The social life of a membrane protein; It’s complex.

Isolde Palombo
Cover: Illustrating the social life of membrane proteins. Derived from Nighthawks by Edward Hopper from 1942.

©Isolde Palombo, Stockholm University 2013


Printed in Sweden by Universitetsservice AB, Stockholm 2013
Distributor: Department of Biochemistry and Biophysics
To myself, for hanging in there.
List of publications

I  Stephen Toddo, Bill Söderström, Isolde Palombo, Gunnar von Heijne, Morten H. Nørholm and Daniel O. Daley
Application of split-green fluorescent protein for topology mapping membrane proteins in *Escherichia coli*.
*Protein Science* (2012) **10**, 1571-1576

II  Isolde Palombo and Daniel O. Daley
Heme incorporation into the cytochrome *bo* occurs at a late stage of assembly
*FEBS Lett.* (2012) **586**, 4197-4202

III  Isolde Palombo, Daniel O. Daley and Mikaela Rapp
The periplasmic loops provides stability to the open state of the CorA magnesium channel

IV  Isolde Palombo, Daniel O. Daley and Mikaela Rapp
Why is the GMN-motif in the CorA/Mrs2/Alr1 superfamily of magnesium channels conserved?
*Submitted*
Abstract

Membrane proteins are key players in many biological processes. Since most membrane proteins are assembled into oligomeric complexes it is important to understand how they interact with each other. Unfortunately however, the assembly process (i.e. their social life) remains poorly understood. In the work presented in this thesis I have investigated when and how membrane proteins assemble with each other and their cofactors to form functional units. We have shown that that cofactor insertion in the hetero-tetrameric cytochrome bo$_3$ occurs at an early state in the assembly process. We also found that the pentameric CorA magnesium ion channel is stabilised by different interactions depending on the magnesium ion concentration in the cell. These studies indicate that the assembly of a functional unit is a dynamic process, which is a result of many different forces. By studying the assembly of membrane proteins we have obtained a deeper insight into their function, which cannot be explained by static crystal structures.
Contents

List of publications ..............................................................iv
Abstract ................................................................. v
Abbreviations ................................................................. viii
The model organism, Escherichia coli ........................................ 9
The E. coli cell envelope ......................................................... 9
Types of membrane proteins in E. coli ...................................... 11
The life of an α-helical membrane protein in Escherichia coli ........ 13
How does an α-helical membrane protein reach its destination? .... 14
How does an α-helical membrane protein fold? ......................... 16
Functional units in the E. coli inner membrane ......................... 21
Why are α-helical membrane proteins assembled into complexes? .. 23
How do α-helical membrane proteins assemble into complexes? ... 24
How do we detect α-helical membrane protein complexes? .......... 27
Objectives of thesis .............................................................. 30
Short summary of papers ....................................................... 31
Concluding remarks ............................................................ 37
Populärvetenskaplig sammanfattning på svenska ......................... 39
Acknowledgements ............................................................. 41
References ................................................................. 44
Abbreviations

<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>BN-PAGE</td>
<td>Blue Native polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>DDM</td>
<td>n-dodecyl β-D-maltoside</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SEC</td>
<td>Size exclusion chromatography</td>
</tr>
<tr>
<td>SRP</td>
<td>Signal recognition particle</td>
</tr>
<tr>
<td>Tat</td>
<td>Twin arginine translocation</td>
</tr>
<tr>
<td>TM</td>
<td>transmembrane</td>
</tr>
</tbody>
</table>
The model organism, *Escherichia coli*

*E. coli* is a gram-negative bacterium that was discovered by Theodor Escherich in 1885. It is commonly found in the intestine of warm-blooded animals where it contributes to the normal flora. It is beneficial for the host as it produces vitamin K and prevents the establishment of pathogenic bacteria (1,2). However, some strains of *E. coli* can cause food poisoning, which recently was the case in Europe where the pathogenic strain EHEC quickly spread and caused severe disease (3).

*E. coli* is heavily used as a model system in biological research as many fundamental biological processes are conserved from *E. coli* to humans. Much of our knowledge about biochemistry, molecular genetics and structural biology is therefore based on research done with this organism (4,5). In my work I have used *E. coli* as a model system for investigating how α-helical membrane proteins fold and assemble into functional units.

The *E. coli* cell envelope

The envelope of *E. coli* is an interface with the extracellular milieu. It contains two membranes (termed inner and outer) that are separated by a periplasm and a layer of peptidoglycan (Figure 1). The peptidoglycan layer acts as an external cytoskeleton, maintaining the physical integrity of the cell (6,7). It is constructed from glycan strands of N-acetylmuramic acid (Mur-NAc) and β1,2-linked N-acetylglucosamine (GlcNAc) that are connected by β-1→4 bonds. A mesh of cross-linked peptides connects the long glycan
strands. The whole structure is mainly connected to the outer membrane via a peptide bond to a C-terminal lysine of Braun’s lipoprotein (LPP) (8,9).

Figure 1. A representation of the *E. coli* cell envelope. (Reprinted from (10) with permission from Nature Publishing Group)

The two membranes differ in lipid composition. The inner membrane and the inner leaflet of the outer membrane contain about 75% phosphatidylethanolamine (PE), 20-25% phosphatidylglycerol (PG) and 5% cardiolipin (CL) (11), while the outer leaflet of the outer membrane is comprised of lipopolysaccharides (LPS) (12-14). The two membranes also differ in protein composition and therefore have many functional differences. The inner membrane contains a more diverse repertoire of proteins than the outer membrane. It has been predicted that there are >1000 different integrated inner membrane proteins and together with the peripherally attached, they make up about 50% of the total proteome (15-20). These proteins are involved in a great number of important events in the cell, such as signalling, energy conversion, division and transport of ions, small molecules, nutrients and toxins. There is also a large fraction of predicted inner membrane proteins that have unknown function (18) (Figure 2). In contrast, the outer membrane contains fewer proteins (around 2-3% of total proteome) (21-23). They are involved
in ion transport and passive nutrient intake. Furthermore, they function as membrane anchors and as defence against attack proteins (23).

![Functional categorisation of the E. coli inner membrane proteome. (Reprinted from (18) with permission from AAAS.)](image)

**Types of membrane proteins in E. coli**

Inner membrane proteins can be either attached to or embedded in the lipid bilayer. The latter contains transmembrane (TM) domains that span the membrane with an α-helical structure (Figure 3). These regions have an average length of 19 hydrophobic amino acids that typically are flanked by aromatic residues and connected by polar loop-residues to assemble in α-helical bundles (24,25). There are also proteins that are peripherally attached to the inner membrane, these proteins are loosely associated and have the ability to move along the membrane or detach from it (26). The outer membrane contains lipoproteins and proteins with a β-barrel structure (21-23). β-barrel proteins are composed of an even number of β-strands, which mostly form channels (Figure 3).
Figure 3. Representation of an α-helical membrane protein fold (PDB accession code 1FFT) (left) and a β-barrel structure (PDB accession code 1QJP) (right).

Lipoproteins are mainly found at the inner leaflet of the outer membrane, but can also be found in the outer leaflet or become secreted to the extracellular milieu. They are anchored to the lipid bilayer by a lipid-modified cysteine residue in their N-terminus and play important roles in bacterial physiology and virulence (reviewed in (9)). There are a few examples of lipoproteins that are attached to the periplasmic side of the inner membrane of E. coli. One of those is CyoA, which is one of the subunits of the respiratory enzyme cytochrome bo₃ ((27) and Paper II).
The life of an $\alpha$-helical membrane protein in *Escherichia coli*

The life of a membrane protein in *E. coli* can be broken down into a series of discrete events. Following gene transcription, the mRNA is translated by ribosomes in the cytoplasm. Thereafter, the protein is targeted to the inner membrane, and inserted into the membrane in a co-translational manner. Information regarding the protein’s final destination lies within the N-terminus of the nascent amino-acid chain. The folding of a membrane protein occurs either after or during its insertion into the inner membrane. Most membrane proteins interact with other proteins (membrane embedded and soluble) to form functional hetero- or homooligomeric complexes. Thus, similar to human beings, membrane proteins require a social life to function properly. In this introductory section, I am going to describe the molecular details about the life of an $\alpha$-helical membrane protein in *E. coli* (summarized in Figure 4).

![Figure 4. Steps in the life of a membrane protein (see text for details).](image)
How does an $\alpha$-helical membrane protein reach its destination?

Membrane proteins that are destined to the inner membrane of *E. coli* are usually targeted to the Sec translocon (SecYEG) co-translationally (Figure 5). In order for an integral membrane protein to be targeted and inserted into the inner membrane of *E. coli* it requires a specific signal, the signal anchor sequence, which is part of the N-terminus of the amino-acid sequence. The signal required for targeting to the SecYEG typically has a charged section and a hydrophobic section (28,29).

![Figure 5. The targeting and insertion of an $\alpha$-helical membrane protein.](image)

The signal anchor sequence is recognised by the signal recognition particle (SRP) as it exits the ribosome (30-33). The SRP-system of *E. coli* contains two proteins, which both are essential for growth. Ffh together with a small 4.5S RNA makes up the SRP-protein and FtsY is the SRP-receptor, which can either be attached to the cytoplasmic membrane or free in the cytosol. Upon binding of SRP with the signal anchor sequence, the SRP-ribosome-nascent chain complex is stabilised and directed to the cytoplasmic
membrane (34,35). At the membrane, SRP binds to its receptor, FtsY on the membrane surface and the ribosome-nascent chain complex will associate with the heterotrimeric SecYEG and the nascent chain continues down the translocation pore, while the SRP-receptor complex dissociates (36).

The SecYEG facilitates the integration of hydrophobic polypeptides into the inner membrane by setting the hydrophobicity threshold for integration of membrane proteins. It also provides a pore for proteins that are destined for the periplasm, the outer membrane or the extracellular milieu. (37,38).

Some Sec-dependent proteins of the inner membrane require the insertase/chaperone YidC for proper integration and folding (39-42). Furthermore, YidC can act as an independent insertase in the integration of smaller, and usually monotopic membrane proteins (43). YidC is an integral membrane protein that belongs to the YidC/OxaI/Alb3 family, whose members are found in bacteria, mitochondria, chloroplasts and some archaea (44-46).

In bacteria and the chloroplasts of plants, there is a second translocon, the Twin arginine translocation (Tat) translocon, which will translocate fully folded proteins across the membrane (47). These proteins will be targeted to the Tat translocon by a specific N-terminal signal sequence that contains a conserved twin-arginine motif. Recently, it has been shown that the Tat translocon is co-operating with the SecYEG in the insertion of one α-helical membrane protein (48). The components of the Tat translocon are assembled when the target protein binds to TatC with its N-terminal signal sequence.
How does an $\alpha$-helical membrane protein fold?

*The three-stage model: A framework for understanding membrane protein folding*

Two decades ago Popot and Engelmann carried out refolding studies of proteolytically cleaved and denatured bacteriorhodopsin. Based on their observations they presented a folding model for $\alpha$-helical membrane proteins, named “The two-stage model” (49). The general basis for the model is that individual $\alpha$-helices of a polypeptide reach a thermodynamic equilibrium with the lipids and aqueous solution to form individual TM domains before interacting with other TM domains. Thus, in the first stage, $\alpha$-helical segments are stably formed in the membrane and in the second stage, these helices interact with each other (49,50) (Figure 6). This model was later elaborated into a three-stage model. The additional stage describes how interacting helices create an internal space that is less hydrophobic, and allows for binding of prosthetic groups and polypeptides that are less prone to form regular TM helices (51) (Figure 6). Much of what is we currently know about membrane protein folding can be explained by the framework of the three-stage model.

![Figure 6. The three-stage model on the folding of helices and insertion of prosthetic groups (denoted with a P). The first stage shows the folding of individual helices, which in the second stage interact to form cavities for the partitioning of prosthetic groups in the third stage. (Reprinted from (51) with permission from Elsevier.)](image-url)
Experimental studies of membrane protein folding

The folding of α-helical membrane proteins has been studied extensively experimentally. It is a complex process that is initiated in the ribosome exit tunnel, where the nascent polypeptide chain starts to adopt its secondary structure (52-56). The final folding and orientation (i.e. topology) is determined by three parameters, (1) the translocon, (2) the sequence of the protein and (3) the composition of the lipid bilayer itself (37,38,57-59).

The SecYEG can identify TM domains and orientate them accordingly. Mutagenesis studies have shown that charged residues in the polar loops of the SecYEG interact with charged residues of the signal anchor sequence of the membrane proteins, which sets an orientation that the protein is obliged to follow (60,61). How efficiently a polypeptide segment becomes inserted into the membrane depends on its hydrophobicity, its overall length and the distribution of polar and charged residues (38,62). It is proposed that the TM helices of membrane proteins are inserted into the membrane by lateral movement through a gate in the SecYEG (63). The molecular details of this insertion are yet not fully understood, however, helices can be released into the membrane sequentially (one by one), pair wise, or as a bundle of helices (64). Once in the lipid bilayer, the polypeptide will form its tertiary fold, as a result of protein-protein and protein-lipid interactions, as discussed below.

The N-first insertion is limited to N-terminal signal anchors. Membrane proteins that have an internal signal anchor sequence follow a different insertion process, which is independent of the positive-inside rule, flanking charges and the overall hydrophobicity of the protein. The effect of hydrophobicity and the apparent signal inversion is largely lost when the hydrophobic signal is moved towards the C-terminus of the protein and the hydrophilic N-terminal part is extended (65).
How does the sequence of the protein affect folding?

The structure of an α-helix is stabilised by hydrogen bonds between amino and carboxyl groups of the peptide bonds. These regions need to interact and pack together in the membrane and they do so using general chemistry such as hydrogen bonds, van der Waal contacts and salt bridges (the knobs-into-holes model) (66,67). Specific sequence motifs also mediate strong helix association between TM helices, the most well characterised being the GXXXG motif (68). The occurrence of polar residues within the TM region is very rare and it would be energetically costly to integrate them in the hydrophobic membrane.

How does the lipid bilayer affect folding?

Lipid-protein interactions are important when the protein is being translocated, since the SecYEG requires the presence of anionic phospholipids for proper protein integration (69,70). Furthermore, positively charged residues in the extramembranous regions and negatively charged lipids in the membrane can influence membrane protein insertion and folding (71). It has also been shown that some lipids may act as molecular chaperones in protein folding, based on the observation that proteins can become misfolded in the absence of specific phospholipids in the membrane, even though the targeting to the membrane remains unaffected (72). Additionally, properties of the lipid bilayer (i.e. bilayer thickness, lateral pressure and membrane fluidity) are a driving force for the folding of α-helical membrane proteins (73). The thickness of the bilayer is related to the hydrophobic matching between the helices and the bilayer, where matching stabilises the helix-interaction, and mismatching might stabilize unspecific TM aggregation.
Representations of a membrane protein fold

The fold of a membrane protein can be visualised by either a topology map or an X-ray/NMR structure. A topology map is a 2D-representation of the protein in the lipid bilayer (Figure 7). These maps simply define the location of TM helices and the orientation of the soluble domains relative to the membrane. They can be predicted theoretically or determined experimentally, or more commonly, by a combination of the two methods (18). Theoretical predictions are based on the identification of topology determinants in the primary amino-acid sequence, such as hydrophobic stretches, charged amino acids (according to the positive-inside rule), and aromatic amino acids (74-77). Unfortunately, not all membrane proteins have strong topology determinants and most theoretical models need to be validated by experiments. Experimental data can also be used to improve a theoretical prediction, if it is available a priori (78).

Figure 7. Representation of a membrane protein fold as a 3D-crystal structure (left) or a 2D-topology map (right top and bottom). (Reprinted from (79) with permission from AAAS.)
Experimental data on the topology of a membrane protein is usually obtained by fusing a reporter protein that has an activity that is limited to a specific compartment of the cell. The C-terminus of the protein has been shown to be an efficient fusion point for experimental topology studies using alkaline phosphatase (PhoA) and the green fluorescent protein (GFP) as reporter proteins (80,81). However, as reporter proteins are large, they can interfere with the targeting of a membrane protein if fused to the N-terminus. Thus, there has long been a need for a reporter system that can be used on the N-terminus of membrane proteins without disrupting the function of the signal sequence. In paper I, we show that the split-GFP system (82) can fulfil this task. In the system, GFP is divided into two parts. One of these parts is only 16 amino acids long, and can be fused to the N-terminus of a membrane protein without perturbing its insertion and topology. When both parts of the split-GFP are located in the cytoplasm, they will fuse and yield a detectable fluorescent signal.

A 3D representation of a membrane protein fold provides many levels of information that cannot be obtained from a 2D-topology map (Figure 7). Apart from showing the orientation of TM helices relative to the membrane, they show how TM helices are packed with each other, how TM helices sit in the membrane (i.e. perpendicular or tilted), and whether there are re-entrant loops. Furthermore, 3D-structures can show different conformational states of a membrane protein, where the 2D-topology maps would remain the same. This extra information contributes to a better understanding of the function of the membrane protein and allows for more accurate structure-function analysis. To date (March 2013) the 3D crystal structure of > 1000 proteins from E. coli have been solved. However, only ~50 of these are membrane proteins (http://www.rcsb.org/pdb/results) (http://blanco.biomol.uci.edu/mstruc/listAll/list).
Functional units in the *E. coli* inner membrane.

The available structural and biochemical data on α-helical membrane proteins indicate that most of them assemble into complexes, or functional units. The inner membrane of *E. coli* contains a number of well-characterized functional units. Notable examples include the complexes of the respiratory chain, the ATP-synthase, the ATP-binding cassette (ABC) transporters and a number of ion channels. In this thesis I have studied the assembly and stability of two of these functional units: The CorA magnesium channel (Papers III-IV) and the cytochrome bo$_3$ complex (Paper IV).

The CorA magnesium channel is a homopentamer (Figure 8). Each monomer has two TM regions, of which TM1 makes up the pore that allows magnesium permeation. The TM helices are connected by a short, and very well conserved periplasmic loop, which makes the first contact with the magnesium ion and appears to create the selectivity filter. CorA also has a large soluble N-terminus at the cytoplasmic side of the membrane. In the available crystal structures of CorA, two magnesium-binding sites were identified in the interface between the cytoplasmic domains of the protomers and the binding of magnesium in these sites is thought to stabilise the pentameric channel in a closed conformation (83-86). Taken together, all parts of the protein have an important role in function by 1) selecting for magnesium ions (periplasmic loop), 2) translocating magnesium through the membrane (TM region), and 3) regulating the influx of magnesium (cytoplasmic domain). Even though there is a lot of structural information available for CorA, little is known about how the pentamer is assembled and what parts of the protein are important for structural stability. In papers III and IV, I have investigated how the periplasmic loop contributes stability to the pentameric assembly.
Cytochrome $b_{O_3}$ is a terminal oxidase of the *E. coli* respiratory chain. It consists of four protein subunits (with a total of 25 TM-helices) and three cofactors, namely heme $b$, heme $o_3$ and Cu$_b$. Subunit I also holds all three cofactors and is the catalytic core of the enzyme (87-90) (Figure 9). The protein subunits of cytochrome $b_{O_3}$ assemble in a specific order, which is dependent on the insertion of heme cofactors (91). However it is not known when or how the heme cofactors are inserted during the assembly process. In Paper II, I have determined the point of heme incorporation during the assembly process of cytochrome $b_{O_3}$, by investigating the heme content of its assembly intermediates.
Why are $\alpha$-helical membrane proteins assembled into complexes?

Generally speaking, there are a number of advantages in assembling membrane proteins into oligomeric complexes as compared to individual proteins. Parts of the complex can be replaced in isolation if damaged (i.e. like one would change damaged parts of a car engine rather than exchanging the whole engine). For example, proteins involved in photosynthesis, mainly photosystem II, are susceptible to damage caused by UV-light, which will lead to damage and inactivation of the reaction centre D1 subunit. Upon damage, the D1 subunit will be phosphorylated and triggered for degradation by proteases. Thereafter, photosystem II will dissociate from the antenna complexes, move to another region of the thylakoid membrane where a new D1 subunit will assemble with the partially disassembled photosystem II, which will then recombine with the antenna complex (92). Another reason why functional units contain many proteins relates to complexity. Different protein modules can be added in response to a cellular requirement, and thereby provide additional regulatory control, as for example proteins involved in signalling. The Tat translocase also assembles in response to cellular cues. The protein components are assembled upon recognition of the targeting signal to the central component TatC, which is embedded in the membrane. Thereafter, TatA and TatB are recruited to form the translocon (47,93,94). Here, the assembly of the components of the Tat translocon is carried out in a regulatory fashion, i.e. when the target protein binds to TatC with its N-terminal signal sequence. This shows how assembly of membrane proteins into complexes is tightly coupled with the function of the complex.

In some instances, multiple proteins contribute to a protein fold. For example, in many channels the subunits are organised around a central pore (95) (Figure 10, left). In other instances it is not clear why the functional unit is comprised of many proteins. Aquaporins assemble into functional homo-
tetramers even though each individual monomer forms an independent channel (96) (Figure 10, right).

Figure 10. The crystal structure of the magnesium channel CorA (left) shows how the five protomers are assembled to create a central pore for translocation. Each individual monomer of AqpZ (right) forms a channel, even though they assemble into a functional tetramer. (Reprinted from (86,97) with permission from Biochemical Society and Elsevier.)

It has been postulated that the reduction of surface area that comes with larger proteins is an additional reason why larger protein complexes have evolved. The advantages of this are many, but the most important are thought to be the increased stability in solvent and protection from degradation (98). One example of a protein complex that has evolved to a larger structure is the cytochrome c oxidase, which has 13 protein subunits in mammals, compared to its bacterial homologue cytochrome bo$_3$, which has only 4 protein subunits (88,99). However, with the advantages of complexity comes a downside: Complex functional units have to be assembled.

How do α-helical membrane proteins assemble into complexes?

In many ways, assembly of α-helical membrane proteins can be viewed as an extension of folding (see previous sections discussing the three-stage model). This point is best exemplified by studies on the assembly of the
AcrB efflux pump, where individual monomers are clearly folded before being assembled into a functional homo-trimeric unit (100). However, there are exceptions to this theory. In the case of the ABC transporter BtuCD, assembly clearly precedes folding of individual subunits (101). Even though folding of membrane proteins has been studied extensively, their oligomerisation into complexes has not, which is why our knowledge regarding this event is limited. In the following sections I will describe what is known about the assembly of functional units in the inner membrane of *E. coli*.

Assembly of proteins is often an ordered event, following steps or reactions in one direction (102). This point is supported by the assembly of the Tat translocon (see above), cytochrome *bo*₃ and the *E. coli* divisome, which all assemble in a sequential manner (91,103,104).

The dynamic and transient interactions between α-helical membrane proteins is further exemplified by the discovery of a large number of peripheral membrane proteins (PIMs) that bind reversibly to membrane protein complexes (20). Sequential assembly patterns of a membrane protein complex can therefore regulate function. Membrane protein complexes can also assemble into super-complexes, which is the assembly of functional units involved in the same biological process. A well-characterised super-complex consists of proteins involved in oxidative phosphorylation, called a respirasome (105). This super-complex is thought to promote the efficient transfer of electrons between the membrane bound complexes and minimise the occurrence of electron leakage and reactive oxygen species. In *E. coli*, a super-complex comprised of formate dehydrogenase and cytochrome *bo*₃, both involved in the respiratory electron transport, has been discovered (106). Moreover, there are super-complexes involved in other biological processes, e.g. *F₀F₁*-ATPase, the *K*⁺-transporting system Trk and formate hydrogen lyase, which have been shown to work together in a super-complex during anaerobic conditions as an ion-exchanging *H*⁺/*K*⁺-pump (107).
Protein-lipid interactions have an impact on the assembly of membrane protein complexes.

The hydrophobic environment in the lipid bilayer also influences the assembly of membrane proteins into complexes. The presence of unfolded and small proteins in the membrane requires an organisation of lipids that need to be ordered around the protein surface. The ordering of lipids will reduce the entropy in the membrane. However, when two or more proteins oligomerise, there is a gain of free energy due to the increasing entropy of the lipids, making oligomerisation energetically favourable (96). Furthermore, hydrophobic mismatch between the TM region and the lipid bilayer may also have an impact on oligomerisation, folding and conformational changes, due to self-association/aggregation of proteins. The increase in free energy due to a protein-lipid mismatch can be lowered by eliminating protein-lipid contacts and replace them by protein-protein interactions (discussed in previous sections and in (73,108)).

Cofactor insertion has been shown to be important for the assembly of complexes.

Many protein complexes have cofactors that are required both for function and stability (109-111). The E. coli succinate:ubiquinone oxidoreductase (SQR) contains a heme molecule within its membrane domain that is not essential for the function of the complex. However, a heme-lacking variant of the complex is destabilised in detergent, suggesting a structural role of the heme (112). In paper II we show that heme incorporation into cytochrome bo3 is required for the assembly of the complex and that assembly of protein subunits is stalled if the heme cofactors are unable to bind. Taken together, the oligomerisation of membrane proteins into functional units is influenced by forces made by protein interactions, the lipid environment, cofactor insertion and conformational changes.
How do we detect α-helical membrane protein complexes?

Two-hybrid based approaches

Many studies attempting to decipher the interactome of proteins have been carried out using the yeast two-hybrid system. It was developed over 20 years ago and the first system allowing for studies of protein-protein interactions in vivo (113). In this system the protein GAL4 from Saccharomyces cerevisiae is split and fused to putative protein-protein partners and upon their interaction GAL4 is activated. The yeast-two hybrid system is based on the targeting of proteins to the nucleus and is therefore not suited to study interactions between α-helical membrane proteins, however it set the scene for the development of other complementation systems.

The complementation techniques better suited for membrane proteins are for example split-GFP and split-ubiquitin and the GALLEX system (114-117). When designing experiments using these systems, the topology of the inserted target protein has to be already determined. The GALLEX system was developed to follow the homo- and heterodimerisation of E. coli inner membrane proteins, by coupling each protein of interest to a split LexA. Dimerisation of the proteins of interest results in dimerisation of the LexA TM domains, leading to the repression of β-galactosidase.

The drawback of the complementation techniques presented above is that they are limited to study pairwise interactions. Even though, most protein complexes are dimeric, there has been a need to find methods enabling studies of higher oligomeric membrane protein complexes.
**Size Exclusion Chromatography**

SEC or gel filtration separates macromolecules according to their hydrodynamic volume. The proteins travel through a mesh, usually consisting of polymer beads having pores of different sizes. As the proteins travel through the column, small proteins travel longer through the column since they more often enter the pores of the beads compared to larger proteins. As the proteins exit the column, they are detected spectrophotometrically by absorbance reading, generating an elution profile shown as a graph. Unfortunately, not all membrane protein complexes are stable in detergent and there might be more than one oligomeric species present in the same sample. The graph will provide information about the occurrence of different oligomeric complexes and if the protein is suitable for x-ray crystallisation (determined by the monodispersity of the curve of interest, discussed in (118,119)). It is hard to determine the exact size of an oligomeric membrane protein complex by SEC, since a small difference in size would not make a big difference in the elution volume. Furthermore, the size of the detergent micelle needs to be taken into account, which is not an easy task (118).

**BN-PAGE**

An alternative method to determine the size and protein content of a membrane protein complex, which does not require a high protein concentration, is the use of native gel-based systems. Blue Native Polyacrylamide Gel Electrophoresis (BN-PAGE) was developed by Schägger and von Jagow more than 20 years ago and was applied to the analysis of cytochrome bc/bf complexes in mitochondria and chloroplasts respectively. It enabled the separation of both acidic and basic water-soluble proteins ranging from 10 kDa to 10 000 kDa at a fixed pH of 7.5, but most importantly, it allowed for analysis of membrane protein complexes without engineering tags. In this method the protein samples are solubilised by mild detergents. Not all proteins have the same detergent preferences for BN-PAGE, but the most commonly
used are n-dodecyl-ß-D-maltoside (DDM), digitonin and Triton X-100) (120). When the protein complexes are embedded in the detergent micelles, their net charge is neutral, which is not suitable for gel electrophoresis. Therefore, the negativ molecule Coomassie (G250) is added to the sample before running it on the gel, which will give the complexes a net negative charge. The cathode buffer that is running through the gel during electrophoresis also contains Coomassie and the negative molecules will substitute the detergent that is surrounding the hydrophobic membrane protein complex as the electrophoresis is taking place, which is why BN-PAGE is denoted a charge-shift method (121). BN-PAGE can be combined with a second dimension sodium dodecylsulfate electrophoresis (SDS-PAGE), which provides an analytical method for deciphering the content of each membrane protein complex. Furthermore, the molecular mass and oligomeric state of non-dissociated complexes can be determined, as well as subunit composition and detection of subcomplexes (121,122). In Paper III and IV, I have used BN-PAGE to analyse the oligomeric state of the magnesium channel CorA from T. maritima. By introducing different substitutions in the periplasmic loop, and thereafter analysing their oligomeric state, we were able to assign structural roles for certain residues in this very well conserved region. A very important advantage of BN-PAGE is that the proteins will remain in their native state during electrophoresis, as compared with SDS-PAGE where cofactors, such as heme, will dissociate during the run. In Paper II, I took advantage of this by trapping assembly intermediates of cytochrome bo₃ and analysing both oligomeric states and heme-content by BN-PAGE.
Objectives of thesis

Most biological functions are carried out by protein complexes (102). This statement is supported by the fact >80% of the conserved and essential proteins of *E. coli* have been shown to assemble with at least one protein (123). These complexes can be either homo-oligomeric (*i.e.* many copies of one protein), hetero-oligomeric (*i.e.* two or more different types of proteins interact). In *E. coli*, 79% of the protein complexes are homo-oligomeric whereas 21% are hetero-oligomeric (98). In certain cases the protein complex may also require a metal cofactor(s) in order to function, or to stabilise the structure. The assembly of proteins and cofactors generates biologically functional units.

To date research on membrane protein biogenesis has mainly focused on the targeting of membrane proteins, as well as the insertion and folding of individual TM helices into the lipid bilayer. Surprisingly, there is an inadequate amount of knowledge of the process between the insertion and folding of TM regions and the assembly of membrane proteins into complexes. During my thesis work, I have aimed to better understand the latter process, by studying how membrane proteins and their cofactors assemble into functional units.
Short summary of papers

Paper I

Application of split-green fluorescent protein for topology mapping membrane proteins in *Escherichia coli*.

Stephen Toddo, Bill Söderström, Isolde Palombo, Gunnar von Heijne, Morten H. Nørholm and Daniel O. Daley

Protein Science (2012) **10**, 1571-1576

Here we focused our attention on the early steps of membrane protein folding. Even though all membrane proteins that are inserted cotranslationally enter the SecYEG with their N-terminus, the α-helices are able to flip during insertion in order to attain the correct topology (74,124). The final topology of a membrane protein can be predicted theoretically by analyzing the amino-acid sequence, but the results need to be verified experimentally (18). Available reporter proteins can only be used on the C-terminus of membrane proteins since fusion to the N-terminus could disrupt targeting and insertion. In Paper I, we show that the split-GFP system (82) can be used to determine the location of both the C-terminus and N-terminus of a membrane protein, since one of the parts is only 16 amino acids long. In the study three membrane proteins with known topology were chosen, ZipA, PpiD and YfgM. The N- or the C-terminus of these proteins was fused to the 16 amino-acid peptide of GFP (called [11H7]) and the remainder of GFP ([1-10OPT]) was co-expressed in the cytoplasm. The location of 11H7 and,
thus, the topology of the membrane protein could then be determined from the fluorescence profile.

Paper II

Heme incorporation into the cytochrome $bo_3$ occurs at a late stage of assembly.

Isolde Palombo and Daniel O. Daley


Cytochrome $bo_3$ of the respiratory chain of *E. coli* belongs to the superfamily of heme-copper respiratory oxidases. It is composed of four different protein subunits and three cofactors (heme b, heme $a_3$ and CuII) that are all buried in the hydrophobic core of subunit I (88). It has been shown that the four protein subunits of cytochrome $bo_3$ follow a specific order in their assembly process (91). However, it is not known when or how the heme molecules are inserted. Due to the hydrophobic nature of the hemes and the fact that they are buried deep within the cytochrome $bo_3$ complex, it has mostly been speculated that the insertion occurs at an early stage of assembly (125). In Paper II we expressed different assembly intermediates of the cytochrome $bo_3$ (subunit I monomer, I-III-IV-intermediate and cytochrome $bo_3$ complex) and followed the incorporation of $^{55}$Fe labelled heme into these intermediates. We observed full incorporation of heme in the I-III-IV intermediate, but only very low levels in the monomer and therefore concluded that the incorporation of heme occurs at a later step in the assembly process than anticipated. Our data also indicate that the association of cofactors into subunit I is a requirement for the assembly of the functional unit. Our data allow us to
propose a model for the assembly pathway for both proteins and hemes of cytochrome $bo_3$ (Figure 11).

![Figure 11. The assembly pathway of protein subunits and cofactors in cytochrome $bo_3$. (Reprinted from Paper II with permission from Elsevier).](image)

Ordered assembly and late heme incorporation is probably a way to protect the cell from harmful reactive oxygen species that could be an effect of a partly assembled enzyme with reactive cofactors. Additionally, a heme-delivery protein must interact with the cytochrome $bo_3$ and/or the assembly intermediate (I-III-IV) at some point during biogenesis. By determining the timing of heme insertion we have narrowed down the alternatives for when a heme-delivery protein would interact with the protein, which will be helpful in future studies.

Paper III

The periplasmic loops provide stability to the open state of the CorA magnesium channel

Isolde Palombo, Daniel O. Daley and Mikaela Rapp


Ion channels are membrane embedded protein structures that regulate the influx/efflux of ions and transport them according to the cell’s needs. To do
so they must adopt different conformations, but it has proven difficult to capture these different conformational states in structural studies (95). The prokaryotic magnesium ion channel CorA is one of the most studied ion channels. In 2006 three crystal structures of the channel from *T. maritima* (*TmCorA*) were solved (83,84,126). They were all solved in the presence of high concentration of Mg\(^{2+}\) (100 mM), and two Mg\(^{2+}\) binding sites in the interface between each protomer, termed M1 and M2 could be identified (described previously and illustrated in Figure 8). The binding of Mg\(^{2+}\) was thought to stabilise the pentameric channel in a closed conformation. The periplasmic loop, connecting the two TM helices of each monomer, is the most conserved part of the protein and it remained unresolved in all three structures. The loop has been shown to be essential for function and thought to make up the selectivity filter for Mg\(^{2+}\) as well as the entry point of the channel (127,128). In one study it was shown that two of the conserved residues of the loop had an impact on pentamer stability (129). This finding could not be explained by the 3D-structures where the structural stability of the pentamer appeared to be mediated by metal coordination to the cytoplasmic M1/M2 sites. We decided to take a closer look at this latter finding by investigating the role of each loop residue in pentamer stability. We determined, by BN-PAGE, SEC and chemical cross-linking that eight residues of the loop were essential for pentamer stability in the absence of Mg\(^{2+}\). Seven out of these eight protein variants were also shown to be non-functional *in vivo*. Since the crystal structures were all resolved in 100 mM Mg\(^{2+}\), and this was suggested to stabilise the pentamer in a closed conformation, we wanted to investigate if the addition of Mg\(^{2+}\) would have any impact on our mutated channels. We therefore added 100 mM of Mg\(^{2+}\) to our protein variant and reanalysed their oligomeric state by SEC. Expectedly, all destabilised variants shifted from lower oligomers to pentamers by the addition of Mg\(^{2+}\), suggesting that the binding of magnesium indeed stabilised the pentameric structure by other interactions than those mediated by the periplasmic loop.
We could furthermore determine that this shift was reversible and brought about by the binding of Mg$^{2+}$ to the cytoplasmic M1/M2 sites. Combining the results, we could draw the conclusion that the interactions mediated by the periplasmic loop are structurally important in the absence of Mg$^{2+}$, thus, in a putative open conformation of TmCorA channel.

Paper IV

Why is the GMN-motif in the CorA/Mrs2/Alr1 superfamily of magnesium channels conserved?

Isolde Palombo, Daniel O. Daley and Mikaela Rapp

Submitted

The short periplasmic loop of CorA magnesium channels carries the family signature motif, GMN, which is universally conserved in all phyla and essential for protein function (127,128). The crystal structure of CorA from Methanocaldococcus jannaschii and T. maritima are solved in the presence of Mg$^{2+}$ and show that the periplasmic loops are arranged to form a concavity at the pore entrance. At the bottom of the concavity the asparagines (N) of the GMN triplets form a five-fold N-ring that, together with the main-chain oxygen of the adjacent glycines (G), is believed to act as the selectivity filter for partially hydrated Mg$^{2+}$ (84-86,130). The methionine (M) side chain of the GMN triplet appears to contribute to a hydrophobic core that stabilizes the GN-selectivity filter. Thus the crystal structures imply that the GMN motif plays a central role in Mg$^{2+}$ selectivity. What is not clear from these structures (or from any experimentation) is why the GMN motif has been so well conserved in the CorA/Mrs2/Alr1 super family. In Paper III we discovered that single alanine substitutions at several positions in the conserved loop-motif of TmCorA disrupted the pentameric state when Mg$^{2+}$ was absent
from the cytoplasmic metal binding sites. The interactions provided by these residues probably played a prominent structural role in supporting the open conformation. In Paper IV we were able to identify the chemical constraints of these interactions by identifying residues with similar chemical properties that could mediate the same interactions. However, the conserved GMN motif was restricted by both functional and structural constraints. These constraints provide a plausible explanation for why the periplasmic GMN has been conserved throughout the CorA/Mrs2/Alr1 super family.

These studies show that the difference between structure and function might not always be easy to distinguish. It has been shown that $Tm$CorA is also able to transport other divalent cations (131) and it would be interesting to see if we could change specificity for any of these cations by changing the N-ring to a structurally stable Q-ring. If so, it would deepen our understanding on how the selectivity filter in the periplasmic loop actually works.
Concluding remarks

Membrane proteins are key players in many biological processes and are hotly pursued targets for both drugs and antibiotics. The process of finding drugs to target these proteins requires that we understand every step in their life; i) how they are synthesised, ii) how they are inserted into the membrane, iii) how they are folded, iv) how they are assembled into functional complexes and v) how they function. In this thesis, I have studied these processes for a number of different proteins. These studies provide information about critical events in the life of the proteins.

In Paper I, we developed a method that enables us to better understand the orientation of individual membrane proteins (their topology). Experimental topology studies have been limited by the fact that available reporter proteins can only be used on the C-terminus of proteins. Our studies showed that the split-GFP system can be used at both the C- and N-terminus. This system is currently being used to determine the topology of membrane proteins with unknown function, which will greatly facilitate future studies.

In Papers II, III and IV we studied the folding and assembly of membrane proteins into functional units. Most of what is currently known about this process is based on interactions between isolated transmembrane helices and does not provide the full picture of how proteins interact with each other and with cofactors. Our studies on the assembly of the CorA magnesium channel from T. maritima identified interactions required to stabilise the putative open state. These interactions were not captured in any of the available crystal structures. Thus these studies resulted in a better understanding of the transport cycle of CorA. Our studies on the cytochrome bo₃ showed that it is assembled in a stepwise manner, through a set of non-functional intermedi-
ates. This sequence of events most probably protects the cell from harmful intermediates. These findings have immediate implications for the human cytochrome c oxidase, which is also within the family of heme-copper oxidases.

As humans are shaped into mature individuals by their social interactions, membrane proteins are shaped by interactions with other proteins and cofactors to form mature functional units. In this thesis we have analysed and identified a number of interactions that are important in the life of a membrane protein. These interactions are not easily predicted and do not follow a consensus pathway, which is why the assembly of membrane proteins is a complex story.

Jag har under min doktorandtid studerat hur membranprotein sätts samman till större fungerande enheter i *Escherichia coli*. *E. coli* är en encellig bakterie som främst återfinns i våra tarmsystem och har använts i över hundra år som modellorganism i forskning för att öka vår kunskap om hur liv är uppbyggt. Mina resultat visar att det är vissa specifika interaktioner mellan de olika membranproteinerna i komplexet som är viktigare än andra för att enzymet ska fungera. Jag har också studerat effekten av metaller och andra
metallinnehållande molekyler och deras betydelse i att stabilisera membranproteiner när de sätts samman till större komplex.

För att förstå hur membranprotein fungerar är det viktigt att man har kunskap om varje steg i deras liv, från att de syntetiseras till att de blir mogna och funktionella enheter. Precis som vi människor har membranproteiner ett socialt liv där de interagerar med andra membranproteiner. Jag har i min avhandling studerat detta sociala liv för få ökad kunskap och förståelse om hur den slutgiltiga funktionella enheten fungerar.
Acknowledgements

I have in my thesis shown that the social life of a membrane protein is important to yield a functional unit. This conclusion is applicable to my PhD studies as well. Without all the people I have interacted with along the way, I would not have been able to mature as a person and as a scientist. So thanks to all of you, with a special thanks to;

Daniel Daley, you have been an excellent supervisor by always sharing your knowledge and interest in science with me, and the rest of the group. I really appreciate all the support you have given me and that your door has always been open, both for laughter and tears.

Gunnar von Heijne, my co-supervisor and the one who took me in at DBB and inspired me to stay for a PhD together with Susanna Seppälä, you are really an inspiration and working with you was so much fun.

My siblings in the DD-family: Stephen Toddo, for teaching me about training and eating and your funny (sometimes too quiet) jokes. Jörg Götzke for intense work and lots of fun working together in the lab and for presenting the daily surprise at lunch (the sandwiches are always the same, but the desserts change). Claudio Muheim for bringing new energy to the lab and for adopting my baby (the cyo-project). Our stepbrother Bill, for always being cheerful and eager to learn. Yet, you never learned to have lunch at normal times. Former members Kalle Skoog, for endless discussions (or fights) about everything, I have learned a lot from you. Filippa Bruzell, I miss your words of wisdom and your ability to motivate. Your calmness in the lab
should be an inspiration to everyone. Gianluca, for being more stressed and loud than me and for your positive energy.

Erika, my lovely wife. We have shared almost everyday (and every lunch) for nine years together, need I say more? I could not have survived these years without you (and all the broccoli and tuna). To the rest of the EG-group, Elzbieta, Catarina (Miss Sunshine), Beata, Pedro and Hasse; thank you for adopting me to your group and for all the fika. Former members and present friends Nyosha and Anna-Karin, we had so much fun at the department, on the ranch in Katy, Texas and in NYC.

The dynamic GvH cluster, so many good people have come and gone and contributed to the great energy and atmosphere. Joy, Marie, Salomé, Roger, Johannes, Morten, Carmen, Nurzian, Joanna, Jan-Willem, David, Mirjam, Susan, Anna, Thomas, Karin, Florian, Patricia, Nina, Rickard, Ola, Rob, Diogo, Johan, Dan and finally Mikaela, you got me in at DBB before you left, resulting in a fun and inspiring collaboration that I hope will continue.

Åke Wieslander, for being such a charismatic and generous person. You know everything there is to know about everything. Something I really admire.

PB, PÅ, MO, MO2, MH, LM, AN and IC. All these letters represent even more people. You all contribute to the fantastic (and famous) environment at DBB.

Ann, Maria, Lotta, Heidi och alla ni andra som jobbar och har jobbat så hårt för att institutionen ska fungera. Ni har ett imponerande tålamod med alla oss förvirrade forskare som ställer samma dumma frågor gång på gång på gång…. 

Stefan Nordlund, you (are a) rock! I told you during my first biochemistry course (2005!) that I was aiming for a PhD in biochemistry. Your compelling introduction to the cell and its metabolism is what convinced me. Thank you for saving me from organic chemistry.

Min familj och mina vänner, även om ni fortfarande inte helt förstår vad jag har gjort de senaste åren så har jag inte kunnat göra det utan er.

John, för att du finns och för att du påminner mig om att jag kan göra allt jag vill, när jag vill.
References

71. van Klompenburg, W., Nilsson, I., von Heijne, G., and de Kruijff, B. (1997) EMBO J 16(14), 4261-4266
76. von Heijne, G. (1986) EMBO J 5, 3021-3027