to the insertion of transmembrane helices into the endoplasmic reticulum. *Proc Natl Acad Sci U S A* **105** (11), 4127 (2008).
[93] Enquist, K. et al., Membrane-integration characteristics of two ABC transporters, CFTR and P-glycoprotein. *J Mol Biol* **387** (5), 1153 (2009).
[94] Gafvelin, G. and von Heijne, G., Topological "frustration" in multispanning E. coli inner membrane proteins. *Cell* **77** (3), 401 (1994).

[95] Rapp, M., Granseth, E., Seppala, S., and von Heijne, G., Identification and evolution of dual-topology membrane proteins. *Nat Struct Mol Biol* **13** (2), 112 (2006).
[96] von Heijne, G., Membrane-protein topology. *Nat Rev Mol Cell Biol* **7** (12), 909 (2006).
[97] Seppala, S. et al., Control of membrane protein topology by a single C-terminal residue. *Science* **328** (5986), 1698 (2010).
[98] Yamagishi, M. et al., A few positively charged residues slow movement of a polypeptide chain across the endoplasmic reticulum membrane. *Biochemistry* **53** (33), 5375 (2014).
[99] Meindl-Beinker, N. M. et al., Asn- and Asp-mediated interactions between transmembrane helices during translocon-mediated membrane protein assembly. *EMBO Rep* **7** (11), 1111 (2006).
[100] Hermansson, M. and von Heijne, G., Inter-helical hydrogen bond formation during membrane protein integration into the ER membrane. *J Mol Biol* **334** (4), 803 (2003).
[101] Zhang, L. et al., Contribution of hydrophobic and electrostatic interactions to the membrane integration of the Shaker K+ channel voltage sensor domain. *Proc Natl Acad Sci U S A* **104** (20), 8263 (2007).
[102] Ota, K., Sakaguchi, M., Hamasaki, N., and Mihara, K., Membrane integration of the second transmembrane segment of band 3 requires a closely apposed preceding signal-anchor sequence. *J Biol Chem* **275** (38), 29743 (2000).
[103] Monne, M., Nilsson, I., Elofsson, A., and von Heijne, G., Turns in transmembrane helices: determination of the minimal length of a "helical hairpin" and derivation of a fine-grained turn propensity scale. *J Mol Biol* **293** (4), 807 (1999).
[104] Hermansson, M., Monne, M., and von Heijne, G., Formation of helical hairpins during membrane protein integration into the endoplasmic reticulum membrane. Role of the N and C-terminal flanking regions. *J Mol Biol* **313** (5), 1171 (2001).
[105] Yamagishi, M. et al., A sugar chain at a specific position in the nascent polypeptide chain induces forward movement during translocation through the translocon. *J Biochem* **149** (5), 591 (2011).
[106] Illergard, K., Kauko, A., and Elofsson, A., Why are polar residues within the membrane core evolutionary conserved? *Proteins* **79** (1), 79 (2011).
[107] Feige, M. J. and Hendershot, L. M., Quality control of integral membrane proteins by assembly-dependent membrane integration. *Mol Cell* **51** (3), 297 (2013).
[108] Fujita, H., Yamagishi, M., Kida, Y., and Sakaguchi, M., Positive charges on the translocating polypeptide chain arrest movement through the translocon. *J Cell Sci* **124** (Pt 24), 4184 (2011).
[109] Jusoh, S. A. and Helms, V., Helical integrity and microsolvation of transmembrane domains from Flaviviridae envelope glycoproteins. *Biochim Biophys Acta* **1808** (4), 1040 (2011).

[110] Bano-Polo, M. et al., Charge pair interactions in transmembrane helices and turn propensity of the connecting sequence promote helical hairpin insertion. *J Mol Biol* **425** (4), 830 (2012).
[111] Kida, Y., Morimoto, F., and Sakaguchi, M., Two translocating hydrophilic segments of a nascent chain span the ER membrane during multispanning protein topogenesis. *J Cell Biol* **179** (7), 1441 (2007).
[112] Ota, K. et al., Forced transmembrane orientation of hydrophilic polypeptide segments in multispanning membrane proteins. *Mol Cell* **2** (4), 495 (1998).
[113] Virkki, M. T. et al., Folding of Aquaporin 1: multiple evidence that helix 3 can shift out of the membrane core. *Protein Sci* **23** (7), 981 (2014).
[114] Kauko, A. et al., Repositioning of transmembrane alpha-helices during membrane protein folding. *J Mol Biol* **397** (1), 190 (2010).
[115] Ojemalm, K. et al., Positional editing of transmembrane domains during ion channel assembly. *J Cell Sci* **126** (Pt 2), 464 (2013).
[116] Virkki, M. et al., Large tilts in transmembrane helices can be induced during tertiary structure formation. *J Mol Biol* **426** (13), 2529 (2014).
[117] Zimmer, J., Nam, Y., and Rapoport, T. A., Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* **455** (7215), 936 (2008).
[118] Tsukazaki, T. et al., Conformational transition of Sec machinery inferred from bacterial SecYE structures. *Nature* **455** (7215), 988 (2008).
[119] Frauenfeld, J. et al., Cryo-EM structure of the ribosome-SecYE complex in the membrane environment. *Nat Struct Mol Biol* **18** (5), 614 (2011).
[120] Menetret, J. F. et al., Ribosome binding of a single copy of the SecY complex: implications for protein translocation. *Mol Cell* **28** (6), 1083 (2007).
[121] Park, E. et al., Structure of the SecY channel during initiation of protein translocation. *Nature* **506** (7486), 102 (2014).
[122] Gogala, M. et al., Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion. *Nature* **506** (7486), 107 (2014).
[123] Voorhees, R. M., Fernandez, I. S., Scheres, S. H., and Hegde, R. S., Structure of the mammalian ribosome-Sec61 complex to 3.4 A resolution. *Cell* **157** (7), 1632 (2014).
[124] Egea, P. F. and Stroud, R. M., Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. *Proc Natl Acad Sci U S A* **107** (40), 17182 (2010).
[125] Mitra, K. et al., Structure of the *E. coli* protein-conducting channel bound to a translating ribosome. *Nature* **438** (7066), 318 (2005).
[126] Becker, T. et al., Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science* **326** (5958), 1369 (2009).

[127] Bischoff, L. et al., Visualization of a polytopic membrane protein during SecY-mediated membrane insertion. *Nat Commun* **5**, 4103 (2014).
[128] Ge, Y. et al., Lateral opening of the bacterial translocon on ribosome binding and signal peptide insertion. *Nat Commun* **5**, 5263 (2014).
[129] Menetret, J. F. et al., Single copies of Sec61 and TRAP associate with a nontranslating mammalian ribosome. *Structure* **16** (7), 1126 (2008).
[130] du Plessis, D. J., Berrelkamp, G., Nouwen, N., and Driessen, A. J., The lateral gate of SecYEG opens during protein translocation. *J Biol Chem* **284** (23), 15805 (2009).
[131] Lycklama, A. Nijeholt J. A., Bulacu, M., Marrink, S. J., and Driessen, A. J., Immobilization of the plug domain inside the SecY channel allows unrestricted protein translocation. *J Biol Chem* **285** (31), 23747 (2010).
[132] Harris, C. R. and Silhavy, T. J., Mapping an interface of SecY (PrlA) and SecE (PrlG) by using synthetic phenotypes and in vivo cross-linking. *J Bacteriol* **181** (11), 3438 (1999).
[133] Saparov, S. M. et al., Determining the conductance of the SecY protein translocation channel for small molecules. *Mol Cell* **26** (4), 501 (2007).
[134] Junne, T., Schwede, T., Goder, V., and Spiess, M., The plug domain of yeast Sec61p is important for efficient protein translocation, but is not essential for cell viability. *Mol Biol Cell* **17** (9), 4063 (2006).
[135] Li, W. et al., The plug domain of the SecY protein stabilizes the closed state of the translocation channel and maintains a membrane seal. *Mol Cell* **26** (4), 511 (2007).
[136] Junne, T., Schwede, T., Goder, V., and Spiess, M., Mutations in the Sec61p channel affecting signal sequence recognition and membrane protein topology. *J Biol Chem* **282** (45), 33201 (2007).
[137] Trueman, S. F., Mandon, E. C., and Gilmore, R., A gating motif in the translocation channel sets the hydrophobicity threshold for signal sequence function. *J Cell Biol* **199** (6), 907 (2012).
[138] Duong, F. and Wickner, W., Sec-dependent membrane protein biogenesis: SecYEG, preprotein hydrophobicity and translocation kinetics control the stop-transfer function. *EMBO J* **17** (3), 696 (1998).
[139] Hanada, M., Nishiyama, K. I., Mizushima, S., and Tokuda, H., Reconstitution of an efficient protein translocation machinery comprising SecA and the three membrane proteins, SecY, SecE, and SecG (p12). *J Biol Chem* **269** (38), 23625 (1994).
[140] Oliver, J. et al., The Sec61 complex is essential for the insertion of proteins into the membrane of the endoplasmic reticulum. *FEBS Lett* **362** (2), 126 (1995).
[141] Gorlich, D., Hartmann, E., Prehn, S., and Rapoport, T. A., A protein of the endoplasmic reticulum involved early in polypeptide translocation. *Nature* **357** (6373), 47 (1992).

142 Sauri, A., McCormick, P. J., Johnson, A. E., and Mingarro, I., Sec61alpha and TRAM are sequentially adjacent to a nascent viral membrane protein during its ER integration. *J Mol Biol* **366** (2), 366 (2007).
143 Tamborero, S. et al., Membrane insertion and topology of the translocating chain-associating membrane protein (TRAM). *J Mol Biol* **406** (4), 571 (2011).
144 Pfeffer, S. et al., Structure of the mammalian oligosaccharyltransferase complex in the native ER protein translocon. *Nat Commun* **5**, 3072 (2014).
145 Kelleher, D. J. and Gilmore, R., An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **16** (4), 47R (2006).
146 Aebi, M., N-linked protein glycosylation in the ER. *Biochim Biophys Acta* **1833** (11), 2430 (2013).
147 Gavel, Y. and von Heijne, G., Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng* **3** (5), 433 (1990).
148 Hartmann, E. et al., A tetrameric complex of membrane proteins in the endoplasmic reticulum. *Eur J Biochem* **214** (2), 375 (1993).
149 Sommer, N. et al., TRAP assists membrane protein topogenesis at the mammalian ER membrane. *Biochim Biophys Acta* **1833** (12), 3104 (2013).
150 Dalbey, R. E. and Von Heijne, G., Signal peptidases in prokaryotes and eukaryotes--a new protease family. *Trends Biochem Sci* **17** (11), 474 (1992).
151 Dalbey, R. E. and Wickner, W., Leader peptidase catalyzes the release of exported proteins from the outer surface of the Escherichia coli plasma membrane. *J Biol Chem* **260** (29), 15925 (1985).
152 Meacock, S. L., Lecomte, F. J., Crawshaw, S. G., and High, S., Different transmembrane domains associate with distinct endoplasmic reticulum components during membrane integration of a polytopic protein. *Mol Biol Cell* **13** (12), 4114 (2002).
153 Hori, O. et al., Deletion of SERP1/RAMP4, a component of the endoplasmic reticulum (ER) translocation sites, leads to ER stress. *Mol Cell Biol* **26** (11), 4257 (2006).
154 Lang, S. et al., Different effects of Sec61alpha, Sec62 and Sec63 depletion on transport of polypeptides into the endoplasmic reticulum of mammalian cells. *J Cell Sci* **125** (Pt 8), 1958 (2012).
155 Dudek, J. et al., ERj1p has a basic role in protein biogenesis at the endoplasmic reticulum. *Nat Struct Mol Biol* **12** (11), 1008 (2005).
156 Erdmann, F. et al., Interaction of calmodulin with Sec61alpha limits Ca2+ leakage from the endoplasmic reticulum. *EMBO J* **30** (1), 17 (2011).
157 Schauble, N. et al., BiP-mediated closing of the Sec61 channel limits Ca2+ leakage from the ER. *EMBO J* **31** (15), 3282 (2012).
158 Dalbey, R. E., Wang, P., and Kuhn, A., Assembly of bacterial inner membrane proteins. *Annu Rev Biochem* **80**, 161 (2011).

159 Samuelson, J. C. et al., YidC mediates membrane protein insertion in bacteria. *Nature* **406** (6796), 637 (2000).
160 Beck, K. et al., YidC, an assembly site for polytopic Escherichia coli membrane proteins located in immediate proximity to the SecYE translocon and lipids. *EMBO Rep* **2** (8), 709 (2001).
161 Celebi, N. et al., Membrane biogenesis of subunit II of cytochrome bo oxidase: contrasting requirements for insertion of N-terminal and C-terminal domains. *J Mol Biol* **357** (5), 1428 (2006).
162 Sachelaru, I. et al., YidC occupies the lateral gate of the SecYEG translocon and is sequentially displaced by a nascent membrane protein. *J Biol Chem* **288** (23), 16295 (2013).
163 Tsukazaki, T. et al., Structure and function of a membrane component SecDF that enhances protein export. *Nature* **474** (7350), 235 (2011).
164 Schulze, R. J. et al., Membrane protein insertion and proton-motive-force-dependent secretion through the bacterial holo-translocon SecYEG-SecDF-YajC-YidC. *Proc Natl Acad Sci U S A* **111** (13), 4844 (2014).
165 Chang, C. N., Blobel, G., and Model, P., Detection of prokaryotic signal peptidase in an Escherichia coli membrane fraction: endoproteolytic cleavage of nascent f1 pre-coat protein. *Proc Natl Acad Sci U S A* **75** (1), 361 (1978).
166 Paetzel, M., Structure and mechanism of Escherichia coli type I signal peptidase. *Biochim Biophys Acta* **1843** (8), 1497 (2014).
167 Dalbey, R. E., Wang, P., and van Dijl, J. M., Membrane proteases in the bacterial protein secretion and quality control pathway. *Microbiol Mol Biol Rev* **76** (2), 311 (2012).
168 Antonoaea, R., Furst, M., Nishiyama, K., and Muller, M., The periplasmic chaperone PpiD interacts with secretory proteins exiting from the SecYEG translocon. *Biochemistry* **47** (20), 5649 (2008).
169 Gotzke, H. et al., YfgM is an ancillary subunit of the SecYEG translocon in Escherichia coli. *J Biol Chem* **289** (27), 19089 (2014).
170 Deville, K. et al., The oligomeric state and arrangement of the active bacterial translocon. *J Biol Chem* **286** (6), 4659 (2011).
171 Bauer, B. W., Shemesh, T., Chen, Y., and Rapoport, T. A., A "push and slide" mechanism allows sequence-insensitive translocation of secretory proteins by the SecA ATPase. *Cell* **157** (6), 1416 (2014).
172 Wu, Z. C., de Keyzer, J., Kedrov, A., and Driessen, A. J., Competitive binding of the SecA ATPase and ribosomes to the SecYEG translocon. *J Biol Chem* **287** (11), 7885 (2012).
173 Plath, K. et al., Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. *Cell* **94** (6), 795 (1998).
174 Mackinnon, A. L. et al., An allosteric Sec61 inhibitor traps nascent transmembrane helices at the lateral gate. *Elife* **3**, e01483 (2014).

[175] Zhang, B. and Miller, T. F., 3rd, Direct simulation of early-stage Sec-facilitated protein translocation. *J Am Chem Soc* **134** (33), 13700 (2012).
[176] Osborne, A. R., Rapoport, T. A., and van den Berg, B., Protein translocation by the Sec61/SecY channel. *Annu Rev Cell Dev Biol* **21**, 529 (2005).
[177] Rapoport, T. A., Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450** (7170), 663 (2007).
[178] Zhang, B. and Miller, T. F., 3rd, Long-timescale dynamics and regulation of Sec-facilitated protein translocation. *Cell Rep* **2** (4), 927 (2012).
[179] Goder, V., Junne, T., and Spiess, M., Sec61p contributes to signal sequence orientation according to the positive-inside rule. *Mol Biol Cell* **15** (3), 1470 (2004).
[180] Gold, V. A. et al., The action of cardiolipin on the bacterial translocon. *Proc Natl Acad Sci U S A* **107** (22), 10044 (2010).
[181] Beckmann, R. et al., Architecture of the protein-conducting channel associated with the translating 80S ribosome. *Cell* **107** (3), 361 (2001).
[182] Andersson, H. and von Heijne, G., Sec dependent and sec independent assembly of E. coli inner membrane proteins: the topological rules depend on chain length. *EMBO J* **12** (2), 683 (1993).
[183] von Heijne, G., Membrane proteins: from sequence to structure. *Annu Rev Biophys Biomol Struct* **23**, 167 (1994).
[184] White, S. H. and von Heijne, G., Transmembrane helices before, during, and after insertion. *Curr Opin Struct Biol* **15** (4), 378 (2005).
[185] Leo, Albert, Hansch, Corwin, and Elkins, David, Partition coefficients and their uses. *Chemical Reviews* **71** (6), 525 (1971).
[186] Junne, T., Kocik, L., and Spiess, M., The hydrophobic core of the Sec61 translocon defines the hydrophobicity threshold for membrane integration. *Mol Biol Cell* **21** (10), 1662 (2010).
[187] Zhang, B. and Miller, T. F., 3rd, Hydrophobically stabilized open state for the lateral gate of the Sec translocon. *Proc Natl Acad Sci U S A* **107** (12), 5399 (2010).
[188] Ismail, N., Crawshaw, S. G., and High, S., Active and passive displacement of transmembrane domains both occur during opsin biogenesis at the Sec61 translocon. *J Cell Sci* **119** (Pt 13), 2826 (2006).
[189] Hou, B., Lin, P. J., and Johnson, A. E., Membrane protein TM segments are retained at the translocon during integration until the nascent chain cues FRET-detected release into bulk lipid. *Mol Cell* **48** (3), 398.
[190] Cross, B. C. and High, S., Dissecting the physiological role of selective transmembrane-segment retention at the ER translocon. *J Cell Sci* **122** (Pt 11), 1768 (2009).
[191] Cymer, F., von Heijne, G., and White, S. H., Mechanisms of Integral Membrane Protein Insertion and Folding. *J Mol Biol* (2014).

192 Xie, K. et al., Features of transmembrane segments that promote the lateral release from the translocase into the lipid phase. *Biochemistry* **46** (51), 15153 (2007).
193 Hecht, S. M., Alford, B. L., Kuroda, Y., and Kitano, S., "Chemical aminoacylation" of tRNA's. *J Biol Chem* **253** (13), 4517 (1978).
194 Baldini, G. et al., Mischarging Escherichia coli tRNAPhe with L-4'-[3-(trifluoromethyl)-3H-diazirin-3-yl]phenylalanine, a photoactivatable analogue of phenylalanine. *Biochemistry* **27** (20), 7951 (1988).
195 Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., and Schultz, P. G., A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244** (4901), 182 (1989).
196 Xie, J. and Schultz, P. G., An expanding genetic code. *Methods* **36** (3), 227 (2005).
197 Wang, L., Xie, J., and Schultz, P. G., Expanding the genetic code. *Annu Rev Biophys Biomol Struct* **35**, 225 (2006).
198 Xie, J. and Schultz, P. G., A chemical toolkit for proteins--an expanded genetic code. *Nat Rev Mol Cell Biol* **7** (10), 775 (2006).
199 Antonczak, A. K. et al., Importance of single molecular determinants in the fidelity of expanded genetic codes. *Proc Natl Acad Sci U S A* **108** (4), 1320 (2011).
200 Murakami, H., Saito, H., and Suga, H., A versatile tRNA aminoacylation catalyst based on RNA. *Chem Biol* **10** (7), 655 (2003).
201 Passioura, T. and Suga, H., Reprogramming the genetic code in vitro. *Trends Biochem Sci* **39** (9), 400 (2014).
202 Xiao, H., Murakami, H., Suga, H., and Ferre-D'Amare, A. R., Structural basis of specific tRNA aminoacylation by a small in vitro selected ribozyme. *Nature* **454** (7202), 358 (2008).
203 Ohuchi, M., Murakami, H., and Suga, H., The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus. *Curr Opin Chem Biol* **11** (5), 537 (2007).
204 Parrish, A. R. et al., Expanding the genetic code of Caenorhabditis elegans using bacterial aminoacyl-tRNA synthetase/tRNA pairs. *ACS Chem Biol* **7** (7), 1292 (2012).
205 Whitley, P., Nilsson, L., and von Heijne, G., Three-dimensional model for the membrane domain of Escherichia coli leader peptidase based on disulfide mapping. *Biochemistry* **32** (33), 8534 (1993).
206 Pfeffer, S. et al., Structure and 3D arrangement of endoplasmic reticulum membrane-associated ribosomes. *Structure* **20** (9), 1508 (2012).
207 Johansson, M., Nilsson, I., and von Heijne, G., Positively charged amino acids placed next to a signal sequence block protein translocation more efficiently in Escherichia coli than in mammalian microsomes. *Mol Gen Genet* **239** (1-2), 251 (1993).
208 Lundin, C. et al., 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** (41), 15702 (2008).

[209] Akiyama, Y. and Maegawa, S., Sequence features of substrates required for cleavage by GlpG, an Escherichia coli rhomboid protease. *Mol Microbiol* **64** (4), 1028 (2007).
[210] Koonin, E. V. et al., The rhomboids: a nearly ubiquitous family of intramembrane serine proteases that probably evolved by multiple ancient horizontal gene transfers. *Genome Biol* **4** (3), R19 (2003).
[211] Ha, Y., Structural principles of intramembrane proteases. *Curr Opin Struct Biol* **17** (4), 405 (2007).
[212] Murakami, H., Ohta, A., Ashigai, H., and Suga, H., A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nat Methods* **3** (5), 357 (2006).
[213] Johnston, S., Lee, J. H., and Ray, D. S., High-level expression of M13 gene II protein from an inducible polycistronic messenger RNA. *Gene* **34** (2-3), 137 (1985).
[214] Hu, N. J., Iwata, S., Cameron, A. D., and Drew, D., Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT. *Nature* **478** (7369), 408 (2011).
[215] Engelman, D. M. et al., Membrane protein folding: beyond the two stage model. *FEBS Lett* **555** (1), 122 (2003).