Marginally hydrophobic transmembrane $\alpha$-helices shaping membrane protein folding –

Tuuli Minttu Virkki de Marothy
Marginally hydrophobic transmembrane $\alpha$-helices shaping membrane protein folding

Tuuli Minttu Virkki de Marothy
Abstract

Most membrane proteins are inserted into the membrane co-translationally utilizing the translocon, which allows a sufficiently long and hydrophobic stretch of amino acids to partition into the membrane. However, X-ray structures of membrane proteins have revealed that some transmembrane helices (TMHs) are surprisingly hydrophilic. These marginally hydrophobic transmembrane helices (mTMH) are not recognized as TMHs by the translocon in the absence of local sequence context.

We have studied three native mTMHs, which were previously shown to depend on a subsequent TMH for membrane insertion. Their recognition was not due to specific interactions. Instead, the presence of basic amino acids in their cytoplasmic loop allowed membrane insertion of one of them. In the other two, basic residues are not sufficient unless followed by another, hydrophobic TMH. Post-insertional repositioning are another way to bring hydrophilic residues into the membrane. We show how four long TMHs with hydrophilic residues seen in X-ray structures, are initially inserted as much shorter membrane-embedded segments. Tilting is thus induced after membrane-insertion, probably through tertiary packing interactions within the protein.

Aquaporin 1 illustrates how a mTMH can shape membrane protein folding and how repositioning can be important in post-insertional folding. It initially adopts a four-helical intermediate, where mTMH2 and TMH4 are not inserted into the membrane. Consequently, TMH3 is inserted in an inverted orientation. The final conformation with six TMHs is formed by TMH2 and 4 entering the membrane and TMH3 rotating 180°. Based on experimental and computational results, we propose a mechanism for the initial step in the folding of AQP1: A shift of TMH3 out from membrane core allows the preceding regions to enter the membrane, which provides flexibility for TMH3 to re-insert in its correct orientation.
List of Papers

The following papers, referred to in the text by their Roman numerals, are included in this thesis.

PAPER I: The positive-inside rule is stronger when followed by a trans-membrane helix
DOI: 10.1016/j.jmb.2014.06.002

PAPER II: Insertion of marginally hydrophobic helix in EmrD
Virki M.T., Peters C., Cristobal S., Elofsson A. Manuscript,

PAPER III: Large tilts in transmembrane helices can be induced during tertiary structure formation
DOI: 10.1016/j.jmb.2014.04.020

PAPER IV: Folding of Aquaporin 1: multiple evidence that helix 3 can shift out of the membrane core
DOI: 10.1002/pro.2483

Reprints were made with permission from the publishers.
The following papers have not been included in this thesis:

PAPER V: **Improved production of membrane proteins in *Escherichia coli* by selective codon substitutions**
DOI: [10.1016/j.febslet.2013.05.063](https://doi.org/10.1016/j.febslet.2013.05.063)

PAPER VI: **Manipulating the genetic code for membrane protein production: what have we learnt so far?**
DOI: [10.1016/j.bbamem.2011.08.018](https://doi.org/10.1016/j.bbamem.2011.08.018)
# Contents

Abstract iv

List of Papers vii

Abbreviations xiii

1 Introduction 15

2 The hydrophobic effect 19
   2.1 The hydrophobic effect and membrane lipids . . . . . . . . 20
   2.2 The hydrophobic effect and proteins . . . . . . . . . . . . 20

3 Biological membranes 23
   3.1 Membrane lipids . . . . . . . . . . . . . . . . . . . . . . . . 23
      3.1.1 Physical properties of lipids shape the characteristics
           of a biological membrane . . . . . . . . . . . . . . . . 23
      3.1.2 The lipid environment in biological membrane is hetero-
           genic . . . . . . . . . . . . . . . . . . . . . . . . . . 25
   3.2 The fluid mosaic model and beyond . . . . . . . . . . . . . 26
   3.3 Lipid-protein interactions . . . . . . . . . . . . . . . . . . . 28
      3.3.1 Hydrophobic mismatch . . . . . . . . . . . . . . . . . . 28
      3.3.2 Membrane properties shaping protein function . . . . . 28
      3.3.3 Implications of specific lipid properties in membrane
           protein function . . . . . . . . . . . . . . . . . . . 29

4 Protein machineries in the biogenesis of α-helical membrane pro-
   teins 31
   4.1 The SRP-dependent pathway . . . . . . . . . . . . . . . . . . 31
   4.2 The Sec translocon . . . . . . . . . . . . . . . . . . . . . . . 34
   4.3 Associated proteins . . . . . . . . . . . . . . . . . . . . . . . 35
5 Properties of amino acids in membrane protein structure

5.1 Defining amino acid hydrophobicity

5.2 The properties of amino acids in membrane proteins

5.2.1 Hydrophobic and helix-forming amino acids promote membrane insertion

5.2.2 Charged amino-acid side chains in transmembrane helices

5.2.3 Aromatic amino acids are unequally distributed in transmembrane helices

5.2.4 Proline and glycine in transmembrane segments

5.3 Positive inside rule in establishing topology

5.4 The two stage model and topology

5.4.1 The first stage - establishing topology

5.4.2 The second stage - membrane protein folding

6 Non-sequential membrane integration of α-helical membrane proteins

6.1 Marginally hydrophobic helices

6.1.1 Consequences of mTMH in AQP1 folding

6.1.2 Cost of polar groups in the membrane

6.2 Sequence features implicated in the insertion of mTMHs

6.2.1 Positive-inside rule

6.2.2 Specific interactions between polar groups

6.2.3 Repositioning in the membrane

6.3 Can translocon influence hydrophobicity threshold or topology?

6.4 The membrane environment

7 Methodology - studying membrane proteins

7.1 Biochemical determination of topology

7.1.1 in vitro protein expression in the presence of microsomes

7.1.2 N-linked glycosylation in topology mapping

7.1.3 Leader peptidase as a model protein

7.1.4 Determining the propensity to integrate into the membrane by an in vitro assay

7.1.5 Minimal glycosylation distance mapping

7.2 Membrane proteins in silico

7.2.1 Topology predictions

7.2.2 Computational estimation of membrane integration propensity of a transmembrane helix

7.2.3 Understanding protein function with molecular dynamics
8 Summary of papers

8.1 Paper I - The positive-inside rule is stronger when followed by a transmembrane helix

8.2 Paper II - Insertion of marginally hydrophobic helix in EmrD (Manuscript)

8.3 Paper III - Large tilts in transmembrane helices can be induced during tertiary structure formation

8.4 Paper IV - Folding of Aquaporin 1. multiple evidence that helix 3 can shift out of the membrane core

8.5 Final thoughts

Sammanfattning

Tiivistelmä

Acknowledgements

References
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala,A</td>
<td>Alanine</td>
</tr>
<tr>
<td>AQP1</td>
<td>Aquaporin 1</td>
</tr>
<tr>
<td>AQP4</td>
<td>Aquaporin 4</td>
</tr>
<tr>
<td>Arg,R</td>
<td>Arginine</td>
</tr>
<tr>
<td>Asn,N</td>
<td>Asparagine</td>
</tr>
<tr>
<td>Asp,D</td>
<td>Aspartate</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>Cys,C</td>
<td>Cysteine</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DOPC</td>
<td>1,2-Dioleoyl-sn-glycero-3-phosphocholine</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic reticulum</td>
</tr>
<tr>
<td>Gln,Q</td>
<td>Glutamine</td>
</tr>
<tr>
<td>Glu,E</td>
<td>Glutamate</td>
</tr>
<tr>
<td>Gly,G</td>
<td>Glycine</td>
</tr>
<tr>
<td>GTP</td>
<td>Guanosine-5’-triphosphate</td>
</tr>
<tr>
<td>His,H</td>
<td>Histidine</td>
</tr>
<tr>
<td>Ile,I</td>
<td>Isoleucine</td>
</tr>
<tr>
<td>K⁺</td>
<td>Potassium ion</td>
</tr>
<tr>
<td>Leu,L</td>
<td>Leucine</td>
</tr>
<tr>
<td>Lys,K</td>
<td>Lysine</td>
</tr>
<tr>
<td>Met,M</td>
<td>Methionine</td>
</tr>
<tr>
<td>mTMH</td>
<td>Marginally hydrophobic transmembrane α helix</td>
</tr>
<tr>
<td>OST</td>
<td>Oligosaccharyl transferase</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PE</td>
<td>Phosphatidylethanolamine</td>
</tr>
<tr>
<td>Phe,F</td>
<td>Phenylalanine</td>
</tr>
<tr>
<td>PIP2</td>
<td>Phosphatidylinositol 4,5-bisphosphate</td>
</tr>
<tr>
<td>Pro,P</td>
<td>Proline</td>
</tr>
<tr>
<td>R1-H3 loop</td>
<td>A region in Aquaporin 1 containing the re-entrant region 1 and the loop between it and transmembrane helix 3</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acids</td>
</tr>
<tr>
<td>RNC</td>
<td>Ribosome-nascent chain complex</td>
</tr>
<tr>
<td>Sec</td>
<td>Secretase translocon</td>
</tr>
<tr>
<td>Ser,S</td>
<td>Serine</td>
</tr>
<tr>
<td>SPC</td>
<td>Signal peptidase complex</td>
</tr>
<tr>
<td>SR</td>
<td>Signal recognition particle receptor</td>
</tr>
<tr>
<td>SRP</td>
<td>Signal Recognition Particle</td>
</tr>
<tr>
<td>Thr,T</td>
<td>Threonine</td>
</tr>
<tr>
<td>TMH</td>
<td>Transmembrane α helix</td>
</tr>
<tr>
<td>TRAM</td>
<td>Translocon-associated membrane protein</td>
</tr>
<tr>
<td>TRAP</td>
<td>Translocon-associated protein complexes</td>
</tr>
<tr>
<td>Trp,W</td>
<td>Tryptophan</td>
</tr>
<tr>
<td>Tyr,Y</td>
<td>Tyrosine</td>
</tr>
<tr>
<td>Val,V</td>
<td>Valine</td>
</tr>
<tr>
<td>vdW</td>
<td>van der Waal</td>
</tr>
</tbody>
</table>
1. Introduction

The basic unit of life, the cell, got its name from the cell walls of dead cork cells, which looked like the sleeping chambers of monks to Robert Hooke. He thus named these structures after the Latin word "cella" meaning a small room [1]. Soon after, the first living, single-cell organisms were discovered by Anton van Leeuwenhoek who called them animalcules [2]. From the observations made during the following decades, the observations made by Schleiden, Schwann, Remak and Vrichow later established that all organisms are composed of one or more cells and that cells only arise from the division from a preexisting cells [3].

Despite the differences between a tiny bacterium Eschericia coli and a huge blue whale (Balaenoptera musculus), the structures of their cell(s) are very similar. They all contain a plasma membrane - a structure separating the interior of the cell, the cytoplasm, from its environment. The cytoplasm consists of water, salts, organic molecules (such as glucose or amino acids) and macromolecules (such as proteins and lipids). Many of the metabolic pathways - series of chemical reactions occurring within a cell - are shared, as are the proteins regulating and catalyzing these reactions.

"... anything found to be true of E. coli must also be true of elephants - only more so."
Jacques Lucien Monod

![Figure 1.1: The basic structure of a) a prokaryotic cell and b) an eukaryotic cell (by Minttu Virkki).](image)
Deoxyribonucleic acid (DNA) is a macromolecule encoding genetic instructions used in the development and functioning of all known living organisms. It is a long polymer made from (repeating units called) nucleotides. The order of these nucleotides specifies the genetic code - a set of rules specifying which regions corresponds to genes, how these genes are transcribed into ribonucleic acids (RNA) then finally translated into proteins by the ribosomes. Depending on how DNA is stored, cells can be grouped into eukaryotes (DNA is contained within a membrane-enveloped nucleus) and prokaryotes (DNA is not segregated, i.e. no nucleus). While much of the cellular structures are shared between pro- and eukaryotes, there are some differences (See Figure 1.1). Eukaryotes contain organelles, membrane enclosed subunits with specific function.

![Figure 1.2: A biological membrane depicted according to the fluid-mosaic model. Adapted from an illustration by Mariana Ruiz Villarreal.](image)

If the cell is viewed as a room, the membrane would be its walls. However, without a special class of protein - membrane proteins - the cell would not be alive. Membrane proteins are the equivalent of water pipes, electric cables, windows and doors for the cell. They take up nutrients, keep track of extracellular conditions, catalyze chemical reactions and communicate with other cells. A typical organism devotes about a quarter of its genes to produce integral membrane proteins [4–6]. Their importance is further implied by the fact that more than half of the targets in the pharmaceutical industry are...
 membrane proteins [7][8]. Lastly, the author of this thesis would like to argue, that these seven pillars of life - program, improvisation, compartmentalization, energy, regeneration, adaptability, seclusion - are all dependent on biological membranes [9].

I will begin my thesis by presenting the hydrophobic effect - fundamental to biochemistry and research on membrane protein insertion. In the third chapter I will introduce membrane lipids, describe biological membranes and discuss their dynamics. The protein machinery essential to membrane protein biogenesis is briefly reviewed in the fourth chapter followed by a description of alpha-helical integral membrane protein in the fifth.

In the sixth chapter I will introduce marginally hydrophobic transmembrane helices (mTMHs). These transmembrane segments are deficit in the defining character of transmembrane helices (TMHs) - hydrophobicity. Nevertheless, they are fairly common in membrane proteins. The seventh chapter will introduce the experimental and computational methods relevant to the thesis and finally, the last chapter summarizes the work done in this thesis with some concluding remarks.
2. The hydrophobic effect

Water molecules are electric dipoles, where the electronegative oxygen has a partial negative charge and the small hydrogen atoms have a partial positive charge. Water molecules therefore have dipole interactions with each other as well as to other molecules. The dipole interaction between a hydrogen atom bonded to an electronegative nucleus (nitrogen, oxygen, fluoride and chloride in particular) and a second such nucleus is termed a hydrogen bond. The hydrogen bond is particularly strong, mainly due to the large dipole moment caused by the electronegative nuclei’s polarization of the X-H bond (where X is an electronegative atom and H a hydrogen), but also due to a partial covalent character. Liquid water forms a tightly hydrogen bonded network between the molecules (typically 3 hydrogen-bonds/molecule), and the relative strength of these interactions is demonstrated clearly by the textbook comparison (e.g. Pauling) to hydrogen sulfide, which is gaseous at room temperature. Hydrogen bonds are the strongest intermolecular interaction between overall neutral molecules.

When liquid water forms an interface at the boundary to some other phase to which it does not bond (e.g. a vacuum or a negligibly-interacting gas), the water molecules at this boundary can not form as many hydrogen bonds since they are not surrounded by other molecules on all sides. The orientations of the molecules at the interface are also more restricted, resulting in a decrease in entropy. Molecules at the surface therefore have a higher energy than those in the bulk, resulting an effective force "pulling" the surface molecules towards the bulk - the well-known phenomenon of surface tension.

Another well-known and related fact is the observation that "oil and water don’t mix", known as the hydrophobic effect in chemical terminology. If a non-polar molecule is placed in bulk water, it disrupts the hydrogen-bond network of the water, in effect creating a 'bubble' around the non-polar molecule. The energetic cost of this may be accounted in two parts: First, the energy required by surface tension to create an interface and form a void in the water. Second, the energy gained from the weak van der Waals (vdW) interactions (Debye, London forces) that exist between the non-polar molecule placed into the void and the surrounding water. Since vdW interactions are much weaker than hydrogen bonds, the former term is guaranteed to dominate in the case of a purely non-polar molecule, and it will cost energy to place the non-polar
molecule into the bulk water. The energetically favored situation is for the non-polar molecules to "lump together" into droplets or bubbles, and ultimately a separate phase, in order to minimize the interface area with the water. The non-polar substance is immiscible in water.

The term “hydrophobic” can be misleading. While bulky hydrophobic molecules may form strong intermolecular bonds (e.g. oils have a higher boiling point than water), they do not bond to each other more strongly than they do to water in terms of the interfacial area. Non-polar molecules interact with water through the Debye (induced dipole moment) and London (dispersion) interactions, while non-polar molecules bind to each other through the weaker (but longer-range) London force alone. Therefore, the driving force behind the hydrophobic effect is not that non-polar molecules bind to each other more strongly than to water, but rather that water binds to itself more strongly than it does to non-polar molecules. Although a completely non-polar molecule is immiscible, hydrophobicity is not an either or phenomenon. The more polar and hydrogen-bonding a molecule is in relation to its size, the more the energetic cost of creating a void in the water is offset. Hence methanol is entirely miscible in water, 1-pentanol is weakly soluble, while 1-decanol is normally considered insoluble.

2.1 The hydrophobic effect and membrane lipids

Biological membranes are composed of amphiphatic lipids, consisting of a hydrophilic head group and a hydrophobic acyl chain (see Figure 2.1). Despite the head group, they are still largely immiscible in water, and therefore aggregate spontaneously. They will also arrange themselves with their head groups at the water interface, where those groups may form hydrogen and polar bonds to the surrounding bulk water. This lowers the interfacial energy sufficiently that a stable emulsion can be formed, rather than lipids aggregating into an "oily" phase. Surface tension will lead to the formation of a spherical micelle (Figure 2.1c), unless the acyl chains are sufficiently bulky for their steric repulsion to counter it. In the latter case a stable lipid bilayer membrane is formed, with oppositely-oriented layers of the amphipilic lipids (Figure 2.1d).

2.2 The hydrophobic effect and proteins

Amino acids are polar compounds with good water solubility (Figure 2.2a). On forming peptide bonds, the charged groups become the amide and keto groups of the peptide backbone, which are hydrophilic but substantially less so than the original charged groups (Figure 2.2b). The amino-acid side chains
Figure 2.1: Lipids and the hydrophobic effect. a) The chemical structure represented as bond-line structure for phosphatidylcoline. Both the hydrophobic and hydrophilic region of the lipid molecule are indicated. b) A schematic presentation for a general lipid structure. The head-group is depicted as a sphere and the tails as drawn lines. c) A schematic drawing of a micelle. d) A schematic drawing of a lipid bilayer. By Minttu Virkki.

(denoted as R in (Figure 2.2) may be polar, charged or hydrophobic. In addition, the folding and geometry of the protein can make the various groups inaccessible from the surrounding medium. Therefore, proteins can exhibit a wide range of hydrophobicity, both as a whole and in localized regions of the protein.

Globular proteins, which exist in the cytosol, are hydrophilic in the sense that they are water-soluble, which in turn implies that their surfaces bind strongly to water relative the protein’s interfacial area. The hydrophobic effect will manifest itself as a force working to bring hydrophobic portions of the protein together, away from the bulk solution and more hydrophilic regions. Strongly
Figure 2.2: Amino acids and the peptide bond. a) The chemical structure of an amino acid. R denotes the position of a side chain, which varies between amino acids. b) The peptide bond is formed when the carboxyl group of one amino acid molecule reacts with the amino group of another amino acid, causing the release of a molecule of water. By Minttu Virkki.

Hydrophobic regions are concentrated towards the interior of the protein. This, and other forms of inter- and intramolecular bonding within the protein results in a specific three-dimensional structure of the protein, determined mainly by its primary amino-acid sequence [14]. For membrane proteins, the hydrophobic effect plays a more complicated role and is discussed in more detail in later chapters.
3. Biological membranes

Biological membranes are mainly composed of lipids and proteins. Together they form an essential barrier between living cells and their external milieu. In addition, they allow compartmentalization of intracellular organelles within eukaryotes. While membrane lipids have been seen as merely solvents for membrane proteins, the picture today is more complex; Biological membranes are dynamic structures, that vary in their lipid, protein and carbohydrate composition co-evolving and functioning together [15]. Indeed, the role of lipids in membrane protein structure [16, 17], topology [18], function [19] and in signaling [20] have become evident.

3.1 Membrane lipids

The bulk of biological membrane consists of lipids. The lipid composition determines the physical properties of the membrane, defining the surface charge, thickness, fluidity and curvature. All these characteristics must be maintained within an appropriate range and can be adjusted depending on the changes in environmental conditions [21-23]. To achieve this, different types of lipids are needed, albeit membrane lipids share the same general structure with one polar and one non-polar region.

Phospholipids are the major component of all cell membranes. Glycerocephospholipids consist of a glycerol backbone with two hydrophobic acyl chains attached via ester linkage to the first and second carbons. The third glycerol carbon is attached to a polar or charged head group through a phosphodiester linkage [21]. The fatty acid chains vary; in general carbon one is usually linked to a C16 or C18 saturated fatty acid whereas the second carbon usually bears a C18 or C20 unsaturated fatty acid [24], see Figure 3.1.

3.1.1 Physical properties of lipids shape the characteristics of a biological membrane

The lipid composition of membranes varies between different organisms, cell types and in time. In addition to the plasma membrane, eukaryotic cells contains a number of internal membranes, each with a specialized set of lipids and
Figure 3.1: Structure of a phospholipid. The structure of phospholipids is illustrated with phosphatidylcholine. The head group consists of the glycerol backbone (in green), linked through a phosphate (in cyan) to a polar or charged head group (in magenta). Here this head group is choline, but it can be replaced by for instance ethanolamine in phosphatidylethanolamine. The two remaining carbons of the glycerol are attached to two fatty acids through ester linkage. The fatty acids can be saturated or unsaturated (in yellow). The saturation status as well as the properties of the head group determine the physiochemical properties of a phospholipid. By Minttu Virkki.

Figure 3.2: Cylindrical lipids are prone to forming bilayers and are abundant in biological membranes. However, non-bilayer lipids are also present in the membrane; Cone-shaped lipids have head groups with a cross-sectional area smaller than their acyl chains. They promote a negative curvature in membranes whereas lipids with a inverted-cone shape tend to form micelles and promote positive curvature in membranes [28] (Figure 3.2). Ultimately, the ratio between bilayer- and non-bilayer forming lipids determines the intrinsic curvature of proteins [23]. Organisms can devote up to 5% of their genome towards lipid metabolism [23, 25] and the number of different lipids range from several hundreds in bacteria to up to thousands in eukaryotes [26]. While the combination of different lipid head groups shape the characteristics of the membrane surface, the physical properties of the membrane are largely dependent on the head groups and various fatty acid side chains [27].
the membrane as well as shape and integrity [19; 21; 29]. Additionally, it influences the membrane flexibility, which is important for fusion/fission events as well as for membrane protein function [15; 30; 31].

Membrane fluidity depends on lipid types and acyl chain composition. Saturated fatty acids have a linear acyl chain allowing it to pack tightly, whereas unsaturated fatty acids contain kinks resulting in increased fluidity. In addition, incorporation of rigid steroidal lipids, such as cholesterol, confer stability [21]. Membrane fluidity is also dependent on temperature and some organisms can adjust their lipid composition in response to changes of temperature [32]. As an example, some bacteria can maintain membrane fluidity at both higher and lower growth temperatures by increasing the number of saturated and unsaturated fatty acids, respectively [33; 34].

3.1.2 The lipid environment in biological membrane is heterogenic

The structure for a 1,2-Dioleoyl-sn-glycero-3-phosphocholine (DOPC) lipid bilayer has been solved using X-ray and neutron scattering data, for illustration see Figure 3.3. The 30 Å thick hydrophobic core is lined with a 15 Å thick interface region on each side. Transition from the core to the interfaces is not sharp and should be viewed as a zone with gradual change in hydrophobicity [36]. The interfacial regions can also vary between the two leaflets. While the lipids are uniformly distributed on both faces of the bilayer in the endoplasmic reticulum membrane and the Golgi apparatus, the two leaflets of plasma membranes are asymmetric [23].

A common feature in many animal cells is that the extracytoplasmic leaflet contains most of the phosphatidylcholine, sphingomyelin, and glycosphingolipids, while phosphatidylserine and phosphatidylethanolamine are enriched
in the cytoplasmic leaflet \[37\]. How the asymmetry is established and maintained is not well understood but three major players have been reported; Aminophospholipid translocase, or flippase, moves phosphatidylserine and phosphatidylethanolamine to the cytosolic leaflet of plasma membranes while floppases move phosphatidylcholine, sphingolipids and cholesterol in the opposite direction - both requiring ATP for their function. A third protein component, scramblase, shows less substrate specificity and moves lipids against their concentration gradient in an ATP independent manner \[38\].

In addition to the differences between leaflets, membranes can also exhibit lateral asymmetry. Membrane domains can be defined as short-range ordered structures (0.001 and 1.0 \(\mu\)m in diameter), enriched in particular lipids and proteins. This lateral heterogeneity has been related to characteristic functional properties \[39\] and both protein- and lipid-based mechanisms for membrane domain formations have been described (reviewed in \[40\]). For some time, lateral heterogeneity was assumed to be an eukaryotic feature but lipid domains were later also found in bacteria \[41\].

3.2 The fluid mosaic model and beyond

The fluid mosaic model describes how proteins and lipids are organized in biological membranes. It depicts the membrane as a two-dimensional viscous lipid matrix, within which freely diffusing membrane proteins are embedded
in (see Figure 1.2 in Introduction). The model assumes that membrane proteins are interacting with lipids mainly due to hydrophobic forces [42]. However, more recent results have added new layers of complexity to this model. Specialized membrane domains and extramembranous structures can limit motion and lateral diffusion of membrane components and protein-lipid interactions can be essential for protein structure and function [40, 43].

Biological membranes are rich in protein - reported lipid:protein ratios vary depending on cell type but range between 18 - 80% protein by mass (reviewed in [44]). Membrane proteins can be crudely divided into two groups, the peripheral and integral membrane proteins. Peripheral membrane proteins attached to lipids or proteins in the membrane through non-covalent interactions. They dissociate from the membrane when treated with a solution with an elevated pH or high salt concentrations. Some peripheral membrane proteins are covalently linked to a lipid- or glycolipid anchor [45].

\[\text{Figure 3.4: Integral membrane proteins. a) A cartoon representation of an } \alpha \text{-helical bundle type of membrane protein (PDB 1A0T). b) A cartoon representation of a } \beta \text{-barrel membrane protein (PDB 1A0T).}\]

Integral membrane proteins have domains that cross the membrane (transmembrane domains) and can be extracted from the membrane with the help of detergents. They can be grouped into two structural classes: (I) \(\beta\)-barrel membrane proteins with amphipathic \(\beta\)-strands and (II) \(\alpha\)-helical membrane proteins with an \(\alpha\)-helices crossing the membrane (Figure 3.4). The vast majority of integral membrane proteins are \(\alpha\)-helical. It has been suggested that the amino-acid sequence of transmembrane domains would reflect on the properties of the bilayers in which they reside. Indeed, the amino acid composition do seem to vary based on the membrane where a particular helix is found em-
3.3 Lipid-protein interactions

So far the characteristics of a lipid bilayer have been discussed in relation to the physical and chemical properties of lipids. However, an interesting consideration is also how membrane proteins and lipids interact with each other. Several features of the membrane have been described to influence the function of membrane proteins: the thickness and phase of the lipid bilayer and the presence of a specific phospholipid [47].

3.3.1 Hydrophobic mismatch

An ideal transmembrane domain would span the hydrophobic core of the membrane, which would require an α-helix with 20 amino acids. However, both longer and shorter transmembrane segments exist and do not match the thickness of the lipid bilayer, a phenomenon termed - hydrophobic mismatch. If the transmembrane helix is too long, hydrophobic sequences are exposed to the aqueous environment. And if it is too short, hydrophilic amino acids have to reside within the membrane. Both situations are energetically unfavorable.

To minimize the free energy of the system, long transmembrane helices may either tilt or distort through coils to compensate for the mismatch, thus preventing the exposure of hydrophobic side chains to the aqueous environment [48]. Another alternative is that the lipid tails around the protein extend by acyl chain ordering [49]. When the transmembrane domain is shorter than the hydrophobic core of the bilayer, the tails of the surrounding lipids may compensate by chain disordering, resulting in compression [50–52]. However, it remains unclear whether lipids in living cells can change the thickness of the membrane [53]. Nevertheless, mismatch has been implicated in the functionality of several proteins [43, 54, 55] and may also be reflected in membrane protein sorting in eukaryotes [46, 56].

3.3.2 Membrane properties shaping protein function

Anionic phospholipids can regulate protein activity and structure [57–59] as well as influence targeting to the membrane [60]. Interaction with anionic phospholipids can be rather nonspecific. Unstructured regions enriched in arginine and lysine residues can form electrostatic interactions with a negatively charged lipid domain in the membrane [61]. Rhodopsin function, in turn, can be related to the requirement for a relatively unstable membrane environment, brought about by lipids with small head groups and bulky acyl chains [62, 63].
The mixing of bilayer prone and non-bilayer lipids can cause the tail region to over-pack while the head group region remains under-packed. This curvature stress results in a higher lateral pressure in the middle of the membrane, which in turn may affect membrane protein structure. A transmembrane protein may relieve this stress by adopting an hourglass shape, a feature found in a large number of membrane protein structures \[64\]. In addition, lipid composition of the membrane can direct the topology of at least some proteins in a non-specific manner \[27\] - a phenomenon discussed in more detail in chapter six.

3.3.3 Implications of specific lipid properties in membrane protein function

Some lipids interact specifically with a protein regulating its function. Since membrane proteins are generally solubilized in detergent solutions before crystallization, lipids are probably underrepresented in crystal structures. In addition, the lipid molecules present in the crystal may represent a few tightly bound lipid molecules and do therefore not represent typical protein-lipid interactions.

Regardless, examples where the structure and function of a membrane protein is linked to a specific interaction do exist \[47; 61\]. For instance, the full activation of Protein Kinase C is dependent on a specific interaction between phosphatidylserine \[65\]. Likewise, a conserved binding site for Phosphatidylinositol 4,5-bisphosphate (PIP2) has been identified in K\(^+\) channels and its functional importance was demonstrated in the Kv7.1 channel, where the coupling of the Kv7.1 channel voltage-sensing domain and the pore domain requires PIP2 \[57\]. A more general feature, such as the lipid head-group size, can also be important for protein function \[19\].
4. Protein machineries in the biogenesis of $\alpha$-helical membrane proteins

Membrane proteins, like the vast majority of proteins, are synthesized by ribosomes in the cytosol. While soluble cytosolic proteins begin to fold as they emerge from the ribosome, proteins destined for either secretion or membrane integration need to be directed to the ER (eukaryotes) or plasma membrane (prokaryotes and archaea). Furthermore, secreted proteins have to cross the membrane and membrane proteins need to transfer from the ribosome and the aqueous environment into the hydrophobic milieu of the lipid bilayer. Even though some $\alpha$-helical proteins are able to spontaneously insert into the membrane, the membrane integration in vivo is generally aided by protein machineries [66, 67].

Transport, translocation, and insertion of both membrane and secretory proteins is largely dependent on the Signal Recognition Particle (SRP) pathway [68]. An overview of this pathway is shown in 4.1. The major components are the cytosolic SRP, its membrane bound signal recognition particle receptor (SR) [69] and the core of the secretase (Sec) translocon - all universally conserved in the three domains of life [68]. The Sec translocon is also involved in post-translational translocation of proteins. In short, proteins destined for post-translational secretion or membrane integrations are bound by cytosolic chaperones and the driving force for translocation through Sec is provided by the cytosolic SecA protein and ATP hydrolysis [70, 71]. In yeast, an additional Sec62/63 complex is required for the post translational pathway [72]. Here, the translocating polypeptide moves through the pore by Brownian motion and the luminal protein BiP binds to the protein preventing backward movement [73].

4.1 The SRP-dependent pathway

Both membrane and secretory proteins contain an N-terminal signal sequence, crucial for transport to the target membrane [68, 74]. The signal sequence is
characterized by a hydrophobic core, surrounded by positively charged amino acids at its N-terminus and polar amino acids at its C-terminus [68]. Since the hydrophobic nature of the signal sequence core is the most important feature for SRP recognition [75], the first transmembrane domain of a membrane protein can also act as SRP substrate [69].

Figure 4.1: A cartoon over SRP-dependant co-translational secretion of soluble proteins (A) and insertion (B) of a membrane protein. From [74].

SRP is recruited to a ribosome-nascent chain complex (RNC) as soon as the signal sequence emerges from the ribosome exit tunnel [68], in eukaryotes, this leads to stalling of the ribosome. Formation of the RNC-SRP complex is followed by its association with the ER (eukaryotes) or plasma membrane (prokaryotes), mediated by the GTP dependent interaction between SRP and its receptor. Once at the membrane, the RNC complex is transferred on to the translocon, followed by GTP hydrolysis in SRP and SR resulting in their dissociation [69; 74]. These events are presented schematically in Figure 4.1.

The RNC and translocon are associated with each other such, that the ribosomal exit tunnel is aligned with the translocon pore and the polypeptide synthesized by the ribosome can be fed into the translocon [76]. The translocon possesses a dual nature, allowing both a passage across the membrane and into the membrane. Transmembrane domains are recognized and shunted sideways into the membrane bilayer, while polar domains either remain in the cytoplasm or translocate to the periplasm/ER lumen. Since the translocon itself is passive, a driving force is required. In the co-translational mode, this force
is provided by GTP hydrolysis during polypeptide synthesis \[35\]. Knowledge of the translocon as well as the properties of amino acids are both needed to understand how transmembrane regions are recognized \[35\]. The structure of the translocon is described below while the fifth chapter will focus on the molecular code for translocon mediated recognition of transmembrane segments.

Figure 4.2: The SecYEG translocon (PDB 1RH5) in cartoon representation. a) A top view from the periplasmic side. SecY is shown in green and purple, to highlight the clamp-shell like shape. SecE is shown in cyan and Sec B in grey. b) Side view from behind. The hinge-region between TMH5 and TMH6 connecting the two halves is prominent. c) The lateral gate (purple and green) and the plug-domain (blue) are highlighted.
4.2 The Sec translocon

The Sec61αβγ-translocon is a heterotrimer and facilitates the translocation of secretory proteins across, and insertion of membrane proteins into the ER membrane (shown in Figure 4.2). The functional homolog in bacteria is SecYEG and in archaea SecYEβ, both which reside in the plasma membrane. Sec61α and γ share high sequence similarity with SecY respectively SecE. A crystal structure of the *Methanococcus jannaschii* SecYEβ revealed fascinating details about the channel structure [77]. The model for translocon function in membrane protein biogenesis is based on both structural data as well as many biochemical studies (reviewed in [78, 79]).

The major component of the translocon is the α-subunit, a membrane protein with 10 transmembrane helices. When viewed from the side, it has an hourglass-shape with a narrow pore composed of hydrophobic amino acids [77], through which secretory proteins and the loops of membrane proteins cross the membrane [80]. The exact dimensions of the pore have remained controversial and the dimensions may adjust to the nascent protein chain [81]. From top view, Secα has a clam-shell like shape with two halves formed by TMH1-5 and TMH6-10 (Figure 4.2 a and b). The loop between TMH5 and 6 functions as a hinge, allowing the opposite side of the protein to open up towards the lipid bilayer [77].

The lateral gate is mainly formed by TMH2 on one side and by TMH7 on the other [77] (Figure 4.2c). The opening allows exposure of a translocating polypeptide to the hydrophobic environment of the lipid bilayer, with the possibility of said polypeptide to partition into the membrane [82]. The exact mechanism is unclear, but experimental and molecular dynamics simulation data suggests that the binding of a prospective transmembrane domain to the translocon could stabilize the open conformation of the lateral gate [83, 84]. The membrane barrier is maintained in part by the narrow pore ring and its interactions with a short helix on the non-cytosolic side, namely TM2a [85], which forms a plug (Figure 4.2c) blocking the passage of small molecules [77]. In eukaryotes, a lumenal protein BiP has also been implicated in preventing leakage [86].

The mammalian Secγ and the prokaryotic SecE are single-spanning membrane proteins in most species. Like Sec61α/SecY, they are essential. Studies in yeast suggest that the γ-subunit is important for translocon stability [87] and it is found associated to Sec61α on the back/hinge side in the crystal structure [77]. The third, non-essential subunit (Sec61β/SecG) is usually a single-spanning protein in archaea and eukaryotes but has two transmembrane domains in bacteria.
4.3 Associated proteins

The active translocon has been observed to be a multimer of Sec heterotrimerers and associated with other proteins. However, the actual number of subunits and their identities have remained controversial [88–90]. Cryo-EM reconstructions of native ribosome-translocon complexes suggested a complex with two dimers of the Sec61 heterotrimer and two tetrameric translocon-associated protein complexes (TRAP) [88], while cross-linking studies indicate that a single Sec61 heterotrimer is responsible for protein translocation [90]. Perhaps the controversy is in part due to versatility of the Sec-translocon, where the number and identify of interacting proteins is at least occasionally dependent on its substrate [91–93].

The co-translational insertion of many membrane proteins and bacteria seem to require YidC, a membrane protein with six transmembrane domains [94; 95]. In eukaryotes, the monomeric translocation-associated membrane protein (TRAM) may fulfill a similar function [96; 97]. Both TRAP and TRAM have been cross-linking to nascent peptide as they emerge from the translocon, possibly influencing integration into the bilayer [96; 98] or the topology [99].

In the ER, two modifying proteins are also found in the close vicinity of the translocon namely the signal peptidase complex (SPC) [100] and Oligosaccharyl transferase (OST) [101]. The SPC cleaves signal sequences of some membrane proteins and from secreted proteins, releasing them to the exoplasm [102]. OST is a membrane protein complex in the ER membrane with its active side on the lumenal side. It recognizes a consensus sequence of Asn-X-Ser/Thr and catalyzes the N-linked glycosylation of the protein sequence; attaching a sugar moiety to the asparagine [35] in a co-translational manner [103].
5. Properties of amino acids in membrane protein structure

The structure of a protein can be described in terms of secondary-, tertiary-, and quaternary structure. The two most common secondary structures are $\alpha$-helices and $\beta$-sheets [104]. The tertiary structure describes the spatial arrangement and interactions between the secondary structure elements in one polypeptide chain. Interactions between at least two folded chains result in a protein complex described by the quaternary structure [24].

Topology is often used to describe $\alpha$-helical membrane proteins. In this context it refers to the number and orientation of transmembrane helices as well as the location of the N- and C-termini of a protein. Membrane proteins tend to follow a predictable pattern of topological organization; anti-parallel transmembrane helices cross the membrane from one side to the other with hydrophilic loops alternating between cytoplasmic and non-cytoplasmic location. By now many properties directing the topology of a nascent polypeptide chain have been reported (reviewed in [105,107]); Transmembrane helices are recognized through sufficiently long and hydrophobic sequences of amino acids and the orientation is determined mainly by the distribution of positively charged amino acids according to the positive-inside rule [35].

5.1 Defining amino acid hydrophobicity

Just as there is an energetic cost associated with introducing a non-polar molecule (or amino-acid side chain) into an aqueous environment, there is an energetic cost from introducing a polar molecule or side chain into the nonpolar membrane interior, due to the loss of polar or hydrogen bonding with water. The polar peptide back-bone opposes membrane insertion, even though the free energy cost can be reduced by secondary structure formation [108,110]. Recent experiments with non-proteinogenic amino acids have demonstrated that the hydrophobic surface area of an amino-acid side chain is directly proportional to membrane insertion [111]. In other words, the hydrophobic effect will promote partitioning of a peptide when said peptide contains a certain number of hydrophobic amino acid side chains.
In the cell, transmembrane helices are recognized by the translocon based on the average hydrophobicity of a stretch of amino acids. Several experimental and computational studies of artificial systems, as well as in vitro ([108, 109, 112]), in vivo ([112–114]) and statistical ([115, 116]) methods have been used to assess amino acid hydrophobicity. The biological hydrophobicity scale, also known as the Hessa scale, defines the individual contributions of amino-acid side chains in a position specific manner ([112; 117]) (Figure 5.1 a). For the work presented in this thesis, the biological hydrophobicity scale and tools for predicting the change in free energy upon insertion based on this scale were used.

Correlation between the different hydrophobicity scales is in general good. Compared to other experiential scales, polar and charged side chains are tolerated unexpectedly well in the membrane ([112; 117]). Concurrently, the hydrophobicity of proline varies significantly. It is hydrophobic in the GES and Wimley-White scale ([118, 119]), but rather hydrophilic according to the Hessa scale. The later being understandable, given its helix-breaking nature ([120]. Overall the differences between hydrophobicity scales may be attributed to the high abundance of proteins in biological membranes, compared to the uniformly hydrophobic environment composed of membrane mimics such as organic solvents ([119].

**Figure 5.1:** Translocon mediated recognition of transmembrane α-helices. a) The Hessa scale for each amino acid when placed in a central position of a transmembrane helix. b) The probability of membrane integration plotted against the number of leucines in a transmembrane helix conform to Boltzmann distribution. Adapted from [35].
5.2 The properties of amino acids in membrane proteins

To establish the Hessa scale, a series of artificial, potential transmembrane segments were designed and presented to the translocon in microsomal membranes. The in vitro expression system (described in detail in chapter seven) allowed a quantitative assessment of membrane insertion efficiency of the test segments \[112\]. The probability of insertion conformed to a Boltzmann distribution, suggesting that the translocon-mediated insertion is an equilibrium process (Figure 5.1b). Hence, the apparent free energy of insertion (\(\Delta G_{\text{app}}\)) can be calculated and used to express the potential for membrane insertion of a given polypeptide \[35; 112\]. By systematically varying the amino acid composition of these test segments, the molecular code for translocon mediated membrane insertion was deciphered \[112; 117; 121\](Figure 5.1a). Similar studies were carried out in the \(E.\ coli\) inner-membrane \[114\], baby hamster kidney cells \[112\] and yeast \[113\].

5.2.1 Hydrophobic and helix-forming amino acids promote membrane insertion

A transmembrane helix is exposed to a varying milieu in the lipid bilayer, also reflected in the statistical difference in amino acid distribution within it as depicted in Figure 5.2 \[64; 115\]. Which amino acids are present in transmembrane helices is not only determined by hydrophobicity and bulkiness but also in their abilities to form interactions with the protein itself and the environment it resides in.

Isoleucine, leucine, phenylalanine and valine promote membrane integration and dominate in the central region of a transmembrane helix (Figure 5.2 a). Cystein, methionine and alanine have a \(\Delta G_{\text{app}} \approx 0\) kcal/mol, placing them at the threshold between those amino acids that promote membrane integration and those that do not (Figure 5.1a). However, alanines are good \(\alpha\)-helix-formers and often found in transmembrane helices \[115\] \[122-125\]. Polar and charged residues are rare in the membrane core, but some of them show biased distribution in membrane proteins (Figure 5.2 b). These amino acids and their role in membrane proteins are discussed below.

5.2.2 Charged amino-acid side chains in transmembrane helices

The increased number of crystal structures have resulted in many findings of charged residues and irregular secondary structure within the hydrophobic core \[126\], despite polar residues having high \(\Delta G_{\text{app}}\) values \[112; 117\].

Long, aliphatic side chains allow charged amino acids to orient their side chains towards the interfacial region and to interact with lipid head groups.
Figure 5.2: Statistical distribution of amino acids in a transmembrane helix. a) An illustration of the distribution of various amino acids at different positions in a transmembrane helix (by courtesy of Linnea Hedin Barkå). b) The position specific contribution towards $\Delta G_{\text{app}}$ plotted against the position within a 19-residue segment is shown in blue. Amino acids with interesting traits discussed in sections below are shown. The red line indicates position-specific statistical distributions calculated from three-dimensional structures of membrane proteins. Adapted from [117].

This so-called snorkeling may explain why polar groups are better tolerated in biological membranes than expected [130]. Simulations suggest, that snorkeling also allows the polar side chains to create polar microenvironments for themselves by pulling water into the membrane core [131]. In addition, intramembrane salt bridges have been found in some membrane proteins, with both structural and functional importance [132–134].

Positively charged residues close to transmembrane helices are strong topogenic signals, which will be discussed in detail under section 5.3. In contrast, acidic residues are much less potent topology determinants and show no
statistical preference for loops on either side of the membrane \[122\] \[135\]. However, they have been reported to influence topology under special conditions, such as when negative charges are present in high numbers \[136\], in close proximity of marginally hydrophobic transmembrane helices \[137\] or when present within seven flanking residues from the end of a transmembrane helix \[138\].

### 5.2.3 Aromatic amino acids are unequally distributed in transmembrane helices

Tryptophan and tyrosine are enriched near the ends of the helices, often referred to as the aromatic belt (Figure 5.2) \[64\] \[122\] \[139\]. They are believed to interact favorably with the lipid head groups and have been shown to anchor and stabilize tilt angles of transmembrane helices relative to the bilayer \[140\] \[141\]. Phenylalanine, on the other hand, is entirely hydrophobic and thus more abundant in the central core region of transmembrane helices (Figure 5.2b) \[64\] \[141\].

### 5.2.4 Proline and glycine in transmembrane segments

Proline is a unique amino acid as its amine nitrogen is part of a ring structure bound to two alkyl groups. Its rigid structure disrupts an α-helix \[120\] introducing either a kink in a transmembrane domain or promoting the formation of helical hairpins (two closely spaced TMHs with a tight turn) in sufficiently long hydrophobic segments. Proline induced kinks may be critical for the proper structural stability and/or function of membrane proteins \[142–144\] and the transmembrane domains of membrane proteins contain more prolines than α-helices in globular proteins \[145\].

Amino-acid sequences rich in prolines and glycines tend to form coils, that is, regions lacking regular secondary structure. The presence of coils allows a higher degree of structural flexibility, creating swivels and hinges \[144\] \[146\] as well as reentrant regions \[64\] \[126\] \[147\]. They are especially common in channels and transporters and are often required for function \[147\].

Glycine is also involved in helix-helix interactions. The GxxxG motif is fundamental in helix-helix associations \[148–151\]. Here, the two small glycine residues are separated by one turn creating a groove on the helix surface. This groove serves as a contact surface for another helix with the same motif. The GxxxG or variations of it (G/A/S)xxxGxxxG and GxxxGxx(G/S/T) occur in more than 10% of all known membrane protein structures \[148\]. An antiparallel version of the motif has been suggested to be even more common (16%) in helix packing in membrane proteins \[152\].
5.3 Positive inside rule in establishing topology

Studies on membrane proteins with known topology revealed a bias for positive charged residues in cytoplasmic loops. This uneven occurrence in amino acid distribution is generally referred to as “the positive-inside rule”, and has been shown to hold for most organisms [4, 122, 153]. The presence of positive charges influence both the insertion and orientation of a transmembrane helix. For instance, a single Arg or Lys residue placed downstream of a transmembrane segment in $C_{in}$ orientation can lower the apparent free energy of insertion by $\approx 0.5$ kcal/mol. The effect is additive and dependent on the distance from the positive charge to the transmembrane segment [154]. However, the effect may also contribute globally, as a single positive residue placed at the very C-terminus of the dual topology protein EmrE, was able to flip the topology of the entire protein [155].

The exact mechanisms behind the positive inside rule are not fully understood. Since the cellular conditions for membrane protein insertion vary within the three domains of life, the effect of positive charges may even differ between Archaea, Bacteria and Eukarya. One explanation to why the retaining effect of positive charges is more pronounced in E. coli than in microsomes [156] is the membrane potential. The electrochemical potential across the bacterial inner-membrane is stronger than that of the ER membrane. However, the positive-inside rule does apply both in the ER and in the bacterium Sulfolobus acidocaldarius, with a reversed membrane potential [157]. Likewise, if the membrane potential was responsible, negatively charged residues would be expected to have topogenic effect. This does not seem to be the case [64, 122, 135, 158].

A more likely suggestion is that the anionic phospholipids prevent membrane passage of positive charges [159]. Basically, the negative headgroup of anionic phospholipids form electrostatic interactions to positively charged residues in protein domains retaining the loop in the cytoplasm [61, 160, 161]. As anionic phospholipids are present in all membranes [61], this might explain the ubiquity of the positive inside rule.

Specific interactions with the translocon have also been reported to contribute to the orientation of the signal sequence [162, 163]. Mutagenesis studies on yeast Sec61p identified three charged residues, which influenced the topology of some membrane proteins [162]. Further indications of the role of translocon in membrane topogenesis, came from studies where substitutions in the lateral gate altered the topology of membrane proteins [163].
5.4 The two stage model and topology

The knowledge of properties of transmembrane domains, topology signals and folding of a membrane protein are often combined into the two-stage model (Figure 5.3). In this model, topology is established during the first step, as helices are inserted individually into the membrane. The second stage consists of interactions between the membrane embedded helices, folding into the final tertiary structure [164-166]. Although this model applies to many proteins [165], others show more complex behavior [126,167-171], which is the focus of chapter five.

![Figure 5.3: The two-stage model for membrane protein folding. In the first stage, membrane proteins are inserted into the membrane. The transmembrane helices and their topology is decided at this step. In the second stage, the membrane protein folds. This includes the formation of tertiary structures, incorporation of prosthetic groups and re-positioning of transmembrane helices. Adapted from [166].](image)

5.4.1 The first stage - establishing topology

Topology is in general established during co-translational membrane integration [172]. Hydrophobicity and the positive-inside rule appear to be the main determinants but other more subtle features may contribute as well [168,173-177].

5.4.2 The second stage - membrane protein folding

Unlike soluble proteins, membrane folding of membrane proteins occurs in an environment that is different in its character (the aqueous extra- and intracellular environment and the hydrophobic core of the membrane) [166]. To add to that, the orientation of transmembrane domains is considered at least relatively fixed. One might think that the loops would be important in keeping the membrane protein together but transmembrane helices are known to assemble into a functional protein without the presence of loops [178,180]. This suggests,
that the native fold of membrane proteins is mostly stabilized by interactions between transmembrane domains.

VdW - (specifically dispersion) - forces have been suggested to be major driving force for membrane protein folding [139, 181]. The amino-acid residues in transmembrane regions are in general more buried than residues in soluble proteins or extramembrane regions [166, 182], resulting in higher numbers of vdw interactions even if not necessarily stronger ones.

The backbones of transmembrane helices can also be involved in “non-conventional” αC-H...O hydrogen bonds, and have been suggested to contribute to helix-helix interactions [183, 184]. While their influence on stability remains controversial [185, 186], they might be important for helices at close distances and be suitable for maintaining stability but also provide structural flexibility [187, 188].

Occasional salt bridges may also contribute towards membrane protein structure and stability. Salt bridges are defined as electrostatic interactions between ions of opposite charge. Due to the low di-electric constant, such interactions are strong [132, 134]. Further, in an oxidative environment, two cysteines can form a disulfide bond. In vivo this occurs in the ER of eukaryotes and in the periplasm of gram negative bacteria and is catalyzed by enzymes [189, 190].
6. Non-sequential membrane integration of $\alpha$-helical membrane proteins

The textbook version of a membrane protein is an $\alpha$-helical bundle, where each of the hydrophobic transmembrane helices cross the membrane in more or less perpendicular orientations [191]. The topology is assumed to form sequentially and to be relatively fixed once established (See Figure 6.1a). However, the increasing number of solved membrane protein structures show far more variation in the structural elements of membrane proteins [192]. Transmembrane segments can vary significantly in length [64, 129, 173, 193, 194], adopt strongly tilted conformations (Paper III), have kinks, polar groups and non-helical regions in the middle of the membranes [195] as well as re-entrant regions - helices that span only a part of the membrane before looping back.

![Figure 6.1](image-url)
Transmembrane segments are not necessarily recognized as such by the translocon, which results in non-sequential membrane insertion. As a consequence, some proteins have been reported to initially insert in an intermediate topology (See Figure 6.1 b). Also, the initially inserted transmembrane regions may differ from the membrane embedded regions in the final structure. In both of these cases, the final topology and functional structure is achieved through later repositioning events. Concurrently, an increasing number of results suggest that topology may not be as fixed as previously thought. Membrane proteins can undergo dramatic re-organizations, including inversion of transmembrane domains and translocation of extracellular loops, in response to changed lipid composition in vitro and in vivo. Since this can occur without the involvement of a cellular machinery, the activation energy must be low for re-assembly. 18 196 197. This is an important observation, not least considering dual topology membrane proteins. Dual topology membrane proteins adopt two opposite topologies 198 199 and a single positively charged residue at the very C-terminus can be sufficient to convert an established topology to an opposite one 155.

6.1 Marginally hydrophobic helices

A marginally hydrophobic transmembrane helix (mTMH) can be defined as a transmembrane domain unable to insert into the membrane by itself. A surprisingly large fraction (>30 %) of transmembrane helices in multi-spanning proteins of known three-dimensional structure have a high $\Delta G_{\text{pred}}$ and may thus be dependent on extrinsic sequence characteristics for membrane insertion 117. While at least some of these helices can insert in the membrane surprisingly well, those with even higher $\Delta G_{\text{pred}}$ above tend to require other parts of the same protein for efficient membrane integration 168 170 171 191 200 201.

Even though polar residues in transmembrane regions hamper membrane integration, they are more conserved than other residues in transmembrane segments 195. This is in part explained by their functional involvement and their tendency of being buried within the structure 195 202. Interestingly, polar residues are more common in polytopic membrane proteins with many transmembrane domains. In addition, the first 195 and last transmembrane helices tend to be more hydrophobic than the helices in between (Paper I).

Membrane integration of mTMHs depend on sequence features extrinsic to the hydrophobic segment itself. Some features have been identified (illustrated in Figure 6.2) and include charged amino-acid residues flanking the hydrophobic segment 117 154 168, interactions between polar residues in adjacent TMHs 203 205, neighboring helices with strong orientational preference
Figure 6.2: Local sequence context can help insert a mTMH. a) Positive charges in cytoplasmic loops can contribute towards the free energy for membrane integration and promote membrane insertion of a mTMH. b) Orientational preference, induced by e.g. positive charges, can pull a mTMH into the membrane. c) Specific interactions between a mTMH and an other transmembrane helix can allow insertion of the helix-pair. d) Repositioning can allow a segment of lower hydrophobicity to enter the membrane, even though a more hydrophobic part is initially recognized by the translocon. From Paper II.

[206] and repositioning of TMHs relative to the membrane during folding and oligomerization [126]. However, the mechanisms by which these events take place and how frequent they are is still unclear. This section is devoted to describing both the function and problems polar groups in transmembrane helices can bring by, as well as how they can be successfully integrated into the membrane despite high energetic cost.

6.1.1 Consequences of mTMH in AQP1 folding

Aquaporin 1 (AQP1) is part of the ubiquitous family of water channels with six transmembrane helices and two re-entrant loops with a central water conducting pore [207]. When AQP1 topology was studied in Xenopus oocytes, it was shown to initially insert as a four-helix intermediate before folding into its final structure at a later stage [208; 209] (see Figure 6.3). These results were regarded with some skepticism based on contradicting results from experiments in mammalian cells [210]. However, this apparent controversy has been solved by the observation that the intermediate is less stable in mammalian cells [209].

Biogenesis of the close homolog Aquaporin 4 (AQP4), occurs more conventionally, with sequential and co-translational insertion of each transmem-
Figure 6.3: Topological rearrangements in AQP1. AQP1 is initially inserted into the membrane as a four-helix intermediate (top panel). The gray shading roughly depicts the hydrophobicity of each segment to highlight the low hydrophobicity of mTMH2, much due to Asn49 and Lys51. For AQP1 to obtain its 6 transmembrane domain topology (bottom panel), the reorientation of TMH3 is required. A proposed “R1-H3 shift” is depicted in the middle panel. According to this model, TMH3 can spontaneously shift out of the membrane core, initiating topological reorganization and folding of AQP1. Adapted from Paper IV.

brane segment. The sequence features underlying the different behavior of AQP1 have attracted a lot of attention. In essence, the second transmembrane helix is not sufficiently hydrophobic to integrate into the membrane [168, 211, 212], which in turn results in altered behavior of the rest of the protein [212]. For one, TMH3 will be inverted in its opposite orientation while TMH4 is unable to integrate into the membrane, due to a number of positively charged residues at its C-terminus. Transition from this intermediate to the final topology requires the 180° rotation of TMH3 as well as the translocation of the loops between helices 3-4 and 4-5 (Figure 6.3). An occurrence requiring the presence of the transmembrane helices four, five and six [209, 211]. These observations imply interesting and novel features in membrane proteins: a dilemma regarding features required for function and correct folding as well as the unexpected flexibility in membrane protein structures.
AQP1 demonstrates clearly the consequences of a mTMH in a membrane protein. The residues rendering mTMH2 unable to integrate to the membrane are functionally important [212]. At the same time, correct orientation is a prerequisite for adequate exposure of extramembranous domains on the functionally relevant side of the membrane. How does the cell solve the dilemma of keeping the unfavorable residues within the mTMH yet obtaining functional protein? After all, not only is the production of a non-functional protein costly [213–215], it also poses a burden for cellular quality control along with protein degradation systems. Clearly there must exist a mechanism to ensure proper membrane integration - some sequence features and cellular machineries have indeed been implicated and are described below.

AQP1 folding also proposes significant plasticity in membrane protein topogenesis and folding. Similar behavior has been observed for several polytopic membrane proteins [171; 200; 201; 216]. The ability to reorient transmembrane segments may not be restricted to membrane protein folding. E. coli SecG has been reported to undergo reversible topology inversion as part of its function [217]. Even more so, at least three membrane proteins (LacY, PheP and GabP) are known to undergo dramatic reorganization upon changes in membrane lipid composition [18; 218; 219].

The dramatic differences in the topogenesis of AQP1 and AQP4 are dependent on surprisingly few differences in their primary sequences. Two residues (Asn49 and Lys51) in AQP1 mTMH2 are responsible for its hydrophilic nature. When exchanged to the corresponding residues in AQP4 (a methionine and leucine), the transmembrane domains of AQP1 are inserted in a sequential manner [212]. A previously uncharacterized difference lies in the nature of the region just before TMH3, which contains the helical section of reentrant region 1, the loop between the re-entrant region and TMH3, along with the N-terminal part of TMH3. In AQP4, this segment is hydrophilic (ΔGpred 4 kcal/mol) while the corresponding region in AQP1 is hydrophobic (ΔGpred 0 kcal/mol). We suggest that this difference may account for the flexibility in AQP1 topogenesis since the third transmembrane helix of AQP1 may shift out of the membrane core simultaneously bringing in the preceding “R1-H3 loop” into the membrane (Paper IV).

6.1.2 Cost of polar groups in the membrane

The hydrocarbon core of biological membranes has been considered to be a “forbidden zone” for charged amino acids [220]. However, the membrane structure derived from X-ray and neutron scattering data depicts a rather dynamic structure. Only the very center of the membrane is entirely hydrophobic, the rest is a mixture of polar lipid head-groups and water molecules [36].
Consequently, the membrane may partially permit the presence and passage of polar groups \([36]\).

In regards to mTMHs and the ability of extramembraneous regions to cross the membrane it is of interest to learn how unfavorable polar groups in the membrane really are. Arginine is often thought of as the most hydrophilic of amino acids, yet it is frequent in transmembrane helices \([220]\). Continuum electrostatic models predict a free energy penalty of tens of kcal/mol for the presence of a charged group in the membrane \([221]\), molecular dynamics estimations suggest a barrier around 17 kcal/mol \([222]\) and experimental scales report values ranging between 12 to 1.8 kcal/mol \([118]\). However, direct comparison of different hydrophobicity scales is not straightforward as they are often normalized in different ways \([119]\).

According to the Hessa scale, the cost of an arginine in the middle of an hydrophobic helix is only \(\approx 2.5 \text{ kcal/mol} \ [112; 223]\) and strongly position specific \([223]\). Several mechanisms seem to decrease the free energy penalty for a charged group in the membrane. Charged residues with long side chains can snorkle toward the membrane interface and polar groups can draw lipid head-groups and water deep into the core \([195; 222; 224]\). Additional, arginine (and other charged residues) may partition into the membrane as a sufficient number of hydrophobic amino acids can to overcome the cost \([220]\).

### 6.2 Sequence features implicated in the insertion of mTMHs

#### 6.2.1 Positive-inside rule

Flanking loops and neighboring helices can be important for the membrane integration of mTMHs \([168]\). In accordance to the positive-inside rule, arginines and lysines in cytoplasmic loops contribute towards the apparent free energy of membrane insertion by \(\approx 0.5 \text{ kcal/mol}\) of a transmembrane helix, see Figure 6.2a \([154]\). However, cytosolic loops with positively charged residues do not always improve insertion of a mTMH \([168]\). Instead, the successful membrane integration of a mTMHs can depend on a neighboring transmembrane segment; the positive inside rule is stronger when the positively charged residues are followed by a transmembrane helix (Paper I).

Characteristics in one transmembrane helix can influence the insertion propensity of another. This phenomenon was first observed for the human band 3 protein, where the strong orientational preference of a transmembrane helix resulted in membrane integration of an upstream region \([171]\). Previous studies have implicated both positive charges and hydrophobicity in the likelihood of a transmembrane helix adopting a certain topology \([106; 174; 176; 194]\). A recent, systematic study showed how orientational preference of a neighboring
helix could both increase and decrease the insertion of a mTMH [206] (see Figure 6.2b).

What then, gives a transmembrane helix an orientational preference - apart from cytoplasmic positively charged amino-acids residues? Long and hydrophobic helices may favor \( N_{\text{out}} - C_{\text{in}} \) orientation, while short and less hydrophobic helices instead adopt an \( N_{\text{in}} - C_{\text{out}} \) topology [173-175]. A hydrophobic transmembrane helix may spend a shorter time inside the translocon, thus not having enough time to reorient. A less hydrophobic transmembrane helix, on the other hand, may not move out from the translocon as fast allowing re-orientation. This “hydrophobicity signal” can however be overriden by, positively charged residues at the N-terminus and vice versa [176; 177]. Since these observation strongly rely on artificial model proteins it remains unclear whether this trend is prevalent in native proteins. In addition, several studies suggest that the first transmembrane helix may not be sufficient in determining the topology of the entire protein [225;227].

6.2.2 Specific interactions between polar groups

Specific interaction within or between helices can improve their propensity to insert into the membrane (see Figure 6.2c). As discussed before, polar or charge groups hamper insertion of transmembrane domains. However, helices with charged or polar residues can interact with each other before entering the lipid bilayer, effectively “hiding” the energetically unfavorable groups in inter-helical hydrogen bonds [204;205].

6.2.3 Repositioning in the membrane

One turn in an ideal \( \alpha \)-helix employs \( \approx 3.6 \) amino acids per turn with one turn having the height of 5.4 Å. To span the hydrophobic core of a lipid layer approximately 20 residues are needed to form a sufficiently long helix. However, the length of transmembrane helices vary between 12 and 40 residues promoting reorganization of both proteins and lipids [64;129;173;193;194]. One of such adaptations is the tilting of longer helices relative to the membrane normal [48].

While tilting of transmembrane domains can occur as a consequence of hydrophobic mismatch, it is also possibly that tilting is induced during membrane protein folding due to packing interactions. It may also enable polar and charged side chains to become transmembrane, allowing them to get buried in the protein rather than being exposed to lipids (see Figure 6.2d, Figure 6.4b and Paper III) [228]. The glutamate transporter homologue from Pyrococcus horikoshii serves as an example. Its three-dimensional structure is complicated
with both mTMHs and transmembrane helices disrupted by coils. In a previous study, the primary sequence was aligned with known secondary structures and a theoretical hydrophobicity profile. Overlapping regions with more hydrophobic character were identified and at least one those segments was shown to be more prone insert into the membrane [126]. Similar behavior has been reported in the human Band 3 protein and the KAT1 voltage dependent K$^+$-channel [201, 229, 230] where polar interactions within the protein enables membrane integration of transmembrane segments [231].

Reentrant regions are segments that penetrate the lipid bilayer without traversing it, having their N- and C-terminus on the same side. They are enriched in alanine, glycine and proline residues and are mostly found in water and ion channels [147, 232]. Their integration to the membrane can occur both co- and post-translationally [231, 233].

6.3 Can translocon influence hydrophobicity threshold or topology?

The translocon itself may influence the insertion and topology of transmembrane helices. Conserved residues both in the translocon pore ring and the later
gate can influence the insertion of signal sequences [234, 235]. Additionally, the arginine to glutamate substitution at the plug domain of yeast translocon weakens the positive-inside rule [163]. Substitution-sensitive residues have also been identified in the lateral gate, where they could both increase and reduce the threshold hydrophobicity for transmembrane segment. However, this impact is strong only for single-spanning membrane proteins and on the first transmembrane domain of a polytopic membrane protein [236].

Re-orientation of transmembrane helices has been proposed to depend on cellular machineries. Based on previous studies, the ribosome exit tunnel [237] was ruled out as a compartment where an α-helix might change orientation. However, more recently a “folding vestibule” was identified at the end of the ribosomal exit tunnel and may allow reorientation [238]. Whether the translocon is big enough to allow a polypeptide invert has also remained controversial [70; 77; 81; 88; 239]. Since transmembrane helices can stay in close proximity to the translocon [172; 240], it has been proposed that reorientation occurs adjacent to the translocon complex even though not in it [106]. Then again, the high protein content [88–90; 100; 101] alone might favor insertion and re-orientation of membrane protein domains [16; 172].

The time point and location where transmembrane helices can form interactions with each other have also gained a lot of attention. Helical hairpins might form as early as in the ribosome “folding vestibule” [238] and might allow membrane insertion of a mTMH [204]. Can the translocon accommodate more than one polypeptide chain? Experimental studies suggest it might [205; 241]. In fact, many membrane proteins do linger within or close by the channel [172; 242]. The translocon associated proteins TRAM and TRAP have both been cross-linked to nascent polypeptides emerging from the translocon affecting both membrane insertion [96, 98] and topology [99] of their substrates.

The size of the translocon protein conducting channel is important considering the popular model for membrane protein insertion [35, 77, 112]. While the cytoplasmic cavity of *M. jannaschii* SecY has a diameter of 20 to 25 Å, the narrowest part of the channel ≈ 5 – 8 Å wide. In order to accommodate an α-helix, the channel would have to expand [77]. Interestingly, the central pore appears wider (≈ 10 Å) in the *Thermotoga maritima* SecY structure, where SecY interacts with SecA and SecG [70]. The channel size has also been assessed experimentally. In one study, SecYEG was challenged with rigid molecules of varying size fused to known Sec-substrates. SecYEG allowed the unrestricted passage of constructs up to 22-24 Å [239], which is larger than seen in crystal structures or estimated by molecular dynamics simulations [243]. Given the conflicting results and the huge thermodynamic driving force for membrane protein insertion [18, 66, 139] a rather interest-
ing model has been proposed. Perhaps the transmembrane helices never fully enter the translocon channel but interact with the cytoplasmic membrane interface and slide into the membrane utilizing the translocon lateral gate [244]. This would sit well with the observations of neighboring helices enabling the membrane integration of mTMHs.

6.4 The membrane environment

![LacY folding dependent on lipid composition](image)

**Figure 6.5:** LacY folding dependent on lipid composition a) The topology of a native LacY. b) Illustration of LacY topology in the absence of phosphatidylethanolamine (PE). If PE is added to the membrane, TMH2 forms a re-entrant region allowing TMH3, 4, 5 and 6 to reorient. TMH7 can now enter the membrane. Adapted from [31].

The topology of a membrane protein can be exceptionally dynamic and can be altered by changing the environment (Figure 6.5). When the zwitterionic membrane lipid PE is depleted from *E. coli*, the topology of LacY exhibits a partial and reversible topological inversion [197][218] (see Figure 6.5b). This re-assembly is independent on any cellular machineries [18]. While the be-
behavior of LacY is remarkable, it should not be considered an anomaly. Lipid-dependent topological reorganizations have been shown to occur for other proteins as well \[159, 218, 219\] and membrane lipids are important for both insertion and stability of many membrane proteins (reviewed in \[30, 31\]).

Most eukaryotic membrane proteins are initially inserted into the endoplasmic reticulum and subsequently targeted to the membranes of other organelles. For example, the major lipids of the plasma membrane are phosphatidylcholine, sphingomyelin but contains cholesterol and phosphatidylinositol as well. In contrast, the ER membrane is mainly composed of phosphatidylcholine, PE, and phosphatidylinositol with only small amounts of sphingolipids and cholesterol \[23, 245\]. The lipid composition in the different membranes is reflected in the composition of transmembrane segments \[46\] and perhaps also topology \[208-210\].
7. Methodology - studying membrane proteins

The function of a protein depends on its structure and thus, three-dimensional structures of proteins are of utmost interest. The most common techniques to obtain three-dimensional structures are X-ray crystallography [246, 247] and nuclear magnetic resonance (NMR) [248]. Both of them provide resolution down to the atomic level (about 1.5 Å or lower). Electron microscopy (EM) gives images of lower resolution, revealing the general shape of a protein. However, recent advances in both cryo-electro microscopy and single-particle reconstitution allow the determination of macromolecular structures at near-atomic resolution [249, 250].

7.1 Biochemical determination of topology

In the absence of a structure, topology is a vital part in understanding membrane protein function. Luckily, the topology of a membrane protein can be deduced utilizing so-called topology markers; modifications that occur in only one of the compartments separated by the membrane, such as glycosylation [251] or cysteine labeling [252]. Acceptor sites for a particular modifications are introduced at various positions in a protein and depending on whether a modification takes place or not, will allow the deciphering of the topology. Another method uses reporter proteins that are active only on one side of the membrane. However, this method can usually only give information on the location of the protein termini.

7.1.1 in vitro protein expression in the presence of microsomes

Cell integrity is not required for protein synthesis. Translation can occur in crude lysates containing the ribosomes, initiation, exchange and termination factors, energy source, tRNAs, amino acids and a DNA (coupled system) or mRNA template (uncoupled). In the coupled transcription and translation system, a mRNA is transcribed in situ, usually resulting in higher protein yields and faster assay time. In addition to the components mentioned, coupled systems need to be supplemented with molecules required in transcription such
as NTPs and RNA polymerase. During the last four decades a number of both pro- and eukaryotic in vitro systems have been made commercially available [253–255]. To facilitate the binding of eukaryotic ribosomes to the mRNA and ensuring a high rate of translation, Kozak consensus sequence [256] can be introduced upstream of the gene to be expressed.

The reduced reaction volumes and processing time make cell-free system good experimental systems for high-throughput studies. Especially, when working with artificial model proteins or expressing proteins from various different organisms [257]. A drawback is the relatively low yields, mostly due to the efficiency of energy supply and the accumulation of inhibitors. As an example, the reported yield in rabbit reticulocyte is in the range of microgram or in the fraction of a microgram per milliliter [258]. To visualize the newly synthesized proteins, radiolabeling is to prefer in part due to the low amounts of protein but also due to the presence of other proteins in the lysate. Radioactivity also facilitates quantification of the expressed proteins.

The in vitro expression system must be supplemented with a hydrophobic environment when studying membrane proteins. Since biological membranes are complex and several proteins participate in translocon-mediated membrane insertion [259], microsomes are a convenient choice. Being small vesicular fragments of rough ER, they contain most - if not all - components necessary for membrane protein insertion, including signal sequence cleavage and N-linked glycosylation [260]. Further, the protein content has been characterized [259; 261] and the membrane-insertion/translocation processes in this system are well-studied [112; 262–264].

7.1.2 N-linked glycosylation in topology mapping

The enzyme OST, a membrane protein complex in the ER (and microsomes) with the active site on the lumenal side of the membrane. It attaches a sugar moiety to an asparagine residue present in the consensus sequence Asn-X-Ser/Thr, where X can be any amino acid except proline [265; 266]. This glycosylation event occurs co-translationally [103] and increases the molecular mass of the protein with approximately 2 kDa [267]. This allows the separation of glycosylated and unglycosylated forms by SDS-PAGE.

7.1.3 Leader peptidase as a model protein

Leader peptidase B, Lep, (Uniprot P00803) is an inner membrane protease found in E. coli. It has two N-terminal transmembrane helices, TMH1 and TMH2, and a small polar cytoplasmic domain (P1) between the transmembrane domains and a large C-terminal periplasmic domain (P2) [268, 269]. When expressed in the presence of microsomes, Lep inserts co-translationally
into membrane, adopting its native topology with both both the N and C-termini in the ER lumen \([156,264]\).

**Figure 7.1:** The leader peptidase (Lep) as a host protein and the in vitro expression in the presence of microsomes system. Two versions of Lep (shown in grey) were used. a) H-segment (in red) placed in the P2 domain of Lep allows studying a transmembrane region in Nout-Cin orientation. b) H-segment replacing the second Lep transmembrane helix. In both model proteins Asn-Xaa-Thr glycosylation sites were introduced on both sides of the H-segment to allow topology determination. c) Glycosylation by the oligosaccharyl transferase occurs in the microsome lumen. Depending on the number of glycan attached to a protein, the topology of said construct can be deduced. Here, singly glycosylated species arise from proteins with the H-segment inserted into the membrane whereas double glycosylated species arise from a translocated H-segment. d) An example of in vitro translation in the absence (-) and presence (+) of microsomes. Each glycan adds around 2 kDa to the molecular mass of the protein allowing their separation on SDS-PAGE. Band of nonglycosylated protein marked with an open circle and bands corresponding to singly and doubly glycosylated protein species are indicated with filled circles respectively. In this example, about 75% of the protein is doubly glycosylated. By Minttu Virkki.

All proteins in this work are modified versions of Lep, where a prospective
transmembrane helix is introduced as H-segments. The orientation of said H-segment can be determined by using one of the two different Lep variants. In LepI or H3, the fused segment replaces a part of the C-terminal domain (Figure 7.1a). In LepII or H2 it replaces the second transmembrane helix (Figure 7.1b), since the first transmembrane domain is sufficient for membrane targeting and insertion [170]. In both cases, two glycosylation sites are placed along the sequence such, that the topology of the construct can be deduced (Figure 7.1c).

7.1.4 Determining the propensity to integrate into the membrane by an in vitro assay

As described above, unglycosylated, singly and doubly glycosylated protein species can be separated by SDS-PAGE (Figure 7.1d). Since the addition of a glycan increases the molecular mass of the protein, the signal intensities from each band in the gel can quantified and related to the relative amounts of inserted and secreted protein. From this data, the apparent change in Gibbs free energy upon insertion $\Delta G_{app}$ can be calculated for this system assuming a model where the translocon-mediated integration is regarded purely as an equilibrium process [112; 117]. The relative amounts of inserted ($f_1$) and non-inserted ($f_2$) protein can be written both as a probability of insertion:

$$p = \frac{f_1}{f_1 + f_2} \quad (7.1)$$

and as an apparent equilibrium constant

$$K_{app} = \frac{f_1}{f_2} \quad (7.2)$$

which, in turn can be placed in the equation for Gibbs free energy:

$$\Delta G_{exp} = -RT \ln K_{app} \quad (7.3)$$

where $R$ is the gas constant and $T$ the absolute temperature.

7.1.5 Minimal glycosylation distance mapping

The active site of the transferase is located at a fixed distance from the membrane. The distance to the active site has been determined; Half-maximal glycosylation is achieved when the Asn residue is (I) 14 residues at N-terminal or (II) 10 residues away at the C-terminal of the transmembrane segment [264].
Thus, the position of a transmembrane segment can be determined by designing a series of constructs, with the consensus sequence at different positions. The distance at which the acceptor-asparagine is glycosylated at half-maximal efficiency is said to be the minimal glycosylation distance (MGD) [130; 264; 270]. Figure 7.2 attempts to describe the method in more detail, using LepI as an example.

**Figure 7.2:** Determination of the "minimal glycosylation distance" (MGD). MGD is defined as the minimum number of residues required between the luminal end of a transmembrane segment and the active site of oligosaccharyl transferase. a) Shows a cartoon for N-terminal mapping of a H-segment (red) within Lep (grey). Two glycosylation sites (Y) are placed on each side of the H-segment to indicate if it is inserted to the membrane or not. The position of the third glycosylation (bigger and lighter Y) site is placed at various distances away from the end of the H-segment. This cartoon depicts the N-terminal mapping. b) An example of results for MGD determination for MscL TMH1. The doubly glycosylated band indicates the glycosylation status of the MGD-site for an inserted H-segment. In the first panel from right, most of the TMH1 is inserted but only 21% of MGD-sites are modified. When the acceptor asparagine is \(\approx 14\) residues away from the last membrane embedded residue, around 52% of proteins are glycosylated at MGD-site (the middle panel). In the last panel, the MGD-site is modified to 86%. Adapted from Paper III.
7.2 Membrane proteins \textit{in silico}

Experimental characterization of proteins in combination with ever-increasing number of sequence data have created a high demand for computational tools to analyze data and to test models \textit{in silico}. Computational biology focuses on the development for and application of these tools to study chemical and biological phenomenon. It is now possible to both predict the topology of a membrane protein based on sequence data only or to simulate the water-to-membrane transition of a peptide.

7.2.1 Topology predictions

As evident from this thesis, the concept of topology is widely used when discussing membrane proteins and consequently, is also relevant in membrane protein structure prediction. There are a number of different topology prediction methods available today and much effort has been put into measuring the level of performance of these methods. The earlier prediction methods relied on non-biological hydrophobicity plots and initially no attempts were made to assess the orientation of transmembrane domains [192, 271, 272]. Considering the features of amino acids in transmembrane domains discussed in chapter four, it may not be surprising that they performed poorly.

The positive-inside rule allows predictions to include the likely orientation of the prospective transmembrane domains proteins [273, 274]. Further, including information on sequence similarity, statistical and machine learning techniques, even more sequence features could be incorporated [5, 192, 275–277], which has improved prediction success. Despite some features such as re-entrant regions, exact membrane borders [126], [Paper III] or marginally hydrophobic helices [168] remain difficult to predict, the best topology predictors today perform with an accuracy of 70 - 90 % [192].

7.2.2 Computational estimation of membrane integration propensity of a transmembrane helix

The Hessa-scale, described in chapter five, provides information on how hydrophobicity, segment length and position of each amino acid in a prospective transmembrane domain affect its insertion propensity. The equation below combines all this information, which can be used to predict the insertion propensity for a prospective transmembrane domain. The \( \Delta G \) predictor is available at http://dgpred.cbr.su.se/, and can be used for calculating both the \( \Delta G_{\text{pred}} \) for a specific segment or for finding potential transmembrane domains within a longer sequence.
\[
\Delta G_{\text{app}}^{\text{pred}} = \sum_{i=1}^{l} \Delta G_{\text{app}}^{\text{aa}(i)} + c_0 \sqrt{\left( \sum_{i=1}^{l} \Delta G_{\text{app}}^{\text{aa}(i)} \sin 100i \right)^2 + \left( \sum_{i=1}^{l} \Delta G_{\text{app}}^{\text{aa}(i)} \cos 100i \right)^2} + c_1
\]

Where

\[
c_1 = b_1 + b_2 l + b_3 l^2
\]

The first term in this equation sums the individual, position-dependant (i) contributions for each amino acid (aa) in the sequence. The second term takes the helical dipole moment into consideration. The last term is fitted and \(b_1\), \(b_2\) and \(b_3\) were optimized to account for effects due to length (l), based on experimental data [112].

7.2.3 Understanding protein function with molecular dynamics

A wide variety of simulation techniques have been developed to help to understand the properties of small scale interactions that occur within and between biomolecules[278]. The most popular methods for simulating large biomolecules are those that fall under the umbrella of Molecular Dynamics (MD) methods.

The basic approach of MD is that the system of atoms and molecules are treated as a (semi)-classical object. Rather than performing explicit calculations of the electrons, their behavior is modeled as effective forces acting on the atoms. In essence, one has a function \((U(x_1, x_2, ..., x_n))\) which gives the potential energy of the system for any set of spatial coordinates for its atoms \((x_i)\), and equivalently the forces on the atoms are given by the classical equation

\[
F = -\frac{dU}{dx_i}
\]

The potential functions (or 'force field') are given by a set of functions that combine physical approximations and curve-fitting to experiment or quantum mechanical calculations. For example, the London force is typically modeled as a Lennard-Jones potential, known to approach the exact result at long range. As an empirical method, MD simulations can only be as accurate as the parameters used to specify the model. In the MD simulations performed in the current work, a force field according to Berger et al [279] was used, which has been parametrized specifically to reproduce the correct density and heat of vaporization for a lipid membrane.

Once the forces are given, the statistical behavior of a system may be studied. The input essentially being the initial structure/state of the system, for
which random velocities are assigned to the atoms although Boltzmann distribution is maintained. Time is stepped forward in discrete units. For each of which, the velocities positions and resultant forces on the atoms are recalculated using Newton’s equations of motion

\[ F = ma \]

The result is information about the statistical trajectories of the atoms, their conformations and motion, the energy and the behavior of the system in general.

The results of the calculations can be used to calculate the Potential of Mean Force (PMF). By calculating the mean overall force on a particle/molecule/atom/etc over many points in space, one can integrate the equation 7.4 stated above, and determine the energy of that particle/etc as a function of its location in space. Case-in-point: the energy of a transmembrane protein segment as a function of its location within the membrane.
8. Summary of papers

8.1 Paper I - The positive-inside rule is stronger when followed by a transmembrane helix

In this paper we studied transmembrane helices that are not recognized by the translocon but are found embedded in the membrane in the three-dimensional structure i.e. marginally hydrophobic helices (mTMH). Positive inside rule, orientational preferences of and specific interactions with neighboring helices have been shown to aid in the translocon-mediated recognition of mTMHs.

To better understand how this occurs, we studied three naturally occurring mTMHs, which were previously shown to require the subsequent helix (TMH<sub>sub</sub>) for efficient translocon recognition. We found no evidence for specific interactions. Instead, we identify arginines located at the N-terminal part of TMH<sub>sub</sub> to be crucial for the recognition of mTMH. This in line with previous results where the positive-inside rule has been shown to contribute towards the free energy towards membrane insertion.

However, in two of the constructs, these arginines do not aid in the recognition without the rest of the TMH<sub>sub</sub>. Instead, the improved recognition of mTMHs is due to the positive-inside rule being stronger when the positively charged residues are followed by a transmembrane helix. Such a mechanism obviously cannot aid C-terminal helices, and consequently, we find that the terminal helices in multi-spanning membrane proteins are more hydrophobic than internal helices.

8.2 Paper II - Insertion of marginally hydrophobic helix in EmrD (Manuscript)

The positive inside role governs the orientation of membrane proteins in such a way that more positively charged amino acids are found on the inside of a cell. The exact mechanisms how this is achieved is not well known, but it is clear that positively charged residues can facilitate the insertion of transmembrane helices that are not sufficiently hydrophobic to be inserted otherwise. Two features related to the positive inside rule have been described to enable the
insertion of mTMHs, namely the positive inside rule as such and orientation preference of a \( TMH_{sub} \). The distinction between positive-inside rule and orientation preference of a \( TMH_{sub} \) is somewhat unclear. No other clear signals than positively charged residues have been identified to have an effect on orientation preference. It is commonly accepted that negatively charged residues do not influence the orientation.

We showed in a recent study, that the insertion of EmrD mTMH2 is facilitated by positively charged residues, followed by a hydrophobic \( TMH_{sub} \). Surprisingly, the positively charged residues are not sufficient alone. If \( TMH_{sub} \) is deleted after the last arginine - mTMH2 cannot enter the membrane, despite the charge bias. Here we replaced the native \( TMH_{sub} \) with a 7L12A peptide, which allows mTMH to inserts to 74%. Even when the two arginines, which were essential in the context of the native \( TMH_{sub} \), are removed the insertion is still 72%. This supports our earlier results that there are no specific interactions between the mTMH and \( TMH_{sub} \) responsible for the translocon recognition. Further, this also demonstrates how fewer positively charged residues are needed when the mTMH is followed by a hydrophobic \( TMH_{sub} \). In other words, the positive-inside rule appears stronger when the positively charged residues are followed by a hydrophobic \( TMH_{sub} \). However, the hydrophobic peptide alone is not enough to allow mTMH to integrated to the membrane.

8.3 Paper III - Large tilts in transmembrane helices can be induced during tertiary structure formation

The general principles of translocon-mediated insertion of transmembrane helices (TMHs) are fairly well understood. In comparison, much less is known about the folding and oligomerization of alpha-helical membrane proteins in the membrane. Further, while topology predictions can predict the number and orientation of TMHs in a multi-spanning protein quite accurately, they are much less accurate at determining the precise ends of TMHs.

In this paper we address the question whether the tilt is an intrinsic property of a TMH or a trait acquired during folding. Four TMHs from polytopic membrane proteins found in highly tilted conformations in high-resolution structures were studied using a glycosylation mapping technique. We found that all four are inserted into the membrane as shorter membrane-embedded segments when expressed by themselves. This suggests that tilting can be induced by tertiary packing interactions within the protein, after the initial membrane-insertion step.

Further, we note that in all cases the region entering the membrane post-insertion contain polar and charged residues. These polar residues tend to pack
against other parts of the protein or snorkel into the lipid headgroup region. While these results pertain to rather extreme cases - the majority of long and tilted TMHs are hydrophobic - it is nevertheless apparent that TMHs can be repositioned perpendicular to the membrane plane during folding and undergo large tilting motions. This may also allow a more polar region of a TMH to enter the membrane - possibly explaining why some TMHs in membrane protein structures have a surprisingly high $\Delta G_{\text{pred}}$.

8.4 Paper IV - Folding of Aquaporin 1. multiple evidence that helix 3 can shift out of the membrane core.

The folding of most integral membrane proteins follows a two-step process: first, individual transmembrane helices (TMHs) are sequentially inserted into the membrane by the Sec translocon. Followed by folding in to the final conformation. Some proteins, including Aquaporin 1 (AQP1), do not follow this traditional two-stage mechanism.

AQP1 is first insert as a four-helical intermediate, where TMHs 2 and 4 are not inserted into the membrane and TMH3 is inserted in an inverted orientation. The final conformation with six TMHs is formed by TMH2 and 4 entering the membrane and TMH3 rotating $180^\circ$. We propose a mechanism for the initial step in the folding of AQP1. First, a shift of TMH3 out of the membrane core, pulling preceding regions into the membrane. TMH3 can then be re-inserted to the membrane in its correct orientation TMH3 and TMH4 enters the membrane.

In this paper we present three observations supporting this idea. I) The preceding loop before TMH3 contains an additional conserved hydrophobic region and this region can be efficiently inserted to the membrane by itsel. II) Experimental, predicted and simulated hydrophobicity data implies a relatively low barrier between TMH3 and the preceding loop. III) A glycosylation mapping technique further suggest that several alternative segments within the R1-TMH3 region can be recognized as transmembrane by the translocon. Lastly, the positive-inside rule would also favor the shift, since this would allow arginines to move to the cytoplasmic side of the membrane.

8.5 Final thoughts

Transmembrane helices have been viewed as simple anchors composed of mostly hydrophobic amino acids. Further, the influence on membrane protein structure and function was considered mostly independent of the environment (lipid bilayer) they reside in. With more solved three-dimensional structures
the amount of structural variation has become obvious. Some transmembrane helices are surprisingly hydrophilic and are not recognized as transmembrane segments by themselves - these mTMHs are the main focus of this thesis.

Recent studies have identified some features in the local sequence context that aid in the insertion of mTMHs. These are the positive inside rule, orientational preferences of and specific interactions with neighboring helices. However, little is known about how common each of these features are. Further, in paper I we observe some intriguing behavior of seemingly equivalent features. In AcrB, charge-bias does not seem important if a hydrophobic helix is present but, in the absence of the $TMH_{sub}$ the charge-bias alone is enough for efficient membrane integration of its mTMH. In contrast, EmrD mTMH2 is dependent on a hydrophobic $TMH_{sub}$. In paper II, we continued to study this observation further. We were able to show how there may be a relationship between the number of positively charged residues in the cytoplasmic loop and the hydrophobicity of $TMH_{sub}$. We also attempted to look at if any conservation regarding mTMHs and local sequence context could be identified, but did not find anything. In the future, I would very much like to test the same constructs used in Paper II in reverse orientation. Here, the $TMH_{sub}$ inserts with its N-terminus towards the ER-lumen, which may influence its ability to bring mTMH into the membrane.

Another structural feature in membrane proteins are the tilt-angle of TMHs. In paper III, we investigated whether the tilt is an intrinsic property of the a particular TMH or if it is a feature acquired during folding. In all four TMHs we studied, the TMH alone inserted as a shorter segment, suggesting that tilting can be induced by tertiary packing interactions within the protein, after the initial membrane-insertion step. We also noted, that the section of a TMH that was not initially inserted but is membrane-embedded in the high-resolution structures contains many polar and charged amino acids. Perhaps repositioning together with interactions with lipid-headgroups and through burying polar residues in the protein interior allow the insertion of mTMHs as well? This behavior also suggests that membrane bounderies are not necessarily well-defined and implies structural mobility in membrane proteins.

Paper IV is definitely my favorite, as it essentially brings together everything I have worked on in this theses. Aquaporin 1 (AQP1) contains both mTMHs and re-entrant regions. As a consequence, translocon mediated insertion results in an intermediate topology with four transmembrane domains. Later repositioning effects are then required for the formation of a six TMH topology.

We made three interesting observations regarding the sequence of AQP1. Most importantly, the preceding loop before TMH3 contains a conserved hydrophobic region. This region, when presented to the translocon by itself, can
integrate into the membrane very well. This region is absent in a close homolog AQP4, which instead has a hydrophobic TMH2. Further, experimental, predicted and simulated hydrophobicity data also imply a relatively low barrier between TMH3 and the preceding loop. This goes back to my discussion regarding the actual cost of polar groups in the membrane, discussed in chapter 5. The interactions an arginine can make when placed within the membrane as well as the possibility of polar groups from membrane embedded proteins probably explain this. Using a glycosylation mapping technique we could also show that several alternative segments within the R1-TMH3 region can be recognized as transmembrane by the translocon. Lastly, the positive-inside rule might favor this shift as it would allow Arg93 to move to the cytoplasmic side of the membrane. Again, this movement of Arg93 may not pose as high cost as one might intuitively expect.

I tredimensionella bilder av membranproteiner har man sett att en del TMHar är överraskande hydrofila. Dessa kallas för marginellt hydrofoba transmembrana helixar (mTMH) och känns inte igen av translokonet utan närvaro av närliggande delar av samma protein. Under de senaste åren har man identifierat ett par olika mekanismer på hur andra delar av samma protein kan hjälpa ett sådant segment att komm in i membranet. I denna avhandling har vi tittat närmare på dels hur närvaro av en mTMH kan påverka insättning och väckning av ett membranprotein, samt hur närliggande delar av samma protein kan facilitera insättning av ett mTMH. I relation till detta har vi även undersökt hur TMHar kan ompositionera sig i membranet.

Vi visar att i tre naturligt förekommande mTMHar behövs närvaro av en efterföljande TMH. Dock är det inte en fråga om specifika interaktioner som det tidigare föreslagits. Istället kan två snarlika mekanismer för deras insättning urskiljas. I det ena fallet, möjliggör närvaro av positivt laddade aminosyror i den cytoplasmiska slingan att mTMHn sätts in i membranet. I två andra fall är de positiva laddningar inte tillräckliga om de inte följs av en annan, hydrofob TMH. Vidare, det naturliga TMHn kan ersättas med en hydrofob modellhelix i vilket fall positiva-laddningar inte längre är lika viktiga. Vi finner också att


Transmembraaniproteiinien rakenteen voidaan jakaa niin sanottuihin transmembraaneihin osiin - jotka sijoittuvat solukalvon sisälle - sekä solukalvon ulkopuolelle sijaitseviin silmukoihin. Solukalvon sisäiset osat koostuvat hydrofoobisesta aminohapoista ja muodostavat usein spiraalin muotoisen rakenne - ”α-kierteen”. Näitä transmembraanisia α-kierteitä (TMKitä) sitoo yhteen solukalvon ulkopuoliset silmukat, jotka koostuvat pääasiassa hydrofiilistä aminohapoista. TMKt siirtyvät solukalvon proteinisynteesin aikana erityisen kanavan, translokaattorin, kautta. Translokaattorin kohdatessa riittävän pitkän ja hydrofoobisen α-kierteen, se auttaa tämän solukalvon sisälle. Hydrofiiliset silmukat puolestaan voidaan joko jättää solun sisäpuolelle tai translokaattori voi päästä ne kanavaa pitkin kalvon toiselle puolelle - riippuen edellisen TMKn suunnasta (pää- vai häntä-edellä).

Transmembraaniproteiinien rakennetta tutkittaessa kolmiulotteisten kuvien avulla on kuitenkin löydetty epätavallisen hydrofiilisiä TMKt, niin sanottuja marginaalisen hydrofoobisesta transmembraani α-kierteitä (mTMKt). Translokaattorien kohdatessa riittävän pitkän ja hydrofoobisen α-kierteen, se auttaa tämän solukalvon sisälle. Hydrofiiliset silmukat puolestaan voidaan joko jättää solun sisäpuolelle tai translokaattori voi päästä ne kanavaa pitkin kalvon toiselle puolelle - riippuen edellisen TMKn suunnasta (pää- vai häntä-edellä).

Näiden piirteitten tarkempi ymmärtäminen on ollut tämän väitöstutkimuksen kohteena ja voimmekin näyttää, kuinka mTMKt seuraava TMK voi edesaut-
taa mTMKn siirtämistä solukalvon sisälle. Osoitamme kuinka mTMKt kolme-
ta eri proteiinista vaativat myöhemmän TMKn päästämiseen solukalvon sisälle. Kyseessä ei kuitenkaan ole kierteitten välisistä vuorovaikutuksista, kuten aiemmin on ehdotettu. Sen sijaan toteamme jo aiemmin havaitun ilmiön, jossa emäksiset aminohapot solulimassa puoleisessa silmukassa pakottavat mTMKn solukalvoon. Tämä ei kuitenkaan vaikuta olevan yleismaallinen tai ainut mekanismi, sillä kahdessa muussa tapauksessa mTMKta seuraavat emäksiset amino-
hapot itsessään eivät riitä, vaan vaativat lisäkseen hydrofaobisen TMKn. Tämä
luonnollinen TMHn voidaan korvata teennäisellä TMK:lla, jolloin emäksisten
aminohappojen merkitys mTMKn saattamiseksi solukalvoon pienenee. Näytämme
myös, kuinka hydrofiiliset aminohapot voidaan saattaa solukalvon sisälle myöhem-
mässä vaiheessa. Neljässä tutkimamme tapauksessa translokaattori tunnista
hydrofobisimman osan pitkästä TMK:stä ja tämä solukalvoon saatettu
segmentti on huomattavasti lyhyempi kuin mitä havaitaan proteiinin kolmiu-
loitteisessa rakenteessa.
Lisäksi olemme tutkineet kuinka mTMK vaikuttaa transmembraaniprotei-
inin laskostumiseen. Akvaporiiini 1 omaa kuusi TMKta mutta translokaattori
saattaa solukalvoon vain neljä näistä. Lopullisen rakenteensa saavuttamisek-
si on Akvaporiiinin läpikäytävä huomattavia rakenteellisia muutoksia. Ensinnäkin, mTMK on vedettävä kalvon sisälle mikä myös vaatii kolmannen TMKn
suunnan kääntämistä, sillä se saatetaan alunperin solukalvoon ylösalaisin. Hy-
drofobisuuskäyrien, solukalvoon saaton tehokkuksien kokeellisten mittausten
sekä molekyylidynamikka simulatioiden perusteella ehdotamme, kuinka nämä
rakenteelliset muutokset ovat mahdollisia. Tulostemme perusteella kolmas TMK
pystyy kallistumaan jyrkästi ja näin ollen siirtymään pois solukalvon keskust-
tasta. Näin sitä edeltävä silmukka pääsee solukalvon sisään luoden suotuisam-
mat olosuhteet rakenneuudistusten suorittamiselle.
Acknowledgements

A large number of people helped me to do this during my years at DBB - in and outside of academia. You know who you are but the following people deserve a special thank you. This will be long, so consider yourself warned.

My supervisors Arne Elofsson for welcoming me into the lab and for all your help during these years - especially during the last hectic week before my thesis submission. I have greatly enjoyed the freedom I was granted and appreciate the support in finding the next place to land. Susana Cristobal for help and advice in the lab. IngMarie Nilsson for all your knowledge regarding the in vitro assay and for always being so supportive. Gunnar von Heijne for collaborations, letting me join the journal club and for sharing your knowledge regarding membrane protein biogenesis. Dan Daley for a very nice collaboration in rare-codon projects. Stefan Nordlund and Elzbieta Glaser for your valuable support and feedback. Also - including Martin Högbom - thank you for the very nice and interesting discussions during the Big Exam - that was really, really fun!

Dear co-authors - it has been wonderful working with you all! Special thanks to Anni for a stimulating and exciting collaboration - you managed to convince me that Aquaporin1 is a cool protein to work with! Lisäksi opin paljon proteiini rakenteista. Kiitos myös työn ulkopuolisiin keskusteluusiin, pidetään yhteyttä! Stephen (for being so calm and relaxed.. I do not know how you do it!), Christoph (for your bioinformatics experties and understanding exactly how annoying writing a paper can be;) Sara L, Morten, Nitin, Carolina, Elin, Therese, Nanjiang, Kostas and Daniel - it has been great to work with you! Also a special thanks to Karin Ö and Florian for support in the world rough microsomes and in vitro glycosylation and for all the interesting discussions about membrane proteins - best of luck in the future.

To master something you have to teach, right? I have been really happy to have so many great students working with me on various projects - even if they may not have ended up published. I owe a big thank you to Elin, Riccardo, Sara B. and Therese. I wish you all the best in the future, it was really fun working with you and I am so happy to see how you all already have - or intend to - go on doing a PhD of your own. Good luck, but beware: I will stalk you on PubMed.;)
I have also been lucky with having a really nice group to work with - past and present! Sara L., för all sällskap under helgarbete, alla intressanta diskus- sioner, hjälp med diverse forskningsrelaterade ämnen och bioinformatiken- du är en mentor alla doktorander borde ha! Marcin and family. It was great being in the same group as you - I really enjoyed our discussions and geekerries through out the years. Thank you for inviting us to your wedding - visit to Poland was unforgettable! I am also grateful for all those dangerously delicious dinners (the cake is not a lie in your house for sure!) and the company and support. A big thank you for helping with some latex related problems with this thesis. Walter not only are you super fun to be around, you are a fantastic drummer and come with an awesome family! Thank you and Elisa for touring rainy Dalarna, visiting us in Finland and being great friends. It has been so much fun to have someone to share all these rage-flip temptations and fffffffuuuuuuu moments during these years. All the best! Kostas, travelling with you is a recipe for success! Celebrity spotting (Snoop Dog!), gourment dinners (Uniteds Dagens Lobster) and excitement (driving like a police and getting donuts).

Best of luck to the current students in the Elofsson lab: Per, Karolis, Mirco, Oxana, Christoph, Marco, Alba, David and Nanjian.

In the memory of the infamous C4 corridor - you guys were the best! Marcin, Linnea, Henryk, Mike, Sophie, Arjun, Rossen, Szilard, Aaron, Sikander, Anna, Walter, Christoph, Daniel, Susan, Kostas, Patrik and Rauan... my first years at DBB were awesome much thanks to the karaoke, dinners and parties - Leonardo Cohens Hallelujah will always be Michelujah for me! To those of you who also attended the conference in Maratea 2010 - the best conference during my PhD! I am sure our innovative and politically incorrect apps (iBLOW) more than impressed our table companion - professor Bowie.

To the best office ever: I salute the Senior Bitches Maria, Linnea and Anna Amelina. We low-life leftovers eating PhD students - Narges Bayat, Gabriela Danielsson and Daniel - will join them soon! Maria, you were like a rock for students struggling with their projects (and life). Anna, thank you for sharing your knowledge and experience in the lab, and the interest for ballet on the outside. All the best to you in England - you are, indeed, a wonderful person!;) Daniel, I have never met anyone as positive as you - I have a lot to learn in that regard. Thank you for your support and all the best in the future! Linnea, you taught me how to multitask and be super-efficient in the lab and pretty much became the big sister I never had. Thank you for the awesome, crazy year of fikas, karaoke and dancing until the sun came up. And for our conversations - serious and silly.
Narges "look at my eyes" Bayat - it all started by us bonding over our hate for yeast and having our pictures taken, ending in a girl crush de luxe - where movie nights and dinner for four is just basic. For your last months, *Toughen up princess and dance (for bandz)*. You will have time to improve those social-skills later.:) **Gabriela**, tack för allt! Jag har lärt mig otroligt mycket om livet av dig. Du tänker mycket och jag älskar våra reflekterande diskussioner. Du anar inte hur mycket jag uppskattade diskussioner som uppstog när vi pluggade till stora tentan.

And to some of you from the "other" side: **Patricia**, I am so happy we taught the same course- it feels like ages ago. Our super long lunches and occasional inappropriate teaching tactics still make me smile. You have definitely saved some dark lab-work days for me with your tele-novella real life stories!:) **Candan**, I have you to thank for introducing me into the fantastic world of lipids - and for correctly foreseeing my future.;) You are a great scientist and a wonderful friend. I had so much fun with you both in the lab, during coffee breaks (...because good science comes with a belly). One day we’ll have offices next to each other.;) **Matthew**, you were so very, very welcome to the almost empty lab on C4. And the music definitely got better with you there. Thank you for all your advice, suggestions and proof-reading this thesis! **Julia**, thank you for the nice dinners! I just wished they were more numerous and will continue. All the best to you and thank you for listening and being there for me last year. **Göran**, thank you for sharing your knowledge in the lab and for providing a post-PhD perspective on a lot of typical(?) PhD-problems. **Anirudh**, thank you for sharing your knowledge regarding organic chemistry and for always being so much fun to talk to! Still doing everything on time?;) **Björn**, thank you for great travel company in Virginia! Instead of dying out of boredom at the airport I almost died from laughing too much. And the donut trip was epic! Good luck with your PhD! **Marco M.** on Scilife floor 4, it was great fun going through the big exam, tibs and pre-dissertation with you. Thank you for those more or less fun times and good luck in future.

I also want to thank all my co-workers, administrators and technical staff at DBB and Scilifelab for making it such a friendly and fun environment to work with!

... and of course a special thanks to everyone in the **GvH, JWdG, DD, RD** complex for being so knowledgeable and kind. It has been great sharing the labs with you and good luck to you all! A big thanks to **Ann Nielsen** and **Maria Sallander** for answering all questions and concerns plus for always being so nice and understanding!

Finally, I would like to acknowledge people outside of work. **Astrid**, jag vill tacka dig för våra diskussioner där jag fått sätta kemi i vardagsperspektiv (ätbara bensenringar!). Vi ses för afternoon tea för lite snobberier, på vårt
favorit häng - T-centralen! **Cata** - Louisiana hot sauce, too PhD for you and lets stretch bitches! Tack för att du bidragit med många skratt under alla dessa år. För att inte tala om hur jag kan nästan komma ner till spagat tack vare våra stretch stunder i tid och otid. We love you! **Karin**, kiitos kanssatuesta pitkine päivineen. Näihin vuosiin mahtuu paljon tapahtumia ja muutoksia - olen iloinen, kun olen saanut jakaa ne kanssasi. Kuka olisi uskonnut, miten elämä muuttuu Zorron ja Ritari assän jälkeen huomattavasti harmoonsempaan suuntaan jossa puutarhurointi ja käsityöt ovat pop. **Kia**, kiitos sinulle, kun jaksat kuunnella ja analysoida kaikkea mitä näihin viiteen vuoteen on mah-tunut - Nauravine ja itkevine paraabeleineen - life, what life? **Riina**, olen usein miettynyt, kuinka harmi oli ettei satuttu olemaan samassa työpaikassa - sen verran on nimittäin ollut tumapelon-tempelissä tapauksia. Mutta onneksi olen sinulta oppinut arvostamaan kuperia kysymyksiä: "puheesi on kuin kärpäsen surinaa korvissani". Josko nyt vaikka ehtisi Turkuunkin?;) My family, both original and extended. Kiitos tuesta ja ymmärryksestä: **Äiti** - olet rakas, **Isä** - lepää rauhassa, **Ville** - kiitos keskusteluajeluista ja musiikista. Joko oot arvannu mun käyttäjäni-men eräällä nimeltämainitsemattomalla nettisivulla? **Pekka** - on ollut mukava keskustella sekä sinun että **Saanan** kanssa, kaikkea hyvää teille molemmille! **Eva, Marva** och **Guy** - tack för att ni välkomnat mig i er familj, kunde inte önska mig en mer underbar svärfamilj - ni är älskade! And of course **Sven**! I’m so glad I met you! You’ve stood by me from the very beginning preparing buffers to the end, being my physical chemistry lexicon and reference manager. Thank you for being genuinely supportive and believing in me when my scumbag brain didn’t. With you, even yeast becomes fun & cute! Not to mention all bad puns... axon-exon.
References


multisubunit complex Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound endoplasmic reticulum-associated degradation in human physiology. Cell, 97:553–564, 1999. 31


The positioning of a transmembrane helix in the microsomal membrane is not easily accommodated in a membrane protein.

Membrane protein topology by a single C-terminal residue in mammalian microsomes is more efficient in Escherichia coli than in mammalian microsomes.

Proline substitutions are not easily accommodated in a membrane protein.

Hinges, swivels and switches: the role of prolines in signalling via transmembrane alpha-helices.

Membrane protein folding and stability: physical principles.

How proteins adapt to a membrane-water interface.

The aromatic residues Trp and Phe have different effects on transmembrane helix packing.

The hunchback and its neighbours: proline as an environmental modulator.

Transmembrane glycine zippers: physiological and pathological roles in membrane proteins.

Transmembrane alpha-helices correlate with the trans-membrane topology.

Positively charged amino acids placed next to a signal sequence block protein translocation more efficiently in Escherichia coli than in mammalian microsomes.


plasmic reticulum Sec61 translocon

signal and signal-anchor sequences are determined by the balance between the hydrophobic
transfer sequence for microsomal membrane
brane protein
brane protein assembly
core evolutionary conserved?
Proteins
family of fluoride-specific ion channels with dual-topology architecture
dual-topology membrane proteins
and functionally important
S5-P-S6
portion of voltage-sensing segments, S3-S4, and independent insertion of pore-forming segments,
Shaker-type K+ channel, KAT1, into the endoplasmic reticulum membrane: synergistic inser-
46

membrane protein assembly
and Asp-mediated interactions between transmembrane helices during translocon-mediated

Identification and evolution of

A family of fluoride-specific ion channels with dual-topology architecture. Elife, 2:e01084, 2013. 46


Integration of Shaker-type K+ channel, KAT1, into the endoplasmic reticulum membrane: synergistic inser-
tion of voltage-sensing segments, S3-S4, and independent insertion of pore-forming segments,
46, 49, 52

Identification and evolution of polar residues within the membrane


Why are polar residues within the membrane


Aquaporin water channels–from atomic structure to clinical medicine. J. Physiol. (Lond.), 542:3–16, 2002. 47


A. WLODAWER, W. MINOR, Z. DAUTER, AND M. JASKOLSKI. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. FEBS J., 275:1–21, 2008. 57


