Integration and topology of membrane proteins

Carolina Boekel
Cover illustration: "Cell membrane" inspired by Mondrian

© Carolina Boekel, Stockholm 2009, pages i-49


Printed in Sweden by Universitetsservice AB, Stockholm 2009
Distributor: Department of Biochemistry and Biophysics,
"If you learn a lot of little things, one day you may end up knowing a big thing"

- B.K.S Iyengar
Contents

Abstract i
Publications included in this thesis ii
Introduction 4
Background 6
  Biological membranes 6
  Lipids 7
  Structural features of membrane proteins 7
    α-helical membrane proteins 8
    β-barrel membrane proteins 9
  Hydrophobic amino acids 9
  Aromatic amino acids 9
  Proline-induced turns 9
  Snorkelling 10
  Hydrophobicity scales 10
  Biogenesis of membrane proteins in the endoplasmic reticulum 12
    Targeting to the endoplasmic reticular membrane 12
    The signal recognition particle 12
    Post –translational translocation 13
    The translocon 14
    Glycosylation - Oligosaccharyl transferase 18
    Signal peptidase 19
  Topology - a structural model of membrane proteins 21
    Membrane protein orientation 21
    Single-spanning membrane proteins 22
    Bioinformatics 23
    Topology mapping 24
Methodology 27
  Model system - Leader peptidase 27
  Expression in vivo 28
  Expression in S. cerevisiae 29
Results and discussion 30
  Summary of papers 30
Paper I: Helix-helix interactions 30
Paper II: Opposite orientation: N_{in}-C_{out} vs. N_{out}-C_{in} transmembrane helices 32
Paper III: Insertion efficiency of the amyloid β-peptide into the ER membrane 34
Paper IV and V: Topology mapping of two eukaryotic membrane proteins 35

Acknowledgements 37

References 39
Abbreviations

aa        amino acids
Aβ        amyloid β-peptide
APP       amyloid-β precursor protein
C-terminal carboxy-terminal
DNA       deoxyribonucleic acid
ER        endoplasmic reticulum
Lep       leader peptidase
N-terminal amino-terminal
OST       oligosaccharyl transferase
RNC       ribosome nascent chain complex
SP        signal peptidase
SR        SRP receptor
SRP       signal recognition particle
TM        transmembrane
TMH       transmembrane helix
TMHMM     transmembrane hidden Markov model
TRAM      translocating chain-associated membrane protein
TRAP      translocon –associated protein complex
Å         Ångström
### Amino acids

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>Alanine</td>
<td>Leu</td>
<td>Leucine</td>
</tr>
<tr>
<td>Arg</td>
<td>Arginine</td>
<td>Lys</td>
<td>Lysine</td>
</tr>
<tr>
<td>Asn</td>
<td>Asparagine</td>
<td>Met</td>
<td>Methionine</td>
</tr>
<tr>
<td>Asp</td>
<td>Aspartic acid</td>
<td>Phe</td>
<td>Phenylalanine</td>
</tr>
<tr>
<td>Cys</td>
<td>Cysteine</td>
<td>Pro</td>
<td>Proline</td>
</tr>
<tr>
<td>Gly</td>
<td>Glycine</td>
<td>Ser</td>
<td>Serine</td>
</tr>
<tr>
<td>Gln</td>
<td>Glutamine</td>
<td>Thr</td>
<td>Threonine</td>
</tr>
<tr>
<td>Glu</td>
<td>Glutamic acid</td>
<td>Trp</td>
<td>Tryptophan</td>
</tr>
<tr>
<td>His</td>
<td>Histidine</td>
<td>Tyr</td>
<td>Tyrosine</td>
</tr>
<tr>
<td>Ile</td>
<td>Isoleucine</td>
<td>Val</td>
<td>Valine</td>
</tr>
</tbody>
</table>

**Prediction server**

- **ΔG-pred**
  Prediction of ΔG for TM-helix insertion
  

- **Sfinx metaserver**
  Prediction of transmembrane topology
  
  [http://sfinx.cgb.ki.se](http://sfinx.cgb.ki.se)
Abstract

Membrane proteins comprise around 20-30% of most proteomes. They play important roles in most biochemical pathways. All receptors and ion channels are membrane proteins, which make them attractive targets for drug design. Membrane proteins insert and fold co-translationally into the endoplasmic reticular membrane of eukaryotic cells. The protein-conducting channel that inserts the protein into the membrane is called Sec61 translocon, which is a hetero-oligomeric channel that allows transmembrane segments to insert laterally into the lipid bilayer. The focus of this thesis is how the translocon recognizes the transmembrane helices and integrates them into the membrane.

We have investigated the sequence requirements for the translocon-mediated integration of a transmembrane α-helix into the ER by challenging the Sec61 translocon with designed polypeptide segments in an in vitro expression system that allows a quantitative assessment of membrane insertion efficiency. Our studies suggest that helices might interact with each other already during the membrane-insertion step, possibly forming helical hairpins that partition into the membrane as a single unit. Further, the insertion efficiency for N_in-C_out vs. N_out-C_in transmembrane helices and the integration efficiency of Alzheimer’s Aβ-peptide fragments has been investigated. Finally, detailed topology mapping was performed on two biologically interesting proteins with unknown topology, the human seipin protein and *Drosophila melanogaster* odorant receptor OR83b.
Publications included in this thesis

This thesis is based on the following publications, which will be referred to by their Roman numerals.


Reprints were made with permission from the publishers.
Additional publications:


Introduction

All living cells contain proteins that carry out specialized functions within the membrane or aqueous spaces. Half of all proteins in a typical cell are transported across or into a membrane. How are proteins synthesized in the cytoplasm of the cell and inserted across or into the membranes?

The eukaryotic cell contains both a plasma membrane and internal membranes. These internal membranes create vesicles and organelles such as the nucleus, endoplasmic reticulum (ER), mitochondrion, chloroplasts, peroxisomes, Golgi and lysosome/vacuoles, each with a specialized function.

George Palade defined the basic structure of the secretory pathway in eukaryotic cells and in 1970, Günter Blobel performed experiments on the translocation of proteins across membranes. He discovered that many proteins have a short amino acid sequence at one end that functions like a postal code for the target organelle, a signal sequence, and the “signal hypothesis” was born (Blobel and Sabatini, 1971).

In the ER, the N-terminal postal code (signal sequence) of the protein is recognized by the signal recognition particle (SRP) while the protein is still being synthesized by the ribosome. The synthesis pauses while the ribosome-protein complex is transferred to an SRP receptor at the ER. There, the nascent protein is inserted into a protein channel (the translocon) that spans the ER membrane. The translocon is also responsible for retro-translocation of misfolded polypeptides destined for degradation in the cytosol by the ER-associated degradation (ERAD) mechanism (Nakatsukasa and Brodsky, 2008; Wiertz et al., 1996).

Proteins that are localized to the Golgi, lysosome/vacuole and plasma membrane are first inserted in the ER. Within the ER, a 14-residue oligosaccharyl core is attached to glycoproteins. Transmembrane proteins are translocated through the membrane by the translocon, until the process is interrupted by a stop-transfer sequence. The amino acid chain of a transmembrane protein can pass back and forth across the membrane one or several times.
This thesis is about how membrane proteins insert into the ER membrane. In papers I-III, we have investigated the sequence requirements for translocation and insertion of artificial and natural transmembrane segments. Very few structures of membrane proteins are known and to broaden our understanding of structure, we have performed topology mapping on two biologically interesting membrane proteins of unknown structure, paper IV-V: the human seipin protein and the *Drosophila melanogaster* odorant receptor OR83b.
Background

Biological membranes

Membranes define the external boundaries of cells and regulate the molecular traffic in and out of the cell. The membrane is a dynamic lipid bilayer that is selectively permeable; it lets some compounds through, while excluding others. In order to do this, the cell membrane contains proteins embedded in the lipid bilayer that act as mediators. These proteins are referred to as membrane proteins. All membrane proteins contain both a hydrophobic part and a hydrophilic part that enable them to be anchored in the membrane and at the same time be in contact with the surrounding environment.

Singer and Nicholson defined the biological membrane (Figure 1A) as a bilayer of amphipathic lipids with proteins floating around freely as icebergs in a sea (Singer and Nicolson, 1972). In addition, some membranes also have carbohydrates (mono- and oligosaccharides) attached to their surface.

Figure 1. A. Schematic representation of a biological membrane of lipids, cholesterol and membrane proteins. B. Chemical structure of a membrane lipid, phosphatidylethanolamin.
Lipids

Membrane lipids are amphipathic molecules with a hydrophilic head and a hydrophobic tail (Figure 1B). For most lipids, the tail region is made up of two acyl carbon chains, typically 16-18 carbons long and differing in the degree of saturation (double bonds).

Due to thermodynamic constrains in an aqueous environment, the lipids self-assemble into a bilayer where the lipid tails point towards the interior, while the headgroups face the hydrophilic exterior on each side of the membrane. Cylindrical lipids promote the formation of a flat lipid bilayer. A lipid is cylindrical when the head group has approximately the same size as the hydrocarbon tail. If the headgroup is larger or smaller than the hydrocarbon tails, the conical shape induces curvature in the lipid monolayer (Gruner, 1985).

The head groups can also vary in polarity; they can be neutral, zwitterionic or negatively charged. Most biological membranes are made of phospholipids, glycolipids and cholesterol, a neutral lipid containing a ring structure that breaks the tight packing of fatty acid chains, creating a more fluid membrane. Biological membranes contain many different kinds of lipids and the lipid composition gives the different membranes their specific characteristics such as curvature, fluidity and rigidity.

Structural features of membrane proteins

Membrane proteins have amphipathic structures that reflect the membrane in which they reside. They have polar surfaces that interact with the aqueous solution and with the lipid head groups, and nonpolar surfaces that interact with the nonpolar interior of the lipid bilayer. As a result of this, they are neither soluble in aqueous solution nor in non-polar solvents.

Membrane proteins can be classified into two categories: integral (intrinsic) and peripheral (extrinsic), based on the type of lipid-protein interactions. An integral membrane protein contains residues with hydrophobic side chains that interact with the fatty acyl chains of the membrane phospholipids and anchors the protein to the membrane in a way that it cannot be extracted by high pH. Most integral proteins span the entire membrane layer and are therefore called transmembrane proteins.

Transmembrane proteins can assume the structure of helix bundles or β-barrels (von Heijne, 1996) due to the stabilization by the internal hydrogen bonds (Figure 2).
Peripheral membrane proteins are associated with the membrane through interactions with the hydrophilic domain of integral proteins and/or with the polar head groups of membrane lipids. Carbonate extraction at high pH is often used to extract peripheral membrane proteins from the membrane.

α-helical membrane proteins

When there are no water molecules available to form hydrogen bonds with the carbonyl oxygen and amide nitrogen of the peptide bond, internally hydrogen bonded α-helices give the most stable conformation (Hermansson and von Heijne, 2003). Genomic analyses of several organisms indicate that 20-30% of all open reading frames code for helix bundle membrane proteins (Krogh et al., 2001).

The most obvious feature of the helix-bundle class proteins is a stretch of around 26 hydrophobic amino acids that will form a transmembrane a-helix, which is roughly perpendicular to the membrane plane (Ulmschneider et al., 2005). Between 15 and 43 amino acids are generally considered the boundaries, in order to successfully form an α-helix that spans the membrane (Granseth et al., 2005). When the polypeptide chain is translocated by the translocon, the hydrophobic segments will partition into the membrane. The favorable partitioning of the hydrophobic side chains into the membrane and the formation of backbone hydrogen bonds overcome the unfavorable entropy of folding the backbone (Faham et al., 2004; Popot and Engelman, 1990). Both single-spanning and polytopic membrane proteins exist.
In prokaryotes, most α-helical membrane proteins are located in the inner membrane, but recently a polysaccharide transporter (Wza) has been shown to form a transmembrane helix bundle in the outer membrane of *Escherichia coli* (Dong et al., 2006).

**β-barrel membrane proteins**

Another possibility for hydrophobic amino acid sequences to span the membrane is the β-barrel, stabilized by hydrogen bonds between antiparallel β-strands. Such proteins are found in the outer membrane of Gram-negative bacteria (Kleinschmidt and Tamm, 1996) and are predicted to be present in the outer membrane of mitochondria and chloroplasts (Schleiff et al., 2003).

**Hydrophobic amino acids**

The hydrophobic domain of a TM α-helix is mainly composed of hydrophobic amino acid residues (Ile, Leu, Val, Ala, Phe and Gly). These residues account for 63% of the residues in TM helices (Ulmschneider and Sansom, 2001).

Alanine is a hydrophobic amino acid on the threshold between those amino acids that promote membrane integration of membrane proteins and those that do not. Previous work (Nilsson et al., 2003a) has shown that Ala has a slight tendency to prefer the lipid-water interface region over the central part of the membrane.

**Aromatic amino acids**

Aromatic amino acids are often located at the lipid-water interface, ±15 Å from the center of the membrane. Trp and Tyr are enriched near the ends of the helices, suggesting that they interact favorably with the lipid headgroups. Phe, on the other hand is more abundant in the central core region of the TM helices. Trp, Tyr and Phe tend to push the transmembrane helix into the membrane when inserted in positions flanking the TM, and Trp pulls the transmembrane helix toward the lipid-water interface when inserted inside the TM segment (Braun and von Heijne, 1999).

**Proline-induced turns**

Previous work has shown that Pro breaks a 40 residue long poly–Leu TM helix when placed in the central ten positions (Nilsson and von Heijne, 1998). The loss of hydrogen bonds and the steric problems caused by the Pro residue makes it energetically more favorable to place Pro in a tight turn near the membrane/water interface than to force it into the center of a long, mem-
brane-embedded TM helix (Nilsson et al., 1998). To be effective as a turn-promoter, the Pro needs to be placed in the middle of a 30 residues long hydrophobic stretch (Monne et al., 1999). Pro has also been shown to be a topology determinant of the integral membrane protein stomatin. Some proteins such as stomatin have a unique hairpin-loop topology, possibly created by the helix-breaking properties of Pro (Kadurin et al., 2008).

Snorkelling

The so-called “snorkel effect” occurs when charged residues with long aliphatic side chains can stretch along the surface of the helix (Chamberlain et al., 2004). This allows the charged terminal group to reach into the lipid headgroup region, even when the Cα is five to six residues below the membrane/water interface.

A previous study (Monne et al., 1998) showed that the negatively charged residues, Asp and Glu, can have a strong effect on the position of a poly-Leu TMH in the ER membrane when they are located in the first turn of the helix. They had almost no effect when placed further into the hydrophobic stretch.

The side chain of Asp is short and cannot “snorkel” in the same way as positively charged Arg and Lys. There may also be some repulsion between the negatively charged residues and negatively charged phospholipid head groups (van Klompenburg et al., 1997).

Hydrophobicity scales

The hydrophobicities of the individual amino acid side chains have been measured experimentally in octanol, which serves as a model for the hydrocarbon core of a lipid bilayer (White and Wimley, 1999; Wimley et al., 1996). The hydrophobicity of the amino acids is defined as the free energy of transfer from water to a nonpolar liquid.

The apparent hydrophobicities of amino acid side chains vary a lot; depending on whether or not polar groups are present. Ionized and polar side chains interact strongly with water and have lower solubilities in nonpolar solvents because of the unfavorable energetics of placing polar groups in a nonpolar environment. According to the octanol scale, transferring the backbone (a Gly) is unfavorable by 1.25 kcal mol⁻¹ per residue. This suggests that side-chain hydrophobicity drives the equilibrium in favor of insertion and that there is a threshold hydrophobicity that would favor insertion into the lipid bilayer (Liu et al., 1996).
A 20-residue polyalanine sequence would not be hydrophobic enough to partition into the hydrocarbon core ($\Delta G=10 \text{ kcal mol}^{-1}$), but replacing five of the Ala with five Leu residues would make the insertion favorable ($\Delta G=-1.25 \text{ kcal mol}^{-1}$). Experiments on translocon-mediated TM insertion indicate that this prediction is remarkably accurate (Hessa et al., 2005).

**Figure 3.** $\Delta G$ scale derived from H-segments with the individual aa placed in the middle of a 19-residue hydrophobic stretch (Hessa et al., 2005).

Based on a large set of data from insertion of designed model segments, the contribution to the free energy of membrane integration ($\Delta G_{\text{app}}^{aa}$) could be estimated for each of the 20 naturally occurring amino acids when placed in the middle of a segment. As seen in Figure 3, Ile, Leu, Phe and Val promote membrane insertion ($\Delta G_{\text{app}}^{aa} < 0$), Cys, Met, Ala have ($\Delta G_{\text{app}}^{aa} \approx 0$) while the polar and charged residues have ($\Delta G_{\text{app}}^{aa} > 0$).
Biogenesis of membrane proteins in the endoplasmic reticulum

Targeting to the endoplasmic reticular membrane

Proteins that are going to be exported (translocated) across or inserted into the ER membrane need both the action of the translating ribosome in the cytoplasm and translocons located in the ER of eukaryotes or the plasma membrane of bacteria. The translocon allows soluble polypeptides to cross the membrane and hydrophobic transmembrane segments of membrane proteins to exit laterally into the lipid phase. Insertion and translocation of membrane proteins can occur both co- and post-translationally and converge at the Sec complex.

Proteins destined for translocation carry an amino-terminal signal sequence that marks them for translocation, which is recognized by targeting factors (Walter and Blobel, 1981a). The signal sequence can be divided into three different regions: a positively charged N-terminal (n-region), a central hydrophobic core (h-region) and a slightly polar C-terminal (c-region), which often contains amino acids with short side chains at the positions closest to the cleavage site (von Heijne, 1983; von Heijne, 1985).

The signal recognition particle

In the event of co-translational insertion, the N-terminal part of the polypeptide emerges from the ribosome and is recognized by the signal recognition particle (SRP). SRP is a ribonucleotide particle, existing in all three kingdoms of life (Keenan et al., 2001). The binding of SRP to the hydrophobic signal sequences arrests protein synthesis (Walter and Blobel, 1981b), this prevents the protein from premature folding or aggregation in the cytosol.

SRP docks onto the ribosomal component L23 located near the ribosomal exit tunnel and binds to the hydrophobic signal sequence as it emerges from the tunnel (Schaffitzel et al., 2006; Valent et al., 1997), forming a ribosome – nascent chain complex (RNC). In eukaryotes SRP then binds GTP and arrests elongation. The RNC-complex binds to the membrane bound SRP receptor (SR), another GTPase, which is associated with the translocon. The binding of SRP to SR activates the GTPase activities of both SRP and SR, whereupon the SRP disengages from the SR and the ribosome, the protein is transferred to the translocon, and elongation of the protein continues (Figure 4) (White and von Heijne, 2004).
The mammalian SRP consists of a 7S RNA and six proteins (SRP9, 14, 19, 54, 68, 72) and it has two domains, the Alu domain (SRP14-SRP9 and 7SL RNA) and the S domain (forked region of 7SL RNA and the remaining four proteins) (Nagai et al., 2003). The Alu domain confers peptide elongation arrest activity. The S domain mediates signal sequence binding and SR docking (Halic and Beckmann, 2005).

A cryo-EM structure of a mammalian SRP bound to an active 80S ribosome (Halic et al., 2004; Wild et al., 2004) shows how the S domain of SRP interacts with the large ribosomal subunit at the nascent chain exit site and binds the signal sequence, while the Alu domain reaches into the elongation-factor-binding site of the ribosome, thereby causing the elongation arrest activity.

**Post–translational translocation**

Post-translational translocation operates by different mechanisms in eukaryotes and bacteria. In yeast ER, the Sec61p complex together with the Sec62p, Sec63p, Sec71p and Sec72p subcomplexes are responsible for post-translational translocation. Mammalian homologues to Sec61p (Sec62) and Sec63p (Sec63) have been identified and proposed to function in post-translational translocation. However, no efficient post-translational translocation has been observed in mammalian ER (Meyer et al., 2000).
Eukaryotes need BiP (Kar2p in yeast), an Hsp70 homologue in the ER lumen, which utilises ATP hydrolysis to pull the polypeptide chain in a ratchet-like manner (Matlack et al., 1999; Misselwitz et al., 1999). Bacteria push proteins through the translocon channel by using a motor protein, the ATPase SecA (Brundage et al., 1990; Hartl et al., 1990). In addition, the reaction requires a cytosolic component called SecB to aid targeting to the membrane and to prevent premature protein folding and aggregation (de Keyzer et al., 2002; Hartl et al., 1990; Kusters et al., 1989; Weiss et al., 1988).

The translocon

In 1975, the hypothesis that secretory proteins are translocated through the ER membrane via an aqueous channel was first presented (Blobel and Dobberstein, 1975). Another popular theory at the time was that the signal sequence directs the spontaneous insertion of the polypeptide into the membrane, while the rest of the polypeptide is translocated through the hydrophobic core of the membrane without any assistance of a channel or proteins (Engelman and Steitz, 1981). The mechanism of protein insertion remained unresolved and was debated for many years due to the lack of experimental data to support either theory. It was not until the beginning of the 1990s that data were presented which supported the involvement of a proteinaceous channel in protein insertion (Simon and Blobel, 1991).

The translocon components that form the protein-conducting channel in the ER membrane were first identified by photocrosslinking (Krieg et al., 1989). Today we know from X-ray crystallography that Sec61 is a heterotrimeric complex (Van den Berg et al., 2004). It consists of $\alpha$, $\beta$, and $\gamma$ subunits (Sec61p, Sbh1p and Sss1p in yeast (Table 1) (Finke et al., 1996)/SecY, G and E in bacteria). All genomes sequenced to date encode at least one homolog of each Sec61$\alpha$$\beta$$\gamma$/SecYEG subunits, suggesting that these proteins play a universal role (Cao and Saier, 2003).

The high resolution X-ray structure of an archeal translocon (Van den Berg et al., 2004) suggests that one copy of the Sec61 heterotrimer serves as a functional translocation channel (Figure 5). The $\alpha$-subunit is the largest with 10 TM helices, forms the central channel, and acts as the main ribosome receptor (Kalies et al., 1994). The $\alpha$-subunit can be divided into two halves (TM1-5 and 6-10), joined at the back of the molecule by an external loop and the single Sec61$\gamma$-subunit. A cavity with the diameter of 20-25 Å is formed at the cytoplasmic side of the $\alpha$-subunit. The tunnel through the channel has a diameter of 5-8 Å in the middle and contains a ring of hydrophobic residues. This ring may form a seal around the translocating polypeptide, hindering the leakage of other molecules.
Figure 5. The structure of the Sec-translocon from *Methanococcus jannaschii* (Van den Berg et al., 2004). Top view of the translocon (left) showing the α-subunit consisting of two halves, TM1-5 and TM6-10, which can open laterally between TM7-8 and 2b-3. The purple cylinder shows the possible position of a nascent chain TM helix that is about to move into the membrane. The channel is closed from the periplasmic side by the green “plug helix”; this plug is thought to move out of the way when the ribosome binds to the translocon.

The Sec61β- and γ-subunits span the membrane once (Esnault et al., 1994; Van den Berg et al., 2004). The function of the Sec61β-subunit, a 10 kD C-tail-anchored transmembrane protein, is still under investigation. An interesting finding is that Sec61β (SecG in prokaryotes) is the only translocon subunit that is not essential for cell viability in yeast and bacteria (Boisrame et al., 1996; Finke et al., 1996; Nishiyama et al., 1994; Toikkanen et al., 1996). Mammalian Sec61β is not required for protein translocation *in vitro*, but has been shown to have an kinetic effect on co-translational translocation (Kalies et al., 1998). In *Saccharomyces cerevisiae*, Sbh1p (Sec61β) is believed to act as the guanine nucleotide exchange factor for the signal recognition particle receptor (SR) (Helmers et al., 2003).

The translocon structure also shows a plug that blocks the pore and presumably closes the channel (Figure 5). Plug displacement will then open the channel and make translocation possible. The structure also suggests that a TM segment in nascent polypeptides can exit the translocon channel laterally between TM2b and TM7 of Sec61. How the opening and closing of this gate is regulated is not known. It seems that when a hydrophobic sequence appears inside the ribosome, it induces a conformational change that would be transmitted to the luminal side of the channel, causing the channel to close by the binding of the protein Bip (Hamman et al., 1998). Another conforma-
tional change would then reopen ribosome-channel junction if the polypeptide chain is further elongated (Rapoport et al., 2004).

There are several lower-resolution electron-microscopy structures of ribosome-bound bacterial and mammalian translocons (Mitra et al., 2005; Morgan et al., 2002) proposing different models for the structure. In one model, the active translocon is proposed to be a single Sec61 complex with the nascent polypeptide chain translocating through a narrow channel. In another model, Sec61 is suggested to be a front-to-front Sec61 dimer with two lateral gates partly open and facing each other, creating a large central channel. A more recent structure shows that a nontranslating ribosome binds a single copy of the SecY complex with the pore of SecY located under the ribosomal tunnel exit (Menetret et al., 2007). Consequently, this copy of SecY could form the channel. Non-translating 80S ribosomes have been shown to associate with three to four copies of trimeric Sec61-complexes, whereas a translating ribosome binds only a single Sec61-complex in the ER (Kalies et al., 2008).

In mammals, additional membrane proteins are present in the vicinity of the translocon complex. These include the signal peptidase (SP), the translocating chain associated membrane protein (TRAM), the translocon-associated protein complex (TRAP) (Table 1) and the oligosaccharyl transferase (OST) (Johnson and van Waes, 1999; Osborne et al., 2005).
Table 1. Targeting and translocon components in the ER membranes of mammalian, *S. cerevisiae* and *E. coli*

<table>
<thead>
<tr>
<th>Mammals</th>
<th><em>S. cerevisiae</em></th>
<th><em>E. coli</em></th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary components</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRP</td>
<td>SRP</td>
<td>SRP</td>
<td>signal recognition particle</td>
</tr>
<tr>
<td>Sec61α</td>
<td>Sec61p</td>
<td>Ssh1p</td>
<td>SecY</td>
</tr>
<tr>
<td>Sec61β</td>
<td>Sbh1p</td>
<td>Sbh2p</td>
<td>SecG</td>
</tr>
<tr>
<td>Sec61γ</td>
<td>Sss1p</td>
<td>Sss1p</td>
<td>SecE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SecA</td>
</tr>
<tr>
<td>Sec62</td>
<td>Sec62p</td>
<td></td>
<td>component for post-translational translocation</td>
</tr>
<tr>
<td>Sec63</td>
<td>Sec63p</td>
<td></td>
<td>component for post-translational translocation</td>
</tr>
<tr>
<td>Sec71p</td>
<td></td>
<td></td>
<td>component for post-translational translocation</td>
</tr>
<tr>
<td>Sec72p</td>
<td></td>
<td></td>
<td>component for post-translational translocation</td>
</tr>
<tr>
<td><strong>Accessory components</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRAM</td>
<td></td>
<td></td>
<td>YidC</td>
</tr>
<tr>
<td>TRAP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oligosaccharyltransferase (OST)</td>
<td>OST</td>
<td></td>
<td>glycosylates proteins in the ER lumen, close to translocon pore</td>
</tr>
<tr>
<td>Bip</td>
<td>Kar2p</td>
<td></td>
<td>lumenal Hsp70 family member</td>
</tr>
<tr>
<td>Signal peptidase (SP)</td>
<td>SP</td>
<td>Lep</td>
<td>signal peptidase</td>
</tr>
<tr>
<td>Calnexin</td>
<td>Calnexin</td>
<td></td>
<td>chaperone activity: binds to glycoproteins</td>
</tr>
<tr>
<td>Calreticulin</td>
<td></td>
<td></td>
<td>chaperone activity: binds to glycoproteins</td>
</tr>
</tbody>
</table>
Glycosylation - Oligosaccharyl transferase

N-linked glycosylation occurs in all eukaryotes and in a few prokaryotes (Kowarik et al., 2006). In eukaryotes, the process is catalyzed by a large (~270 kDa) (Li et al., 2008) transmembrane protein complex called oligosaccharyl transferase (OST).

Many secretory and membrane proteins are N-glycosylated by the OST complex during their translocation across the ER membrane (Figure 6) (Chavan et al., 2005). During co-translational glycosylation, OST is in contact with both the translocon and the ribosome (Menetret et al., 2005; Wang and Dobberstein, 1999). When a polypeptide emerges from the translocon the oligosaccharide is added at the luminal side, while the rest of the polypeptide is still being synthesized by the ribosome in the cytosol (Menetret et al., 2005; Mothes et al., 1994; Whitley et al., 1996).

Figure 6. OST complex and ribosome-translocon complex with translocating nascent chain.
For N-linked glycosylation to occur, OST must cleave the large oligosaccharide group from dolicholpyrophosphate, transfer it to the substrate polypeptide, and catalyze the formation of a covalent bond between the oligosaccharide and the Asn in the N-X-S/T acceptor site.

The composition of the OST complex still remains under investigation. At least nine nonidentical subunits build the multimeric mammalian and S. cerevisiae OST complex. The mammalian OST complex consist of ribophorin I (Ost1p in S. cerevisiae), ribophorin II (Swp1p), OST48 (Wbp1p), STT3A and STT3B (Stt3), N33 and IAP (Ost3p and Ost6p), and Dad1 (Ost2p) (Kelleher and Gilmore, 1997; Kelleher et al., 2003; Shibatani et al., 2005). In addition, two other putative mammalian OST subunits, DC2 and keratinocyte-associated protein 2 (KCP2), have been identified but their function still remains to be determined (Shibatani et al., 2005).

Several laboratories have shown that the STT3 subunit is the catalytic center (Hua et al., 2008; Nilsson et al., 2003b; Yan and Lennarz, 2002). Two STT3 isoforms exist in mammals, STT3A and STT3B. They share 70% sequence similarity, yet they seem to have different functionality (Kelleher et al., 2003; Wilson and High, 2007).

Ribophorin I is thought to be transiently associated with newly synthesized membrane proteins after their departure from the Sec61 translocon. It is thought to retain potential substrates close to the catalytic subunit of the OST, thereby improving the N-glycosylation efficiency (Wilson et al., 2005; Wilson et al., 2008).

**Signal peptidase**

Proteins that are exported to the cell surface are often synthesized with an N-terminal, cleavable signal sequence. The signal sequences are cleaved by a type I signal peptidase (SPase) that has its catalytic domain on the lumenal (periplasmic) side of the membrane and is anchored to the membrane by one or two transmembrane segments located in the N-terminal part of the protein.

SPases have been identified and characterized in Gram-negative and Gram-positive bacteria, mitochondria, and the ER (Paetzel et al., 2002; Tjalsma et al., 2000). Except for the SPase found in the ER, the SPases are serine-lysine proteases (Karla et al., 2005). The active site serine residue, Ser-90 in *E. coli* SPase, is conserved in the homologous subunit within the ER SPase complex. However, the latter differs from the bacterial and organellar SPases by having a His in place of the catalytic Lys. The SPase in the ER may use either a serine-histidine dyad mechanism or the serine-histidine-aspartic acid triad mechanism (VanValkenburgh et al., 1999).
The SPase and OST complexes appear to be in close contact with the translocon because each acts on a nascent chain as it is being translocated, SPase to cleave off the signal sequence and OST to glycosylate the polypeptide. Rough ER microsomes contain approximately equal numbers of ribosomes, SPase and ribophorin I. It is therefore likely that one OST and one SPase are associated with each translocon (Johnson and van Waes, 1999).
Topology - a structural model of membrane proteins

Membrane protein orientation

To get a better understanding of the structure/function relationships that define the activity of a membrane protein, it is necessary to know its membrane topology, i.e. the number of TM segments and their location relative to the ER membrane. Bioinformatic methods are often used for topology prediction (Melen et al., 2003; Möller et al., 2001; Nilsson et al., 2000b). If the location of one terminus is known, the best programs correctly predict the topology of 65–70% of the membrane proteins with known crystal structure (Melen et al., 2003). The topology prediction reliability is increased if the predictions come from several different programs that agree (Melen et al., 2003; Nilsson et al., 2000b).

What are the determining factors for how the translocon orients a signal sequence or a transmembrane segment relative to the membrane? Membrane proteins with cleavable signal sequences are always oriented with their mature N-terminus on the luminal side, whereas two different orientations, N_in or N_out, are possible for membrane proteins with uncleaved signal sequences.

At least three factors determine the orientation of the TMH. First, the distribution of Arg and Lys between the segments flanking the TMH: the more positively charged segments stay in the cytosol. This is the “positive inside” rule (von Heijne, 1989). Genome wide studies have shown that this charge bias is present in almost all living organisms (Nilsson et al., 2005; Wallin and von Heijne, 1998). However, it is not an absolute truth. Cytoplasmic domains with a net negative charge have been found (Allard and Bertrand, 1992), and there are examples of when the “positive inside” rule can be overridden. This can happen when negatively charged residues are present in high numbers (Nilsson and von Heijne, 1990), when negative charges lie within a window of six residues from the end of a highly hydrophobic TMH (Rutz et al., 1999), or when negative charges flank a marginally hydrophobic TMH (Delgado-Partin and Dalbey, 1998). The molecular mechanism behind the “positive inside” rule and the apparent dominance of positively over negatively charged residues remains to be fully understood.

The second factor is the folding characteristics of the N-terminal domain prior to the TMH: only regions without stable tertiary structure can be translocated (Denzer et al., 1995). The third factor is the length of the hydrophobic sequence: longer sequences favor localization of the N terminus to the ER lumen (Sakaguchi et al., 1992; Wahlberg and Spiess, 1997). In addition, the orientation of the first TM in some multspanning membrane protein
decides the orientation for the following TMHs (Wessels and Spiess, 1988). There are also cases where the internal TMHs have a preferred orientation regardless of the preceding TMHs (Gafvelin and von Heijne, 1994; Locker et al., 1992; McGovern et al., 1991; Nilsson et al., 2000a; Sato et al., 1998).

The lipid composition of the bilayer also has influence on the final topological organization. A recent study on E. coli lactose permease (LacY) (Bogdanov et al., 2008) shows that phosphatidylethanolamine (PE) restrains the translocation potential of negative residues in favor of the cytoplasmic retention potential of positive residues, giving an explanation to the dominance of positive over negative amino acids as co- or post-translational topological determinants.

**Single-spanning membrane proteins**

Single spanning membrane proteins can be classified according to their orientation in the membrane (Figure 7). Type I membrane proteins are targeted to the ER membrane by a cleavable N-terminal signal sequence (Blobel, 1983; von Heijne, 1990). Once in the membrane, they are anchored in the membrane with a stop-transfer (ST) sequence and oriented with an exoplasmic N-terminus and the C-terminus in the cytosol.

Membrane proteins of type II have a cytoplasmic N-terminus and a exoplasmic C-terminus. A signal anchor composed of 18-25 mainly apolar residues anchors the protein in the membrane and induces the translocation of the C-terminal end across the membrane.

Type III membrane proteins are inserted into the membrane with an exoplasmic N- and a cytoplasmic C-terminus. This type of proteins is anchored in the membrane by a reverse signal anchor-sequence that induces the translocation of the N-terminal end across the membrane. Both the signal anchor and the reverse signal anchor are recognized and targeted to the membrane by the SRP.
Figure 7. Schematic illustration of type I, II and III membrane proteins.

Bioinformatics

Determining the topology of a membrane protein however, is a greater challenge than determining its amino acid sequence. Thanks to the fast development of computer technology there are now advanced prediction methods for membrane proteins. It is possible to predict whether a protein is an α-helical membrane protein, its topology, if it contains a signal sequence and its organelle localization. The only requirement is that the amino acid sequence is known.

There are many different prediction algorithms in use today. The first prediction methods simply evaluated the hydrophobicity of individual residues; regions with several hydrophobic residues were predicted to be TM domains (Kyte and Doolittle, 1982). The dense alignment surface (DAS) method analyzes the frequency with which groups of amino acids are found in the TM domains of proteins in the test set (Cserzö et al., 1997). The latest generation of topology-prediction programs uses machine-learning algorithms called hidden Markov models (HMM), which are trained by analyzing the residues that tend to occupy defined regions in integral membrane proteins. Two such algorithms, TMHMM and HMMTOP, assess five or seven (respectively) defined regions of an integral membrane protein, such as the helix core, the TM domain boundaries and cytosolic and luminal parts. Instead of looking at the probability of individual or groups of amino acids to populate each region as in TMHMM, HMMTOP assigns topology by comparing the residues found in one region with those found in other regions (Sonnhammer et al., 1998; Tusnady and Simon, 1998). To evaluate a protein, the programs look for distribution of amino acids in patterns similar to those defined in the training set.

Membrane protein topology prediction programs generally give four different kinds of information: (1) whether or not the protein is likely to be an integral membrane protein; (2) how many TMHs the protein has; (3) the ori-
entations of each of the TMHs; and (4) the boundaries of the membrane and non-membrane domains (Ott and Lingappa, 2002). No program is perfect, and incorrect predictions can appear from several different sources. The hydrophobic core of a soluble protein can be misidentified as a TM domain and the program can miss short TM domains or TM domains containing charged residues.

**Topology mapping**

Multispanning membrane proteins cross the membrane in a zigzag way and expose their hydrophilic loops alternately in the different compartments that are separated by the membrane. To determine the protein topology or verify predicted topology models, the TM domains must be confirmed and the hydrophilic loops must be localized to one compartment (van Geest and Lolkema, 2000).

**PhoA**

One of the first proteins to be used as a topology reporter is PhoA, which has been widely used for identifying periplasmic loops of membrane proteins in *E.coli* (Manoil et al., 1990). Two disulfide bonds have to form in order for PhoA to become catalytically active. Once active it catalyses the hydrolysis of phosphate groups located in the periplasm of bacteria. The disulfide bonds can only form when PhoA is located in the periplasm; if the protein is located in the cytoplasm PhoA will not fold correctly and remains inactive. PhoA is resistant to proteases when located in the periplasm, but not when located in the cytoplasm (Roberts and Chlebowski, 1984). Both of these features can be used to map topology.

**GFP**

Green fluorescent protein (GFP) is a (26.9 kDa) protein that originally comes from the jellyfish *Aequorea victoria* (Shimomura et al., 1962). GFP is widely used as a cytoplasmic reporter as it can be fused to the C-terminal end of the protein of interest (Drew et al., 2002; Rapp et al., 2004). GFP must fold correctly in order to become fluorescent and it can only fold correctly when it is located in the cytoplasm of *E.coli* (Feilmeier et al., 2000).

**Glycosylation mapping**

An example of a target site that can be used for topology mapping is the N-glycosylation site that can be introduced at specific positions in the protein by site-directed mutagenesis. N-linked glycosylation can only occur in the ER lumen, due to the location of the active site of OST. Addition of the oligosaccharide to a target site in a protein results in an increase of the molecular mass by ~2 kDa compared to the unglycosylated form and can be detected by SDS-PAGE (Chavez and Hall, 1991).
In the glycosylation mapping approach, the lumenally oriented active site of the OST is used as a fixed point of reference against which the position of a TMH in the ER membrane is measured. Point mutations in a TMH that affect the position in the membrane will change the minimal glycosylation distance (Nilsson and von Heijne, 2000). The minimal glycosylation distance (MGD) is the number of residues in the nascent chain needed to bridge the distance between a given reference residue at the luminal end of the TMH and the OST active site (Monne et al., 1998). For a typical TMH, the MGD is about 10 residues (Figure 8).

![Figure 8. Determination of the "minimal glycosylation distance" (MGD). The MGD is defined as the minimum number of residues required between the luminal end of a transmembrane segment (white) and the active site of oligosaccharyl transferase (Y) (Mingarro et al., 2000).](image)

**Protease-protection assay**

The proteases trypsin and proteinase K can be added to spheroplasts or microsomes in order to confirm the topology of a membrane protein. These enzymes degrade all exposed domains, while unexposed parts localized in the bacterial cytoplasm or ER lumen remain undegraded. The protease-protected part can then be detected by SDS-PAGE.

**Epitope tags**

Epitope tagging is a recombinant DNA method by which a protein encoded by a cloned gene is made immunoreactive to a known antibody. Protein detection is then performed by highly specific anti-tag antibodies, eliminating the need for protein-specific antibodies. Commonly used epitope tags include glutathione-S-transferase (GST), cMyc, hexahistidine (6X-His), FLAG, maltose binding protein (MBP), β-galactosidase, GAL4 and GFP (Patton, 2002).
In yeast the HA-tag (hemagglutinin influenza virus epitope tag (Kolodziej and Young, 1991)) is commonly used as an epitope tag. The HA-tag is a small segment of the viral hemagglutinin coat protein that can be incorporated in gene expression vectors as an epitope tag. Hemagglutinin antibodies then recognize the HA-tag, and the expressed protein can then be determined by Western blotting.
Methodology

Model system - Leader peptidase
Leader peptidase (Lep) is a 233 amino acid long protease located in the inner membrane of *E. coli*, where its function is to remove signal peptides from secretory proteins. It is anchored to the membrane by two N-terminal transmembrane segments, H1 and H2, and has a small polar cytoplasmic domain (P1) and a large C-terminal periplasmic domain (P2) (Wolfe et al., 1983). The H1 segment is more hydrophobic than H2 and can insert into the lipid bilayer in the absence of H2. H2 on the other hand needs the aid of H1 for proper membrane integration (Heinrich and Rapoport, 2003). The X-ray crystal structure of the soluble P2 domain and site-directed mutagenesis studies have shown that Ser 90 and Lys 145 in the P2 domain are essential for the catalytic activity (Paetzel et al., 1998; Sung and Dalbey, 1992; Tschantz et al., 1993).

![Figure 9. Topology of wild-type leader peptidase.](image)

27
Lep inserts co-translationally into microsomes when translated in vitro and adopts its native topology with the P1 loop facing the cytoplasm and both the N and C-termini in the ER lumen (Nilsson and von Heijne, 1993). Lep can be glycosylated in a eukaryotic in vitro transcription/translation system supplemented with dog pancreas microsomes. A single asparagine (N)-linked glycosylation site (Asn215-Glu-Thr) is present in the large periplasmic domain (P2) of Lep (wt).

**Expression in vivo**

There are many expression vectors and host cell systems available for eukaryotic protein expression. Bacterial expression systems, especially those based on E. coli vectors, are often used. However, these have turned out to be less suitable for overexpression of mammalian membrane proteins. The main reason for this is that many post-translational modifications and processing events that are required for correct folding of mammalian proteins, do not exist in prokaryotic cells. Additionally, insertion of recombinantly expressed mammalian membrane proteins into bacterial membranes may cause severe toxicity to the host cells, or induce the formation of inclusion bodies.

Semliki Forest virus (SFV) vectors have been applied for the expression of recombinant integral membrane proteins in a wide range of mammalian host cells (Lundstrom, 2003). The SFV expression system is based on the expression vector pSFV1, derived from the Semliki Forest virus, which is an enveloped single-stranded RNA virus. The vector encodes the single stranded RNA genome of a recombinant form of the virus in one operon, into which the target gene is cloned. The recombinant virus is unable to form viral particles due to the absence of three structural proteins necessary for particle formation (Liljestrom and Garoff, 1991). If the recombinantly produced RNA is synthesized in vitro and subsequently transfected into cells, it will drive a self-replication process and its own capping, leading to massive production of viral and recombinant proteins, while competing out the host protein synthesis.
Expression in S. cerevisiae

*S. cerevisiae* is another expression system for eukaryotic proteins that has many of the same capabilities as mammalian cells to secrete and modify proteins.

An efficient transformation system for *S. cerevisiae* is based on homologous recombination (Oldenburg et al., 1997). *In vivo* recombination takes place between the vector and a restriction fragment whose ends bear homology to plasmid sequences, removing the need for an *in vitro* ligation reaction (Ma et al., 1987).
Results and discussion

Summary of papers
The major aim of this thesis has been to define the sequence requirements for insertion of membrane proteins into the ER membrane. We have also performed topology mapping of two biologically interesting membrane proteins with previously unknown topology, the human seipin protein and the *Drosophila* odorant receptor OR83b.

Paper I: Helix-helix interactions
We wanted to investigate whether and to what extent inter-helix hydrogen bonding can drive the process of translocon-mediated TM helix insertion itself, and if the separation between the two helices within the sequence may influence these interactions.

It is known that hydrogen bonding between successive TM helices can have a favorable effect on TM helix interactions in detergent micelles or model lipid membranes. A single Asn or Asp in a TM helix provides a strong driving force for dimer or trimer formation (Choma et al., 2000; Gratkowski et al., 2001; Zhou et al., 2000). Because dimerization occurs in a region of low dielectric constant, it is likely that the carboxyl-containing side chain of Asp is protonated. It has long been known that carboxylic acids form strong hydrogen bonds, especially in a low dielectric solvent (Pauling and Brockway, 1934; Pauling and Sherman, 1934).
To measure the effect of Asp- or Asn-mediated interactions between H2' and H segments, two constructs were compared for each H segment: one with a 19-Leu H2’ segment (I) and one with an H2’ segment in which one or two of the Leu residues have been replaced by Asp or Asn residues (II).

By analyzing model protein constructs in which zero, one, or two Asn or Asp residues were placed in two neighboring hydrophobic segments (Figure 10), it was found that ΔG_{app} (contribution to the free energy of membrane integration of a marginally hydrophobic H-segment) was significantly reduced only if both the H2’-segment and the H-segment contained two Asn or Asp- residues spaced one helical turn apart in both H2’ and H-segments. We conclude that inter-helix interactions, which are important in natural proteins (Heinrich and Rapoport, 2003; Mitra et al., 2005), can be treated as a second-order correction in the calculation of ΔG_{app} values based on the biological hydrophobicity scale (Hessa et al., 2005).
Paper II: Opposite orientation: $N_{in}$-$C_{out}$ vs. $N_{out}$-$C_{in}$ transmembrane helices

Transmembrane α-helices in integral membrane proteins can have two orientations in the membrane: $N_{in}$-$C_{out}$ or $N_{out}$-$C_{in}$. Previous studies of model $N_{out}$-$C_{in}$ transmembrane segments have led to a detailed, quantitative picture of the “molecular code” that relates amino acids sequence to membrane-insertion efficiency in vivo (Hessa et al., 2007). However, if the same code also applies to $N_{in}$-$C_{out}$ transmembrane helices was not known. In this project, we found that the contributions of individual amino acids to the overall efficiency of membrane insertion are similar for the two kinds of helices, and that the threshold hydrophobicity for membrane insertion can be up to ~1 kcal/mol lower for $N_{in}$-$C_{out}$ compared to $N_{out}$-$C_{in}$ transmembrane helices, depending on the neighboring helices.

The mechanisms of translocon-mediated recognition of $N_{out}$-$C_{in}$ and $N_{in}$-$C_{out}$ TM helices thus seem to be very similar, even though $N_{in}$-$C_{out}$ TM helices must re-initiate translocation of the downstream part of the protein while $N_{out}$-$C_{in}$ TM helices enter the translocon channel as part of a translocating nascent chain (Figure 11).

**Figure 11.** Membrane insertion of $N_{out}$-$C_{in}$ and $N_{in}$-$C_{out}$ transmembrane helices. (A) $N_{out}$-$C_{in}$ TM helices enter the translocon channel as part of a translocating nascent chain. (B) The entry of the H-segment into the translocon triggers translocation of the downstream part of the nascent chain. The H-segment integrates into the membrane with an $N_{in}$-$C_{out}$ orientation.
We also found that when the TM helix serves as an ER-targeting signal it has to interact with SRP before encountering the translocon, consequently the threshold hydrophobicity is markedly higher. This suggests that many TM helices in multi-spanning membrane proteins will not by themselves be able to trigger SRP-mediated targeting to the ER translocon.
Paper III: Insertion efficiency of the amyloid β-peptide into the ER membrane

In this study we wanted to investigate the possible retention in the ER membrane of different amyloid β-peptides (Aβ) from the amyloid precursor protein (APP). We chose different Aβ-fragments corresponding to those known to be produced in vivo from APP due to processing by β-γ-ε- and ζ-secretases.

![Diagram](image)

**Figure 12.** Model for Aβ-fragment integration. Mature APP is cleaved by β-secretases, releasing the soluble N-terminal fragment (APPβ). The C-terminal fragment (CTF) is produced by the ε-secretase cleavage together with the Aβ 48-49 membrane-bound fragments. Longer Aβ fragments are further processed by ζ/γ-secretases, leaving Aβ 46 partly in the membrane while shorter fragments (Aβ 40-45) are released into the ER lumen.

Using *in vitro* translation in the presence of microsomes, we demonstrated that different Aβ segments are integrated into the ER membrane to different degrees, depending on their overall length, hydrophobicity and helix potential (Figure 12). This might also affect their access to the secretases in the ER membrane and the ability of Aβ to form toxic aggregates.
Paper IV and V: Topology mapping of two eukaryotic membrane proteins

Using the glycosylation mapping technique both in vitro and in baby hamster kidney (BHK) cells, we determined the topology of two membrane proteins with unknown topology, the human seipin protein and the *Drosophila* odorant receptor OR83b. We first determined the topology by using multiple prediction programs.

The seipin protein was predicted to be a transmembrane protein with two transmembrane segments between residues 95–117 and residues 294–316 respectively. Insect ORs lack homology to G-protein-coupled receptors in vertebrates (Wistrand et al., 2006) and were predicted to be transmembrane proteins with seven transmembrane segments and an intracellular N-terminus and extracellular C-terminus. These predictions were then tested by in vitro transcription/translation of a series of variants with glycosylation sites added to different parts of the proteins.

Our results suggest a N<sub>cyt</sub>-C<sub>cyt</sub> topology for the 462 residue long seipin protein and a N<sub>cyt</sub>-C<sub>exo</sub> topology for the OR83b (Figure 13). Our result have also been confirmed by later studies, showing that seipin is an ER membrane-resident protein with a luminal loop and with both N- and C- termini facing the cytoplasm (Ito et al., 2008).

Other studies on the OR83b placed its N-terminus in the cytoplasm (Benton et al., 2006) and showed that OR83b is a conserved member of the insect OR family and heterodimerises with other ORs, forming receptor complexes.
Recent work suggests that heteromeric insect ORs comprise a new class of ligand-activated non-selective cation channels with no homology to any previously described ion channel (Sato et al., 2008).
Acknowlegements

Tack till alla som förgyllt min doktorandtid, ett särskilt STORT TACK till:

Gunnar, min handledare, för din oändliga kunskap om membranproteiner och din entusiasm inför alla resultat, bra som dåliga. Tack för all uppmuntran och frihet du har gett mig genom åren.

IngMarie, min andra handledare, för din kunskap, arbetsamhet och hjälpsamhet.

Joy, the “yeast guru”, for your kindness, for reading my thesis and giving me the opportunity to work with you.

Stefan, för ditt stöd, uppmuntran och att du alltid är rättvis.

Agneta, för att jag alltid får nya utmaningar i undervisnings värld.

Peter Brzezinski och Birgitta Norling för att ni varit mina engagerade utvärderare.


Jan-Willem, for importing Jorrit, dankjewel! For stealing pepernoten and always being enthusiastic about science.
Past and present members of the GvH cluster: Isolde, Salomè, Roger, Tara, Marika, Mikaela, Yoko, Dan, David W, Sam, Louise, Mirjam K, Susan, David D, thank you for your support and knowledge. Keep up the good work!

Everyone at CBR and DBB corridors, thank you!

Mina lärare på Yogashala Stockholm, som lär mig både om yoga och livet!

Danshögskolan: Anna Karin, Marre, Kicki, Anette, Monica för den fantastiska tiden på Confidencen!

Mina vänner som stått ut med mig under avhandlingsskrivandet! Särskilt tack till: Karolin och Alex, för alla roliga stunder på molekylärbiologlinjen, Tovetorp, möhippor och bröllop! Minna, för att du har orkat lyssna på mina doktorandproblem genom åren och alltid är en fantastisk vän!


Mamma, för din omsorg och ditt stöd. Tack för alla miljoner gånger du ställt upp för mig och Jorrit och för all uppmuntran.

Jorrit, för att du får mig att skratta. För all kärlek, glädje och lycka! Du är bäst!
References


Finke, K., K. Plath, S. Panzer, S. Prehn, T.A. Rapoport, E. Hartann, and T. Sommer. 1996. A second trimeric complex containing homologs of


Nilsson, I., S. Witt, H. Kiefer, I. Mingarro, and G. von Heijne. 2000a. Distal downstream sequence determinants can control N-tail transloc-


