Protein production in the *E. coli* cell envelope

Thomas Baumgarten

Abstract

Proteins fulfill essential functions in every cell and malfunctioning proteins are often the cause of diseases. On the other hand, proteins like antibody fragments or hormones can be used to treat diseases. Proteins are often produced in the bacterium *Escherichia coli* so that they can be studied to understand their (mal)function or so that they can be used to treat a disease. Unfortunately, producing proteins in the cell envelope of *E. coli*, like integral membrane proteins, which are important drug targets, and secretory proteins like antibody fragments and hormones, often results in unsatisfactory yields. Therefore, the objectives of this doctoral thesis were to identify bottlenecks that can limit the production of recombinant proteins in the cell envelope of *E. coli* and to try to overcome these bottlenecks. In the first study, we isolated and characterized the *E. coli* membrane protein production strain Mt56(DE3). This strain, in which the target gene expression intensity is strongly reduced, outcompetes the standard *E. coli* membrane protein production strains for most targets tested. In the second and third study we focused on the production of secretory proteins, i.e., proteins that are translocated across the inner membrane into the periplasm of *E. coli*. First, we investigated the impact of the targeting pathway used to direct a secretory protein to the translocation machinery on the cell physiology and protein production yields. We found that the co-translational targeting of a produced protein saturates the capacity of the translocation machinery resulting in heavily impaired biomass formation and low protein production yields. In contrast, post-translational targeting of a produced protein did not saturate the capacity of the protein translocation machinery resulting in hardly affected biomass formation and high protein production yields. In the third study we investigated how optimizing the production of a co-translationally targeted protein, by harmonizing its production rate with the capacity of the protein translocation machinery, affects the physiology of the cell. We found that, in stark contrast to the non-optimized condition, the optimized production did not affect the composition of the *E. coli* proteome. This surprising finding indicates that a protein can be produced efficiently in the periplasm of *E. coli* without compromising the physiology of the cell. In the last study we aimed at developing an outer membrane vesicle-based tuberculosis vaccine. To this end, an *E. coli* strain was created that produced outer membrane vesicles coated with different tuberculosis antigens. It was shown that a homogenous population of vesicles was produced, which will hopefully facilitate the isolation of these vesicles on an industrial scale.

Keywords: *E. coli*, protein biogenesis, recombinant proteins, membrane proteins, secretory proteins, protein displays, outer membrane vesicles.
PROTEIN PRODUCTION IN THE E. COLI CELL ENVELOPE

Thomas Baumgarten
Protein production in the *E. coli* cell envelope

Thomas Baumgarten
List of papers included in the thesis

I.  Isolation and characterization of the E. coli membrane protein production strain Mutant56(DE3)
    Sci Rep 2017; 7:45089

II. Post-translational targeting of a recombinant protein promotes its efficient secretion into the E. coli periplasm
    Ytterberg JA, Zubarev RA, Baumgarten T
    Manuscript in preparation

III. Optimizing Recombinant Protein Production in the Escherichia coli Periplasm Alleviates Stress
    Baumgarten T, Ytterberg JA, Zubarev RA, de Gier JW
    Appl Environ Microbiol 2018; 84(12):e00270-18

IV. Decoration of Outer Membrane Vesicles with Multiple Antigens by Using an Autotransporter Approach
    Appl Environ Microbiol 2014; 80(18):5854-65
Additional publications

High-level production of membrane proteins in *E. coli* BL21(DE3) by omitting the inducer IPTG
Zhang Z, Kuipers G, Niemiec Ł, Baumgarten T, Slotboom DJ, de Gier JW, Hjelm A
Microb Cell Fact 2015; 14:142

Bacterial-based membrane protein production
Schlegel S, Hjelm A, Baumgarten T, Vickström D, de Gier JW
Biochim Biophys Acta 2014; 1843(8):1739-1749

Optimizing *E. coli*-Based Membrane Protein Production Using Lemo21(DE3) and GFP-Fusions
Hjelm A, Schlegel S, Baumgarten T, Klepsch M, Wickström D, Drew D, de Gier JW
Methods Mol Biol 2013; 1033:381-400
Contents

List of papers included in the thesis .................................................. iv

Additional publications ................................................................. v

1 Recombinant DNA technology: How it got started ..................... 1

2 E. coli as a host to produce recombinant proteins ................. 4

3 The E. coli cytoplasm .............................................................. 7
   3.1 Protein synthesis and folding in the cytoplasm ....................... 8
   3.2 Protein degradation in the cytoplasm ..................................... 13
   3.3 Targeting of proteins to the cell envelope ............................. 14

4 The E. coli cell envelope ......................................................... 16
   4.1 The inner membrane .......................................................... 17
      4.1.1 Biogenesis of inner membrane proteins .......................... 17
      4.1.2 Degradation of inner membrane proteins ....................... 22
   4.2 The periplasm ...................................................................... 23
      4.2.1 Biogenesis of secretory proteins .................................... 25
      4.2.2 Protein degradation in the periplasm ............................... 29
   4.3 The outer membrane .......................................................... 30
      4.3.1 Biogenesis of outer membrane proteins ............................ 31
5 Recombinant protein production in the *E. coli* cell envelope .... 33

5.1 Producing recombinant proteins in *E. coli* ........................ 33
5.2 Modifying the gene encoding a recombinant protein .......... 38
5.3 Improving inner membrane protein production yields ........ 39
  5.3.1 Directing the evolution of *E. coli* ................................. 40
  5.3.2 Engineering *E. coli* .................................................. 45
5.4 Improving the production of proteins in the periplasm .......... 51
  5.4.1 Improving the targeting of secretory proteins ................. 51
  5.4.2 Improving the yield and quality of secreted proteins ....... 55
  5.4.3 Improving the stability of secreted proteins .................... 59
5.5 Display of recombinant proteins on the cell surface .......... 60
5.6 Export of recombinant proteins into the culture medium ...... 62

Conclusions and future perspectives .................................... 65

Populärvetenskaplig sammanfattning .................................. 69

Acknowledgements ................................................................ 71

References ........................................................................... 73
1 Recombinant DNA technology: How it got started

Recombinant DNA technology enables scientists and engineers to use bacteria as well as other organisms as “factories” to produce valuable products from inexpensive starting materials. Although living organisms have been used since a few thousand years to produce e.g., bread, alcoholic drinks and vinegar, developing recombinant DNA technology less than 50 years ago created the basis for modern biotechnology. In 1974, Stanford University filed a patent application based on three publications from Stanley N. Cohen and Herbert W. Boyer [1]. They described that DNA can be manipulated in vitro, subsequently replicated in the bacterium *Escherichia coli* and used to synthesize functional proteins in *E. coli* that originate from other organisms [2–4]. The full potential of this new technology was recognized immediately. However, also concerns about the risk of biohazards were raised, stalling the process of patenting the recombinant DNA technology. The assessment of benefits and risks of this groundbreaking technology included also issuing new regulations. Moreover, the patent process was further delayed until the US Supreme Court had decided that an organism can actually be patented. Therefore, the patent could only be granted as late as in 1980. However, one of the inventors of the recombinant DNA technology, Herbert W. Boyer, did not wait until the patent issue was resolved. He cofounded the company Genentech already in 1976 to commercialize this technology. Being the first company using recombinant DNA
technology, Genentech reported already in 1977 the production of the human peptide hormone somasostatin using *E. coli* as “cell factory”. One year later Genentech was able to use *E. coli* to produce human insulin, which became the first recombinantly produced therapeutic approved in US and parts of Europe in 1982 [5].

Nowadays, recombinant DNA technology is routinely used in academia and industry, and recombinant products are part of our everyday life (see Table 1).

**Table 1. Examples of products produced using recombinant DNA technology.** All listed products are produced in genetically modified production hosts and they cover a wide range of applications. Year: product approved by the US Food and Drug Administration.

<table>
<thead>
<tr>
<th>Product</th>
<th>Production Host</th>
<th>Application</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human insulin</td>
<td><em>E. coli</em></td>
<td>treat diabetes</td>
<td>1982</td>
</tr>
<tr>
<td>Human growth hormone</td>
<td><em>E. coli</em></td>
<td>treat growth disorders</td>
<td>1985</td>
</tr>
<tr>
<td>Hepatitis B antigen</td>
<td><em>Saccharomyces cerevisiae</em></td>
<td>used as vaccine</td>
<td>1986</td>
</tr>
<tr>
<td>Hydrolases</td>
<td><em>E. coli</em></td>
<td>removal of stains</td>
<td>1988</td>
</tr>
<tr>
<td>Chymosin</td>
<td><em>E. coli</em></td>
<td>cheese production</td>
<td>1990</td>
</tr>
<tr>
<td>Blood clotting factor VIII</td>
<td>CHO cells</td>
<td>treat bleeding disorder</td>
<td>1992</td>
</tr>
<tr>
<td>Herbicide resistant crops</td>
<td>Soy</td>
<td>weed control</td>
<td>1996</td>
</tr>
<tr>
<td>Amylases</td>
<td><em>Bacillus licheniformis</em></td>
<td>starch processing</td>
<td>1999</td>
</tr>
<tr>
<td>Trehalase</td>
<td><em>Aspergillus niger</em></td>
<td>ethanol production</td>
<td>2017*</td>
</tr>
</tbody>
</table>

*application filed to use Trehalase in the production process of ethanol
Alone in the pharmaceutical industry more than 400 recombinant protein-based therapeutics have been approved and by 2015 another 1,300 recombinant drugs were tested in clinical trials to treat e.g., diabetes, multiple sclerosis, cystic fibrosis and cancer [6]. Nevertheless, we have just made a first step towards exploring the full potential of implementing recombinant DNA technology in academia and industry. Current challenges include shortening engineering periods to modify the production host and enhancing low yields of many valuable recombinant proteins.

The aim of my Ph.D. thesis was to study the production of recombinant proteins in the most commonly used bacterial host, E. coli, and to develop strategies to increase recombinant protein production yields in this "cell factory". In particular, I focused on the production of recombinant proteins in the cell envelope of E. coli, because improving the yields of proteins in this compartment is highly desirable but particularly difficult. First, I will give an overview of E. coli as a protein production host and describe protein biogenesis and protein quality control in this bacterium. Then, the results of my research efforts will be discussed with regards to strategies used by others to improve the production of recombinant proteins in the E. coli cell envelope.
2  *E. coli* as a host to produce recombinant proteins

*Escherichia coli* is a rod shaped bacterium with a length of about 2 µm and a diameter of about 1 µm. It is mainly found in the intestines of higher animals. The *E. coli* cytoplasm is separated from the environment by a complex cell envelope, which consists of an inner and an outer membrane and the space between them, which is called the periplasm (Figure 1) [7].

![Figure 1. Schematic representation of the Gram-negative bacterium *Escherichia coli*. The *E. coli* cytoplasm contains the chromosome and the protein synthesis machinery and it is enclosed by the cell envelope, which is composed of the inner membrane, the periplasm and the outer membrane.](image)

Because *E. coli* was one of the first model organisms used to study genetics and cell physiology, this bacterium was evidently also the “host of choice” to produce the first recombinant proteins. Today *E. coli* is appreciated as protein production host due to its fast growth,
low cultivation costs and high success rates when attempting to produce recombinant proteins [8]. Furthermore, *E. coli* is generally considered to be safe and comprehensive knowledge of its genome and physiology is available. This knowledge enables scientists to precisely modify the *E. coli* genome thereby allowing e.g., to control the expression of particular genes and to introduce particular genes into or delete particular genes from the genome.

The simplicity of *E. coli* comes along with some limitations [9]. Disulfide bonds, which are important for the folding of many recombinant proteins, cannot be formed in the cytoplasm of *E. coli*, but disulfide bond formation can be catalysed in the periplasm. Furthermore, *E. coli* is unable to e.g., glycosylate and phosphorylate proteins, but these modifications can be crucial for the proper folding of eukaryotic proteins. In addition, recombinant proteins produced in *E. coli* often form insoluble, non-functional aggregates. This can be beneficial if the recombinant protein is toxic or if aggregate formation facilitates the isolation of an easy to refold recombinant protein. However, the re-folding process is elaborate and often the isolation of soluble, functional recombinant protein is preferred. Another inherent problem using *E. coli* as protein production host is that it contains endotoxins, which can cause a strong immune response. Therefore, endotoxins must be removed from recombinant protein preparations, especially if the product is designated for therapy. Nevertheless, in most cases *E. coli* is still the organism of choice to produce recombinant proteins and strains have been isolated and engineered to overcome many of its limitations.

To produce a recombinant protein in *E. coli*, often the gene encoding the recombinant protein is introduced into the cell by using plasmids as carrier molecules (see 5.1). Notably, the recombinant gene contains all information required to produce a functional protein in the
cytoplasm or the cell envelope of *E. coli* (see 5). However, irrespective of its destination, the biogenesis of every protein begins in the cytoplasm.
3 The *E. coli* cytoplasm

The cytoplasm of *E. coli* is a crowded environment that contains chromosomal DNA, which encodes the information required to synthesize all cellular proteins. Most proteins need to attain a certain three-dimensional structure to fulfill their function and the protein folding process is often assisted by helper proteins called chaperones (see 3.1) (Figure 2). On the other hand, the degradation of proteins regulates their accumulation levels and clears misfolded/aggregated proteins from the cytoplasm (see 3.2). Furthermore, the cytoplasm contains proteins that target cell envelope proteins to the inner membrane, thereby assisting their biogenesis (see 3.3).

Although this thesis focuses on the production of recombinant proteins in the cell envelope, it should be noted that the synthesis of these proteins and their targeting to the cell envelope takes place in the cytoplasm. Finally, knowledge of protein biogenesis, protein degradation and protein targeting is crucial to understand strategies used to improve the production of recombinant proteins in the cell envelope of *E. coli* (see 5).
Figure 2. Protein biogenesis and degradation in the *E. coli* cytoplasm. In *E. coli* all proteins are synthesized in the cytoplasm by ribosomes. Already during their synthesis most proteins interact with trigger factor (TF), which assists protein folding. Membrane and secretory proteins are targeted either co- or post-translationally to the inner membrane (see 4.1.1 and 4.2.1). In the cytoplasm, the folding of proteins can be assisted by *e.g.*, the ATP-dependent chaperone systems DnaKJ-GrpE and GroEL/ES. IbpA/B bind protein aggregates, thereby preventing further protein aggregation and facilitating the refolding of misfolded proteins. Not properly folded proteins can be unfolded by the sequential action of the DnaKJ-GrpE system and the unfoldase ClpB. Unfolded proteins can be either refolded or degraded. Proteases, like ClpXP, ClpAP, HslUV, Lon or the membrane associated protease FtsH, can unfold misfolded proteins in an ATP-dependent manner and subsequently degrade them (see 3.2 and 4.1.2).

3.1 Protein synthesis and folding in the cytoplasm

In all domains of life, proteins are synthesized by ribosomes, which are huge complexes of RNAs and proteins. Ribosomes catalyse the specific connection of particular amino acids based on a specific messenger (m)RNA sequence using transfer (t)RNA molecules as adapters [10]. A tRNA adapter carries a particular amino acid and only if the tRNA matches the mRNA, meaning a particular tRNA se-
sequence is complementary to a particular mRNA sequence, the amino acid is added to the nascent polypeptide chain. In this way ribosomes translate genetic information into a protein. Although translation is well studied biochemically, only recently progress has been made to actually visualize this process using cryo-electron microscopy [11–13]. During or immediately after their synthesis, proteins can either interact with chaperones facilitating protein folding or are targeted to their designated compartment (see 3.3).

In *E. coli*, the chaperone trigger factor (TF) interacts with nascent polypeptide chains at the translating ribosome [14]. Although not essential, TF is highly abundant and well-characterized. TF consists of three domains that form a cavity: the N-terminal ribosome binding domain, which is connected to the peptidy-prolyl cis/trans isomerase (PPIase) domain via a long linker sequence and the C-terminal domain [15,16]. Although all three domains exhibit chaperone activity, the PPIase domain is not crucial for the function of TF and the main chaperone activity was assigned to the C-terminal domain [17]. To bind substrates, TF uses its entire inner cavity, which consists of hydrophobic and polar residues allowing a variety of interactions of TF with a substrate [18]. Additionally, TF is a rather flexible protein, potentially allowing it to adapt its conformation based on the protein substrate bound [19]. This structural flexibility together with the versatile binding cavity may enable TF to assist the folding of most *E. coli* proteins and it was estimated that around two thirds of all proteins released from the ribosome-TF complex obtain their functional structure without interacting with other chaperones [19]. However, proteins that do not fold properly after interacting with TF require the help of other chaperone systems present in the cytoplasm in order to attain their correct three dimensional structure.
One of these chaperone systems is the non-essential DnaKJ-GrpE system. DnaK is an extensively studied, ATP dependent chaperone that cooperates with DnaJ and the nucleotide exchange factor GrpE to promote protein folding [20]. It has been shown that DnaK alone exists in a very flexible open conformation and binding of ATP restricts its flexibility to some extent [21–24]. The co-chaperone DnaJ recognizes misfolded proteins and delivers them to DnaK [25]. DnaK recognizes substrate proteins by a relatively hydrophobic seven amino acid long sequence flanked at one side by a positive charge [26]. Substrate binding to DnaK stimulates ATP hydrolysis and this is accelerated by DnaJ [27]. This converts DnaK into a closed state, which tightly binds the substrate protein [28]. Upon ADP release, triggered by GrpE, DnaK adopts its open state thereby releasing the substrate protein [27]. It has been suggested that during this reaction cycle DnaK selects and therefore stabilizes certain conformations of the substrate, potentially preventing protein aggregation and promoting correct protein folding [29]. In addition, it has been suggested that DnaK disaggregates protein aggregates by a mechanism called entropic pulling [30]. However, detailed mechanistic insight into how DnaK exactly acts on misfolded and aggregated proteins remains to be elucidated. In particular, the role of the protein aggregate binding C-terminal loop in the unfolding processes is an open question in the field [31].

To fully unfold misfolded proteins the DnaKJ-GrpE system cooperates with ClpB and it has been shown that DnaK binds to ClpB indicating a direct substrate handover from DnaK to ClpB [32–34]. ClpB is another non-essential ATP-dependent chaperone and it consists of homohexameric complexes, which form a ring with a central pore [34]. In contrast to the DnaKJ-GrpE system, ClpB does not act through sole binding and release on substrate proteins, but uses the
hydrolysis of ATP to actively pull on misfolded proteins [34]. First, the substrate protein binds to flexible pore loops of ClpB through hydrophobic interactions [35]. Then, the conformation of these loops changes, thereby pulling the substrate protein through the pore, finally resulting in the unfolding of the substrate, a process also known as threading.

Another well-characterized chaperone system that acts downstream of TF is the GroEL/ES system [36]. Like ClpB, GroEL/ES is an ATP-dependent system, but it is characterized by forming large cylindrical complexes of two heptameric GroEL rings [37,38]. The chamber formed by two GroEL rings ends with a pore on each side that can be closed by a GroES heptamer [37,38]. It has been found that around 250 proteins interact with GroEL/ES and approximately 50 proteins are strictly dependent on GroEL/ES to fold properly [39,40]. These substrate proteins are characterized by a complex fold that often involves long range interactions and some of these proteins are crucial for the survival of E. coli, thus making GroEL/ES essential [39]. The GroEL/ES-mediated folding cycle has been studied in great detail. First, a substrate binds trough hydrophobic interactions to the pore region leading to a conformational expansion of the substrate and this is enhanced by ATP binding to GroEL [40,41]. Then, the substrate is sequentially released into the chamber and replaced from the pore region by a GroES heptamer [40,41]. The folding chamber restricts contact of a substrate protein with other proteins in the cytoplasm, thereby preventing protein aggregation during the folding process. Furthermore, the negatively charged inner surface of the folding chamber favours the compact packing of hydrophobic residues into the core of the substrate protein, thereby promoting protein folding [42]. Upon hydrolysis of ATP and release of ADP, GroES dissociates from GroEL and the folded substrate is released into the
cytoplasm [43]. Notably, an incorrectly folded substrate can rebind to
the GroEL pore region, thus entering another chaperone assisted
folding cycle. GroEL/ES does not only act as a passive cage but also
enhances the folding rate of substrate proteins [43]. Most GroEL sub-
strates have a molecular weight below 50 kDa which is consistent
with the size of the folding chamber. Nevertheless, GroEL may assist
the folding of larger proteins by binding such a substrate at the GroEL
pore region, expanding the substrate through ATP binding to GroEL
and finally releasing the substrate without its encapsulation [44].

As mentioned above, chaperones can also mediate the refolding of
aggregated proteins. Protein aggregates are insoluble, macromolecu-
lar structures that contain different misfolded or unfolded proteins and
protein aggregation is elevated under certain stress conditions i.e.,
during a heat shock or when a recombinant protein is produced
[45,46]. In E. coli, the small heat shock proteins IbpA and IbpB are
binding cooperatively to protein aggregates, thereby preventing fur-
ther aggregation [47,48]. Furthermore, IbpA and IbpB keep aggre-
gated proteins in a refolding competent state, enabling the unfolding
of aggregates by other chaperones, like the DnaKJ-GrpE system in
cooperation with ClpB [49]. However, if a protein cannot be refolded
properly it needs to be degraded to prevent its re-aggregation.
3.2 Protein degradation in the cytoplasm

At some point proteins are degraded by proteases and protein degradation is a complex and highly regulated process. The main proteases of *E. coli*, e.g., ClpXP, Lon and HslUV, belong to the AAA+ class (ATPase associated with a variety of cellular activities) and they consist of two functional units [50]. The unfolding domain sits on top of the proteolytic chamber and forms a ring with a pore, thereby preventing access of folded proteins to the proteolytic chamber [51]. Notably, different unfolding domains, e.g., hexameric ClpA or ClpX, can form a functional protease with a proteolytic chamber formed by the ClpP tetradecamer. To enter the proteolytic chamber substrate proteins are unfolded by the ATPase subunit in a way similar to the protein unfolding catalysed by the chaperone ClpB (see 3.1). Then, the linearized part of the substrate enters the proteolytic chamber where cleavage of peptide bonds is catalyzed by an activated serine (in case of ClpP and Lon) or threonine residue (in case of HslV) [52,53].

Proteolysis needs to be highly regulated because it is an irreversible process. The first layer of regulation is linked to the different ATPase subunits, because different ATPases have an overlapping, but not identical substrate pool. Furthermore, various cases have been reported where specific adaptor proteins are required to target a substrate to a ATPase domain [54–57]. Moreover, the binding of a substrate to an adaptor protein can be modulated by post-translational modifications or anti-adaptors [58,59]. Proteins have specific degradation sequences called degrons and also the accessibility of such degrons regulates protein degradation. Because degrons are hydrophobic sequences, they are more likely to be exposed in misfolded proteins, thereby targeting them for degradation. On the other hand, in properly folded proteins, degrons can be made accessible through endoproteolytic cleavage, by inducing conformational changes.
through interaction with a binding partner or by fusing a degron to the substrate protein [60–63]. Notably, it has been shown that the hydrophobic sequences representing degrons differ from the hydrophobic sequences recognized by chaperones [64]. Hence, although proteases and chaperones recognize misfolded proteins, the need for protein degradation or protein refolding of misfolded proteins is monitored differently.

3.3 Targeting of proteins to the cell envelope

About 50% of the *E. coli* proteome is not located in the cytoplasm, but is either inserted into the inner membrane, secreted into the periplasm, inserted into the outer membrane or released into the extracellular space [65]. Irrespective of their final destination, proteins are synthesized in the cytoplasm and proteins that do not remain in the cytoplasm are targeted to the inner membrane. Proteins targeted to the inner membrane are either inserted into the membrane or translocated across it and both processes are mainly facilitated the Sec-translocon, an inner membrane associated multiprotein complex (see 4.1.1). Proteins are targeted to the Sec-translocon by their N-terminal sequence that is recognized by targeting factors and protein targeting occurs either co- or post-translationally [66]. The co-translational pathway, which is used by membrane and some secretory proteins is characterized by coupling of protein synthesis and protein insertion into or translocation across the membrane (see 4.1.1) [67]. In contrast, post-translational protein targeting is characterized by the nearly complete synthesis of the secretory protein in the cytoplasm before being translocated across the inner membrane by the Sec-translocon (see 4.2.1) [68]. Most secretory proteins are designated to this targeting pathway. Proper targeting of cell envelope proteins is crucial be-
cause in the cytoplasm they are non-functional and are likely to induce protein misfolding and aggregation.
4 The *E. coli* cell envelope

The cell envelope encloses the cytoplasm and is composed of three compartments; the inner membrane, the periplasm and the outer membrane (Figure 1). The inner membrane consists of a symmetric lipid bilayer and membrane proteins and it separates the cytoplasm from the periplasm (see 4.1). The periplasm is an aqueous, gel-like environment that is much more oxidizing than the cytoplasm and it contains the peptidoglycan network, which gives the cell its stability (see 4.2). The outer membrane is an asymmetric lipid bilayer that is characterized by its surface exposed layer of lipopolysaccharides (LPS) (see 4.3).

The production of recombinant proteins in the cell envelope is highly desirable for many reasons, e.g., a recombinant protein may only fold properly if it is produced in the cell envelope or if its isolation from the cytoplasm results in persistent contaminations. However, producing high levels of recombinant proteins in the cell envelope is often very difficult (see 5). Hopefully, increasing our knowledge of the biogenesis and quality control of cell envelope proteins may pave the way to improve the production of recombinant proteins in this compartment of *E. coli*. 
4.1 The inner membrane

The inner membrane of *E. coli* consists of lipids and proteins. It has a hydrophobic core and polar or charged moieties form the surfaces of the membrane facing the cytoplasm and the periplasm. Due to the hydrophobic nature of the inner membrane, diffusion of polar or charged molecules across it is restricted. This allows a cell to establish concentration gradients, and these gradients are the driving forces for many essential processes. Inner membrane proteins fulfil different functions in a cell, e.g., they sense the environment, regulate transport across the membrane and catalyse enzymatic reactions. The setup of the inner membrane as a hydrophobic barrier that separates two aqueous compartments requires the action of specialised proteins to mediate the biogenesis of inner membrane proteins and to translocate proteins across the inner membrane into the periplasm.

4.1.1 Biogenesis of inner membrane proteins

It is estimated that about 30% of all *E. coli* genes encode for inner membrane proteins [65]. Inner membrane proteins can be associated with the membrane through interactions of the protein with the charged surface of the membrane (peripheral membrane proteins), a lipid moiety that is covalently attached to the protein (lipoproteins) or by spanning the membrane once or multiple times (integral membrane proteins). Here, only the biogenesis of integral membrane proteins will be described (Figure 3).
Figure 3. Inner membrane protein biogenesis and degradation.
In *E. coli* almost all membrane proteins are inserted co-translationally into the inner membrane. The ribosome translating a membrane protein is directed to the inner membrane via the signal recognition particle (SRP) and its receptor FtsY. At the membrane, the translating ribosome is handed over to either SecYEG/YidC, which mediates the insertion of membrane proteins with big soluble domains, or YidC alone, which mediates the insertion of membrane proteins with small soluble domains. SecA translocates sizable periplasmic loops of membrane proteins through the SecY channel in an ATP-dependent manner (see 4.1.1). QmcA and the HflCK complex can target misfolded membrane proteins for degradation by the endoprotease HtpX and the exoprotease FtsH (see 4.1.2).

Integral membrane proteins span the inner membrane with hydrophobic, α-helical structures, the so-called transmembrane helices. Transmembrane helices are inserted into the inner membrane by the Sec-translocon and/or the insertase/foldase YidC [69,70]. To prevent the aggregation of hydrophobic transmembrane helices in the cytoplasm the translation of integral membrane proteins is coupled to their membrane insertion. The co-translational targeting of integral membrane proteins to the inner membrane is realized by the signal recognition particle (SRP) and its membrane receptor FtsY (Figure 3) [71].

The *E. coli* SRP consists of the protein subunit Ffh and the 4.5S RNA [72]. SRP interacts with newly synthesized proteins when they
emerge from the ribosomal exit tunnel [73,74]. When SRP recognizes the first transmembrane helix of an integral membrane protein, the ribosome bound SRP binds to FtsY, thereby docking the translating ribosome to the inner membrane [72,75]. Subsequently, the translating ribosome is handed over either to YidC or to the Sec-translocon/YidC and both entities can mediate the biogenesis of integral membrane proteins [76].

The Sec-translocon is a heterotrimeric complex of SecY, SecE and SecG in a 1:1:1 stoichiometry [69]. In contrast to other *E. coli* multi-protein complexes, the genes encoding the Sec-translocon components are not organised in an operon, leaving the question how the levels of these proteins are regulated to form functional complexes in the correct stoichiometry. The structure and function of the Sec-translocon has been studied in great detail. The core Sec-translocon subunit SecY spans the membrane ten times and consists of two halves that together form an hourglass shaped channel that can be closed by a plug domain, thereby preventing leakage of ions and small molecules through the channel [77]. Presumably, during translation of an inner membrane protein transmembrane helices pass directly from the ribosomal exit tunnel into the SecY channel and they are subsequently released into the membrane through a lateral gate formed by the SecY transmembrane helices two, three and seven [77]. SecE spans the membrane three times and it binds SecY opposite to the lateral gate [77]. SecE is believed to clench both halves of SecY together, thereby stabilizing SecY and it was shown that SecY without SecE is degraded by the protease FtsH [78]. SecG consists of only two transmembrane helices and is, in contrast to SecY and SecE, not essential for cell survival [79]. Notably, YidC was found to be located at the lateral gate of SecY indicating that it cooperates with
the Sec-translocon to assist the proper insertion and folding of membrane proteins [80,81].

The transmembrane helices of integral membrane proteins can be connected by large periplasmic loops and the translocation of such loops through the SecY channel requires SecA [82]. SecA is the peripheral ATP-dependent motor protein of the Sec-translocon and it mediates the translocation of linear sequences through the Sec-translocon [83]. The role of SecA in protein translocation has been studied extensively, but it remains elusive if the translocation of periplasmic loops of integral membrane proteins follows the same mechanism. Notably, SecA binds to SecY at the same region as the ribosome does [84]. Thus, the ribosome needs to dissociate from SecY before SecA can be recruited. Previously, a structure of the translating ribosome bound to the Sec-translocon showed that the nascent membrane protein was not translated directly into the SecY channel, but formed a loop on the cytoplasmic surface of SecY [85]. Potentially, sequences which are supposed to be translocated into the periplasm are not hydrophobic enough to insert into the membrane and are therefore stalled in the SecY channel. Because protein translation continues, the nascent protein can then form a loop on the cytoplasmic side of SecY leading to the dissociation of the ribosome from SecY thus allowing the recruitment of SecA to the Sec-translocon. Then, SecA could translocate the periplasmic sequence through the SecY channel. Recently, it was shown that SRP does not only bind to the first transmembrane helix of integral membrane proteins, but that it can also recognize succeeding helices [86]. Therefore, after dissociation of SecA from SecY, re-association of the translating ribosome to SecY may be facilitated by SRP. Notably, it has not been investigated yet if FtsY is absolutely necessary in this re-association process.
YidC can also catalyse the co-translational insertion of transmembrane helices [70]. Recently, a high resolution crystal structure of the *E. coli* YidC was published [87]. The *E. coli* YidC consists of six transmembrane helices and a large periplasmic domain. It was proposed that YidC provides a hydrophobic surface that allows transmembrane helices to slide into the membrane [88]. Furthermore, it was suggested that the hydrophilic cavity, present at the cytoplasmic side of YidC, accommodates soluble domains of integral membrane proteins while their transmembrane helices are inserted into the lipid bilayer [89]. Interestingly, it was reported that YidC interacts transiently with its substrates, potentially allowing it to insert transmembrane helices more rapidly into the inner membrane than the Sec-translocon does [90]. Integral membrane proteins, which can be inserted into the membrane by YidC are characterized by small soluble domains of less than 100 amino acids [91]. Integral membrane proteins with larger periplasmic domains may be handed over from YidC to the Sec-translocon to complete their biogenesis [91]. Besides interacting with the Sec-translocon, YidC can also form a complex with the accessory Sec-translocon components SecDF-YajC and the transient formation of these different translocon complexes *in vivo* makes it difficult to assign specific functions to YidC [92].

Integral membrane proteins must also adopt their correct topology, meaning that the transmembrane helices are positioned such that soluble domains end up in their designated compartment, *i.e.*, in the cytoplasm or in the periplasm. It has been shown that topogenesis of integral membrane proteins depends on many factors such as their primary sequence and the lipid composition of the membrane [93]. One decisive factor determining the topology of a membrane protein is the "positive inside rule", *i.e.*, positively charged residues that are located in close proximity to transmembrane helices will be preferen-
tially located on the cytoplasmic side of the inner membrane [94]. However, it is not well understood yet if all integral membrane pro-
teins adopt their correct topology during their insertion into the mem-
brane or if topogenesis can also occur afterwards. Integral membrane
proteins that are not inserted properly into the membrane or that do
not adopt their correct topology are likely to be non-functional and
have to be cleared from the membrane.

4.1.2 Degradation of inner membrane proteins

The misfolding of membrane proteins can have a tremendous impact
on the structure and function of the inner membrane, thereby com-
promising the cell fitness. Membrane proteins are degraded by prote-
ases and in *E. coli* the best studied membrane associated protease is
FtsH (Figure 3) [95]. FtsH is an ATP-dependent zinc metalloprotease,
which spans the membrane two times and its transmembrane helices
are essential for its oligomerization into hexamers [96]. At the cyto-
plasmic side of the membrane the ATPase domain of FtsH and its
protease domain form a module of two consecutive rings [97]. This
architecture is similar to other AAA+ ATPases, allowing assumptions
about FtsH’s mode of action (see 3.2). It was suggested that FtsH
substrates bind to a conserved aromatic phenylalanine at the edge of
the pore of the ATPase subunit and ATP hydrolysis triggers confor-
mational changes, pulling the substrate into the protease domain [95].
Thereby, the substrate becomes unfolded and available for degrad-
ation by the protease domain. Besides unfolding the substrate, FtsH
may use the hydrolysis of ATP to pull the substrate out of the mem-
brane. It has been shown that FtsH functions as an exoprotease that
degrades substrates from the N-terminus, the C-terminus and poten-
tially from cytoplasmic loops [98–100]. Substrates of FtsH include
SecY, SecE, integral membrane proteins that are involved in the biogenesis of outer membrane lipids and cytoplasmic proteins like the alternative sigma factor σ^{32} [101,102].

The activity of FtsH is regulated by the HflKC complex. HflKC forms a membrane integrated hexameric complex with a large periplasmic domain [103]. It was suggested that the periplasmic domain of HflKC can recognize misfolded periplasmic loops of membrane proteins, subsequently targeting the misfolded membrane protein for FtsH-dependent degradation [104]. Misfolded cytoplasmic loops of membrane proteins may be detected by the membrane protein QmcA through its large cytoplasmic domain, subsequently directing the misfolded membrane protein to FtsH [105]. Such a system that senses misfolded domains of membrane proteins on both sides of the inner membrane and targets the respective membrane proteins for degradation appears to be appealing. However, so far little evidence for this hypothesis has been reported.

Another membrane integrated protease is HtpX [106]. In contrast to FtsH, HtpX is an ATP-independent endoprotease cleaving within cytoplasmic loops of membrane proteins [107]. Due to this endoproteolytic activity, HtpX creates new starting points for the FtsH-dependent degradation of membrane proteins.

4.2 The periplasm

The periplasm of *E. coli* is an aqueous, gel-like environment that is enclosed by the inner and the outer membrane (Figure 1). In contrast to the cytoplasm, no energy transferring molecule like ATP is present in the periplasm. Compared to the cytoplasm, the periplasm is an oxidizing environment, allowing proteins to form intramolecular disulfide bonds between cysteine residues. Proteins present in the periplasm...
plasm e.g., participate in the transport of nutritions, detoxify harmful compounds, sense changes in the environment and synthesize the peptidoglycan, a network of carbohydrates and peptides that gives the cell its stability. Chaperones assist the folding of proteins in the periplasm and holdases facilitate the transport of proteins, designated for the outer membrane, across the periplasm (Figure 4). However, to reach the periplasm proteins have to be translocated across the inner membrane, a process known as protein secretion. Importantly, knowledge about protein secretion and the folding of secreted proteins is crucial to be able to design strategies to improve the production of recombinant proteins in the periplasm of *E. coli* (see 5.4).

**Figure 4. Biogenesis of secretory proteins.**
In *E. coli* most secretory proteins are targeted via the post-translational SecA/SecB pathway to the Sec-translocon, which facilitates translocation of protein across the inner membrane and the SecDF-YajC complex may assist in this process. The ATPase activity of SecA drives the translocation of proteins across the inner membrane. On the periplasmic site of the inner membrane the chaperones PpiD and YfgM that can associate with the Sec-
translocon and the chaperones FkpA and Spy can assist the folding of secreted proteins (see 4.2.1). DsbA can introduce disulfide bonds in secreted proteins and electrons are transferred from DsbA via DsbB into the Q-pool. Incorrectly formed disulfide bonds can be re-oxidized by DsbC and the required electrons are transferred from the cytoplasmic thioredoxin (Trx) via DsbD to DsbC. Outer membrane proteins (OMPs) are mainly transported across the periplasm to the Bam-complex by SurA and Skp. The Bam-complex mediates the insertion of OMPs into the outer membrane (see 4.3.1). In the periplasm, proteases like DegP, Tsp and Ptr can degrade misfolded proteins (see 4.2.2).

4.2.1 Biogenesis of secretory proteins

About 20% of the *E. coli* proteome is estimated to be secreted into the periplasm [65]. These proteins cross the inner membrane in a folded or unfolded state and protein secretion is facilitated by integral membrane protein complexes, so-called translocases. The Tat-translocon transports folded proteins across the inner membrane [108,109]. Because the vast majority of secretory proteins cross the membrane in an unfolded state via the Sec-translocon (see 4.1.1) only the biogenesis of Sec-dependent secretory proteins will be described in this section.

Secretory proteins are targeted to the Sec-translocon by their N-terminal signal peptide. Signal peptides are around 20 amino acids long sequences that contain a charged N-terminus, a hydrophobic h-region and a polar C-terminus [110]. The C-terminus contains the signal peptide cleavage site, which is recognized by leader peptidase that clips off the signal peptide upon protein translocation [110,111]. Some secretory proteins are targeted co-translationally to the Sec-translocon like integral membrane proteins (see 4.1.1). However, most secretory proteins are targeted to the Sec-translocon in a post-translational manner [68]. Post-translational protein targeting is characterized by the (almost) complete synthesis of the secretory protein
in the cytoplasm before reaching the Sec-translocon. Post-
translationally targeted secretory proteins can interact with cyto-
plasmic chaperones like SecB. Binding of SecB to hydrophobic
stretches of secretory proteins is believed to prevent their premature
folding, thereby keeping them in a translocation competent state
[112]. It is generally assumed that the signal peptide of a secretory
protein is recognized by SecA and that SecA binds to the SecY sub-
unit of the Sec-translocon, thereby guiding the secretory protein to the
protein conducting channel [113,114]. Besides facilitating protein tar-
getting, SecA also triggers the release of SecB from the secretory
protein and drives protein translocation through the Sec-translocon in
an ATP-dependent manner [115,116]. Already during the transloca-
tion of a secretory protein through the Sec-translocon chaperones in
the peri-plasm can assist its folding and the folding process may ac-
celerate the translocation of a secretory protein.

During the last years progress has been made to better understand
the chaperone assisted folding process of secreted proteins. The
chaperones PpiD and YfgM are each tethered to the periplasmic side
of the inner membrane by a single transmembrane helix and both can
be associated with the Sec-translocon [117,118]. The association of
these chaperones with the Sec-translocon may allow them to interact
first with secretory proteins that are being translocated. YfgM was
identified as a periplasmic chaperone potentially being involved in the
trafficking of outer membrane proteins [119]. PpiD has a peptidyl-
prolyl isomerase domain (PPIase domain), which was shown to be
inactive and not essential for its chaperone activity [120]. In contrast,
the PPIase domain of the periplasmic chaperone PpiA can catalyse
the cis-trans isomerisation of peptide bonds in conjunction with a
proline residue [121]. However, PpiA is not essential for cell viability
or growth and its exact physiological function is unclear.
Important for the proper folding of a variety of secreted proteins is the formation of correct disulfide bonds between cysteine residues. These covalent bonds stabilize protein folds and in the oxidative environment of the periplasm the formation of such bonds is catalysed by the disulfide bond formation (Dsb)-system [122,123]. Oxidized DsbA contains a redox active disulfide bond that is reduced when DsbA introduces a disulfide bond into a secreted protein [124]. To recycle DsbA into its oxidized state, electrons are transferred from DsbA to the integral membrane protein DsbB, which funnels these electrons into the respiratory chain [125,126]. In proteins containing more than two cysteines incorrectly formed disulfide bonds can be reduced by DsbC [127,128]. The electrons required for this reaction are transferred from cytoplasmic thioredoxins to the integral membrane protein DsbD that in turn reduces the cysteines in DsbC [129]. Dimeric DsbC detects incorrectly formed disulfide bonds through hydrophobic residues that are exposed in misfolded secreted proteins [130]. After the reduction of an incorrectly formed disulfide bond by DsbC, the secreted protein can undergo another cycle of disulfide bond formation catalysed by DsbA.

Besides assisting protein folding, periplasmic chaperons also prevent the premature folding and aggregation of proteins designated for the outer membrane, so-called outer membrane proteins (OMPs). The periplasmic chaperone SurA is the major protein that transports OMPs across the periplasm to their insertion site in the outer membrane, and it does not interact with soluble periplasmic proteins [131,132]. SurA consists of a chaperone domain and two PPIase domains [133]. Notably, one PPIase domain is inactive and the other PPIase domain is not essential for the function of SurA [133]. SurA binds with high affinity to aromatic residues exposed by OMPs, thereby preventing their premature folding in the periplasm [134,135].
It has been proposed that SurA hands over OMPs to the Bam-machinery, an outer membrane integrated multiprotein complex that inserts OMPs into the outer membrane (see 4.3.1) [136]. Another chaperone that is important for the biogenesis of OMPs is Skp. Skp is a small chaperone that binds OMPs as well as a few periplasmic substrates [137,138]. To prevent the premature folding and aggregation of OMPs in the periplasm, trimeric Skp forms a cage with three α-helical arms enclosing a substrate [139]. Probably, this cage-like architecture results in the observed higher affinity of Skp to OMPs compared to the binding affinity of SurA to OMPs [135]. Notably, the size of the Skp cage appears to be dependent on the substrate bound and substrates which are too large to fit inside the Skp cage can be sequestered allowing additional Skp trimers to bind [140]. It has been proposed that both Skp and SurA stabilize a dynamic state of unfolded OMPs allowing these OMPs to sample multiple conformations to reach a low free energy state [135]. In that way β-strands, which are the dominant secondary structure in OMPs, may be pre-formed before an OMP is actually integrated into the outer membrane.

In their natural environment cells encounter stress that can affect protein folding like a rapid temperature shift or a change in the salt concentration. Upon such stress, accumulation levels of the periplasmic chaperone Spy are dramatically increased, presumably to avoid protein misfolding and aggregation [141]. The Spy dimer adopts a unique cradle-like structure with a concave positively charged surface [142]. Spy was shown to bind the small model substrate lm7 very fast and independent of the folding state of lm7 [143]. Spy seems to flatten the energy landscape of the folding pathway of lm7, thereby allowing the substrate to fold or unfold completely when bound to Spy [143]. Notably, so far Spy is the only periplasmic chaperone with demonstrated unfolding activity. However, it remains unclear if Spy
cannot only unfold properly folded proteins, but also misfolded or even aggregated substrates. Furthermore, the substrate pool of Spy has not been defined yet.

Taken together, although for many periplasmic chaperones it is not well understood how they assist protein folding, a general concept appears to be preventing protein aggregation rather than refolding misfolded proteins. Presumably, without energy transferring molecules like ATP, the refolding of secreted proteins may be impossible, making protein aggregation in the periplasm a serious threat to the cell viability. Therefore, misfolded proteins, which are prone to aggregation, need to be cleared immediately from the periplasm.

4.2.2 Protein degradation in the periplasm

In the *E. coli* periplasm, protein degradation has to be controlled precisely because protein synthesis and protein translocation across the inner membrane are costly processes. So far, more than 20 periplasmic proteases have been identified, but only a few, e.g., DegP, Ptr and Tsp, have been characterized in more detail [144].

The most prominent periplasmic protease is DegP [145,146]. DegP is a serine protease that may function as a chaperone below 28°C, but acts as a protease at higher temperatures [147]. The protease activity of DegP has been studied in quite some detail. In its inactive state DegP forms hexamers with a disordered active site [148,149]. Upon substrate binding, DegP trimers are formed, which subsequently oligomerize into higher order complexes, thereby forming a proteolytic camber [149]. This closed architecture makes it more probable that a substrate is cut more than once by DegP, thereby preventing the premature release of aggregation prone protein fragments. Furthermore, the substrate binding domain of DegP recog-
nizes similar sequences as its active site does [150]. This allows the substrate binding domain to recapture a recent cleavage product and transfer it to a neighbouring active site for another round of cleavage [150]. It has been shown that DegP can degrade a broad range of periplasmic proteins, including unfolded OMPs [149]. On the other hand, it has been suggested that DegP in its role as a putative chaperone forms a cavity that transports OMPs across the periplasm to the outer membrane [136,149].

4.3 The outer membrane

The outer membrane encloses the periplasm and thereby constitutes the surface of *E. coli* (Figure 1). In contrast to the inner membrane, the outer membrane is an asymmetric lipid bilayer [151]. The periplasmic facing leaflet of the outer membrane contains phospholipids and lipoproteins. The outer leaflet of the outer membrane consists mainly of LPS. The major component of LPS is lipid A, which can cause a strong immune response in humans. The hydrophobic acyl chains of lipid A form the core of the outer leaflet and its long chains of branched carbohydrates face the extracellular space. The carbohydrate structure has a complex and diverse composition, which depends on the *E. coli* strain, the available nutrients and the physiological state of the cell.

The outer membrane contains also outer membrane proteins (OMPs), which can form pores allowing the diffusion of molecules smaller than around 700 Da across the membrane [152]. Therefore, the outer membrane is, in contrast to the inner membrane, not considered to serve as diffusion barrier. Notably, studying the structure and function of OMPs has led to the development of different plat-
forms to display recombinant proteins on the surface of *E. coli* (see 5.5).

### 4.3.1 Biogenesis of outer membrane proteins

OMPs adopt, in contrast to the α-helical structures of inner membrane proteins, amphipathic β-sheet structures that form a barrel-like architecture [7]. As all proteins, OMPs are synthesized in the cytoplasm, but they have to cross the inner membrane and the periplasm before they are finally inserted into the outer membrane. Most OMPs cross the inner membrane via the Sec-translocon and their targeting to the Sec-translocon occurs primarily post-translationally (see 4.2.1). After passing the inner membrane, OMPs are transported by periplasmic chaperones in an unfolded state to the outer membrane (see 4.2.1 and Figure 4). There, OMPs are inserted into the outer membrane by the Bam-complex [153]. The Bam-complex consists of five subunits BamA-E [154]. Only BamA and BamD are essential for cell viability, but *in vivo* all components are necessary for the efficient biogenesis of OMPs [154]. Crystal structures of the core subunit BamA show a membrane inserted β-barrel consisting of 16 strands and five periplasmic polypeptide transport-associated (POTRA) repeat domains [155,156]. POTRA domain five was found to be in close proximity to BamA, thereby closing the barrel [155]. The most intriguing feature of the BamA barrel is its junction between strand 1 and strand 16 [155,157]. These two strands are considerably shorter than in other membrane inserted β-barrels. Therefore, the junction is less tightly closed and the existence of a lateral gate between strand 1 and 16 was proposed [155,158]. Furthermore, a BamA exit pore, which would allow the translocation of extracellular loops, has been identified [158]. The other subunits of the Bam-complex, BamB-E, are all
anchored to the periplasmic side of the outer membrane by an N-terminally attached lipid moiety and they interact with the POTRA domains of BamA [154].

Little is known about the mechanism of Bam-mediated OMP insertion. Potentially, OMPs fold themselves at the outer membrane and they require only a locally disturbed lipid bilayer to insert into the membrane [159]. However, such a mechanism seems unlikely for large and complex OMPs. Alternatively, the insertion of OMPs could be a more sequential process, where the first β-hairpin would use the lateral gate of BamA as a template and the following strands would fold and insert by β-augmentation [160]. After the insertion of the last strand, the new β-barrel could bud off from BamA. How BamB-E assist the insertion of OMPs remains elusive, but it has been suggested that they regulate conformational changes in BamA [154]. Potentially, BamB-E coordinate also the transfer of unfolded OMPs from chaperones, like SurA, to BamA [154].

So far, it is not known if OMPs are completely functional after their insertion into the outer membrane or if they require further processing to reach their functional conformation. It has been found that some OMPs contain disulfide bonds, potentially increasing their stability [123]. Notably, the degradation of OMPs and the regulation of this process have not been studied in detail.
5 Recombinant protein production in the *E. coli* cell envelope

To be functional many recombinant proteins, like membrane proteins and disulfide bond containing proteins, need to be produced in the cell envelope of *E. coli*. Unfortunately, attempts to produce recombinant proteins in the cell envelope often compromise the fitness of the cell and consequently biomass formation and protein production yields are low. Here, a general overview of the setups that are routinely used to produce recombinant proteins will be given (see 5.1). Furthermore, it will be described how a gene encoding a recombinant protein can be modified to facilitate protein detection and isolation as well as to improve protein production yields (see 5.2). Next, strategies will be discussed that have been used to improve the yields and/or the quality of recombinant proteins produced in the inner membrane (see 5.3) and the periplasm (see 5.4) of *E. coli*. Finally, approaches will be presented to display recombinant proteins on the surface of *E. coli* (see 5.5) or to export a recombinant protein into the culture medium (see 5.6).

5.1 Producing recombinant proteins in *E. coli*

In *E. coli*, recombinant proteins are commonly produced using plasmids, which are circular DNA molecules that contain the recombinant gene. To ensure that plasmids are maintained in a cell they usually contain a gene encoding for an antibiotic resistance marker and by
adding the corresponding antibiotic to the culture medium only cells that contain the plasmid can survive. During my Ph.D. studies, I used antibiotic marker proteins that give rise to the resistance to either kanamycin or chloramphenicol. Depending on the origin of replication of a plasmid, the number of plasmids can range from only a few copies up to several hundred copies per cell. Because the synthesis of the plasmid DNA in a cell consumes resources, the plasmid copy number can have a substantial impact on the production of a recombinant protein. Furthermore, the number of plasmids per cell determines the number of copies of the gene encoding a recombinant protein, which can also affect production kinetics and consequently yields of the recombinant protein produced. Throughout my thesis, medium copy number plasmids were used with 15-30 plasmid copies per cell (plasmids with either a pMB1 or a p15A origin of replication).

To regulate the expression of a gene encoding a recombinant protein, different promoter systems can be used (see Table 2).

Table 2. Examples of promoters used to produce recombinant proteins in *E. coli*.

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>T7</em></td>
<td>T7 RNAP required</td>
</tr>
<tr>
<td><em>lac</em></td>
<td>induced with lactose or IPTG</td>
</tr>
<tr>
<td><em>trp</em></td>
<td>induced by Trp limitation</td>
</tr>
<tr>
<td><em>tac</em></td>
<td>hybrid promoter of trp and lac, induced with IPTG</td>
</tr>
<tr>
<td><em>phoA</em></td>
<td>induced by phosphate limitation</td>
</tr>
<tr>
<td><em>rhaBAD</em></td>
<td>induced with rhamnose</td>
</tr>
<tr>
<td><em>araBAD</em></td>
<td>induced with arabinose</td>
</tr>
<tr>
<td><em>Tet</em></td>
<td>induced with tetracycline</td>
</tr>
<tr>
<td><em>L</em></td>
<td>induced at high temperatures</td>
</tr>
<tr>
<td><em>CspA</em></td>
<td>induced at low temperatures</td>
</tr>
</tbody>
</table>
A promoter is a DNA sequence that is recognized by an RNA polymerase and the binding of an RNA polymerase to a promoter can be controlled by regulators. Regulators are proteins that promote or prevent RNA polymerases to bind to a promoter, thereby either enhancing or repressing transcription. Importantly, the conformation or the abundance of the regulator can be altered by different stimuli in such a way that it does not interfere with transcription. Therefore, manipulating a specific stimulus, e.g., the concentration of a certain compound or the temperature, allows regulating the expression of a gene from the respective promoter.

Both *E. coli* K- and B-strains have been used successfully to produce recombinant proteins. B-strains are considered to be more efficient to produce recombinant proteins due to their faster biomass formation in minimal medium, their enhanced amino acid synthesis rates and their reduced production of acetate that, at high concentrations, inhibits biomass formation [161]. *E. coli* B-strains that are most widely used to produce recombinant proteins are BL21(DE3) and derivatives thereof (see Table 3), mainly because in these strains recombinant protein production can be driven by the very strong T7 promoter system.

BL21(DE3) was created by Studier and Moffat in 1986 by integrating the gene encoding the RNA polymerase from bacteriophage T7 (T7 RNAP) into the chromosome of *E. coli* (see Figure 5) [162,163]. The T7 RNAP transcribes much faster than endogenous *E. coli* RNAPs and it recognizes exclusively the T7 promoter that is not recognized by *E. coli* RNAPs [162–164]. Therefore, in the T7 system the recombinant gene is expressed from the T7 promoter. However, in this system recombinant gene expression is controlled indirectly by regulating the amount of the T7 RNAP. The expression of the t7rnap itself is governed by the lacUV5 promoter, which is a stronger variant.
of the *E. coli* lac promoter [165,166]. The lacUV5 promoter can be activated by allolactose or its more commonly used non-hydrolysable analogue isopropyl-ß-D-thiogalactopyranoside (IPTG). Taken together, in the T7 system high protein production yields are achieved by producing high levels of a highly active RNAP, resulting in high levels of mRNA encoding for the recombinant protein. Ideally, this setup should lead to the production of high amounts of the recombinant protein.

### Table 3. *E. coli* strains used for the T7-based production of recombinant proteins

<table>
<thead>
<tr>
<th>Strain</th>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>BL21(DE3)</td>
<td>Δlon, ΔompT</td>
</tr>
<tr>
<td>C41/C43(DE3)</td>
<td>decreased target gene expression (see 5.3.1)</td>
</tr>
<tr>
<td>Mt56(DE3)</td>
<td>decreased target gene expression (see 5.3.1)</td>
</tr>
<tr>
<td>BL21(DE3) pLysE/S</td>
<td>decreased target gene expression</td>
</tr>
<tr>
<td>Lemo21(DE3)</td>
<td>titratable target gene expression (see 5.3.2)</td>
</tr>
<tr>
<td>KRX</td>
<td>T7 RNAP produced from P<em>raBAD</em></td>
</tr>
<tr>
<td>Rosetta(DE3)</td>
<td>increased availability of rare tRNAs</td>
</tr>
<tr>
<td>Origami(DE3)</td>
<td>enables S-S bond formation in cytoplasm</td>
</tr>
<tr>
<td>SHuffle</td>
<td>enables S-S bond formation in cytoplasm</td>
</tr>
<tr>
<td>BLR(DE3)</td>
<td>ΔrecA, increased plasmid stability</td>
</tr>
</tbody>
</table>

Using *E. coli* BL21(DE3) as a host to produce recombinant proteins may also be favourable, because it is devoid of the cytoplasmic protease Lon and the outer membrane protease OmpT, and the absence of these proteases in this strain may increase the stability of a produced recombinant protein [167]. In particular, the absence of OmpT
is preferable, because this protease is very stable and can degrade endogenous and recombinant proteins after cell lysis [168]. All these criteria make *E. coli* BL21(DE3) a popular host to produce recombinant proteins in the cytoplasm and cell envelope and in academia *E. coli* BL21(DE3) and its derivatives dominate the protein production field. Therefore, during my thesis, I focused on this strain and its derivatives to increase the yields of recombinant proteins produced in the *E. coli* cell envelope.

**Figure 5. T7-based production of recombinant proteins in BL21(DE3).** On the chromosome of BL21(DE3) LacI represses the expression of the *t7map* from the *lacUV5* promoter. When the inducer IPTG is added, LacI binds IPTG and due to a conformational change it dissociates from its binding site, thereby allowing the production of the T7 RNAP from the *lacUV5* promoter. The T7 RNAP recognizes the T7 promoter on the pET-plasmid and transcribes the target gene, ultimately resulting in the production of the target protein. T7 lysozyme can be used to modulate the activity of the T7 RNAP.
5.2 Modifying the gene encoding a recombinant protein

A gene encoding for a recombinant protein is often modified to increase the protein production yield and/or to facilitate the detection and isolation of the produced protein. To enhance the translation of a recombinant protein the codons of the corresponding gene can be changed to synonymous codons that allow faster translation [169]. However, at least in some cases it may be beneficial to preserve the positions with a low translation rate in order to allow co-translational folding of a recombinant protein [170,171].

To be functional many recombinant proteins that are produced in E. coli are targeted to the periplasm. To target a recombinant protein to the periplasm it can be fused genetically to an N-terminal signal peptide. Notably, a variety of signal peptides, covering Sec- and Tat-dependent protein translocation have been used to target recombinant proteins to the E. coli periplasm (see 5.4.1). Furthermore, Sec-dependent signal peptides have been used that mediate co- and post-translational targeting to the Sec-translocon.

Fusing the gene encoding for a recombinant protein to a gene encoding for a fusion partner can increase the stability and solubility of a recombinant protein produced as well as facilitate its isolation. Commonly used fusion partners are the E. coli proteins maltose binding protein (MBP), thioredoxin (Trx), glutathione S-transferase (GST) or the jellyfish green fluorescent protein (GFP) and variants thereof [8]. GFP is a prominent fusion partner for integral membrane proteins because it enables the easy detection of the fusion protein by monitoring GFP fluorescence. Furthermore, GFP fluorescence can be used to screen for cells that produce the recombinant membrane protein-GFP fusion more efficiently using e.g., flow cytometry. Notably, GFP fused to the C-terminus of an integral membrane protein becomes only fluorescent if the integral membrane protein is inserted
into the inner membrane, but not if the integral membrane protein aggregates in the cytoplasm [172]. Therefore, I used GFP fluorescence to screen for *E. coli* mutants that insert membrane protein-GFP fusions more efficiently into the inner membrane and this facilitated the isolation of the new membrane protein production strain Mt56(DE3) (Paper I).

Besides using entire proteins as fusion partners, also short peptides, so-called tags, can be fused genetically to a recombinant protein. Tags have a lower risk of interfering with the native structure of a recombinant protein than the aforementioned fusion proteins. Furthermore, tags facilitate the detection of a fusion due to the availability of antibodies and other binders that specifically recognize tags. Moreover, affinity matrices have been developed that bind tags, thus allowing the isolation of a recombinant protein fused to a tag from a complex cell lysate [173]. Examples of frequently used tags are the poly-His-, Strep-, HA-, c-Myc- and Flag-tag [174].

Notably, fusing a recombinant protein genetically to more than one fusion partner is not uncommon, *i.e.*, a signal peptide facilitates the secretion of a recombinant protein into the periplasm and a poly-His-tag facilitates its detection and isolation.

5.3 Improving inner membrane protein production yields

Membrane proteins fulfil essential functions in the cell and therefore they are also important drug targets [175]. In the cell, the amounts of membrane proteins are usually low, making their recombinant production necessary to be able to study their structure and function. However, due to the complex biogenesis of membrane proteins, different factors can severely hamper their production. Problems may include inefficient targeting to the Sec-translocon, slow membrane
insertion or insertion with the wrong topology, impaired translocation of loops into the periplasm, misfolding of soluble domains, and degradation. Therefore, it is not surprising that very low yields are frequently observed when attempting to produce membrane proteins [176]. However, *E. coli* can either be evolved (see 5.3.1) or engineered (see 5.3.2) to improve membrane protein production yields.

### 5.3.1 Directing the evolution of *E. coli*

Directing the evolution of *E. coli* can be a powerful tool to isolate strains with improved membrane protein production characteristics, because knowledge of the factors limiting the production yields of a membrane protein is not required. However, the successful isolation of a strain with improved membrane protein production characteristics is highly dependent on a powerful selection method or a screening approach that enables to test many mutant strains for their ability to produce high levels of the membrane protein of interest.

More than 20 years ago, in the lab of John Walker, growth was used as a selection criterion to isolate the *E. coli* BL21(DE3) derivatives C41(DE3) and C43(DE3) [177]. C41(DE3) was isolated from BL21(DE3) producing the mitochondrial oxoglutarate-malate transporter (OGCP). Because the production of OGCP impairs biomass formation of BL21(DE3) dramatically, in this screen it was attempted to isolate mutants that form colonies on agar plates while producing OGCP. This screen yielded the mutant C41(DE3) that formed a small colony and produced OGCP efficiently and mutants that formed large colonies but did not produce any OGCP. C43(DE3) was isolated from C41(DE3) in a similar setup using the β-subunit of the *E. coli* F-ATPase as target membrane protein. Notably, at that time it was not tested if C41(DE3) and C43(DE3) insert the recombinant membrane
proteins that were used for their isolation in the inner membrane. Later it has been shown that the defining mutations in the C-strains weaken the lacUV5 promoter that governs the expression of the t7map [178]. Notably, the production of many recombinant proteins results in the accumulation of mutations similar to the ones that are defining for C41(DE3) and C43(DE3) [179,180]. Nevertheless, C41(DE3) and C43(DE3), also known as the Walker strains, have become the standard production hosts in the membrane protein research field.

In 2009, the Georgiou lab described the isolation of new membrane protein production strains. They used the random chromosomal insertion of the Tn5 transposon to create a library of E. coli mutants with single gene deletions [181]. Then, they screened their library of approximately 13,000 mutants for strains producing high levels of the G-protein coupled receptor (GPCR) human central cannabinoid receptor CB1 C-terminally fused to GFP using fluorescence activated cell sorting (FACS). Using this selection method two E. coli mutant strains were isolated that produced about eight times more CB1-GFP and both strains were characterized. Interestingly, in both mutants the Tn5 transposon disrupted the dnaJ gene encoding for the DnaJ co-chaperone of the major E. coli DnaKJ-GrpE chaperone system. The authors did speculate that DnaJ either impairs the targeting of CB1 to the Sec-translocon or directs CB1 for degradation. Hence, the absence of DnaJ would improve production yields of membrane inserted CB1. However, it was not tested if deleting dnaJ also improves the yields of other recombinant membrane proteins.

Also in 2009, the Bowie lab reported the isolation of E. coli mutant strains that could produce high levels of the Mycobacterium tuberculosis membrane protein rhomboid-Rv1337 [182]. First, Rv1337 was fused genetically to two different antibiotic resistance markers giving
rise to either trimethoprim or kanamycin resistance. Then, *E. coli* TOP10 harbouring a plasmid encoding for Rv1337 fused to the trimethoprim resistance marker was mutagenized and strains producing the Rv1337 fusion protein were selected for using trimethoprim resistance as selection criterion. These mutants were subsequently transformed with an additional plasmid encoding for Rv1337 fused to the kanamycin resistance marker. Selecting for resistance to both trimethoprim and kanamycin resulted in the isolation of mutants that produced both Rv1337 fusion proteins. Using this setup 47 mutant strains were isolated out of which 17 showed improved production of Rv1337 and five strains of these 17 strains were analyzed further. To cure the isolated strains from the selection plasmids, both plasmids had been engineered in such a way that they contain a cleavage site for the endonuclease I-CreI. Producing I-CreI allows the rapid cleavage of the selection plasmids *in vivo*, thereby lowering the risk of accumulating additional mutations in the chromosome during the curing process. Notably, one of the five strains that showed improved production of Rv1337 produced also five other membrane proteins to considerably higher levels than the parental *E. coli* strain. Unfortunately, this strain was not characterized leaving the underlying mechanism of its improved membrane protein production characteristic elusive.

In 2014, the Poolman lab reported the isolation of *E. coli* strains for the production of membrane proteins [183]. They fused two bacterial membrane proteins, GltP and BcaP, at their C-terminus to a dual fusion protein consisting of GFP and the erythromycin resistance protein ErmC. Selecting for mutant strains that produced more of the fusion proteins was done by increasing the erythromycin concentration gradually, assuming that mutants with increased accumulation levels of the properly inserted membrane protein fusion would with-
stand higher erythromycin concentrations. Notably, both fusions were produced from the relatively weak arabinose promoter and the concentration of the inducer arabinose was kept constant throughout the selection process. This selection procedure resulted in the isolation of four mutant strains that all produced both fusion proteins to higher levels than the parental *E. coli* strain. Notably, all four mutant strains had acquired mutations in *hns*, which encodes for a global transcription factor and it is involved in chromosomal organisation. In the best producing strain the C-terminal end of H-NS was deleted. However, a complete deletion of *hns* in the parental strain did not improve membrane protein production yields considerably and it was not verified if the C-terminal truncation of H-NS alone increases yields. Therefore, it remains unclear if/how mutations in *hns* improved membrane protein production yields.

Recently, the Beckwith lab screened for *E. coli* mutants that produced increased levels of a variant of the eukaryotic membrane protein vitamin K epoxide reductase (VKORc1ΔAAR) [184]. VKORc1ΔAAR can complement for a deletion of *dsbB*, an integral membrane protein involved in disulfide bond formation in the periplasm, and an *E. coli dsbB* knock-out strain is not viable in the presence of disulfide bond breaking agents. Therefore, in the presence of a disulfide bond breaking agent, growth could be used to select for strains from a mutagenized *E. coli dsbB* knock-out library that produced functional VKORc1ΔAAR. This screen yielded 11 mutant strains carrying between five and 40 mutations. By characterizing some of these mutations individually, it was shown that mutations in the genes encoding for the *E. coli* insertase YidC and the protease HslV improved production yields of functional VKORc1ΔAAR. It was proposed that the mutant YidC is able to promote the proper topogenesis of VKORc1ΔAAR and improved yields of VKORc1ΔAAR in a hslV knock out strain indicate
that HslUV is involved in the degradation of VKORc1_ΔAAR. Hence, this study demonstrates that bottlenecks in the biogenesis of a recombinant membrane protein and its degradation can limit protein production yields.

As part of my thesis, we aimed to isolate new *E. coli* BL21(DE3) derivatives with improved membrane protein production characteristics (Paper I). To screen for new membrane protein production strains we used the *E. coli* membrane protein insertase YidC C-terminally fused to GFP as a target protein because it had been shown that the production of YidC-GFP does, in contrast to the production of many other recombinant proteins, not result in the accumulation of mutations that are defining for the membrane protein production strains C41(DE3) and C43(DE3). Since the production of YidC-GFP impairs biomass formation of BL21(DE3) dramatically, mutants were selected for based on their ability to form colonies on agar plates when producing YidC-GFP. Because only membrane inserted YidC-GFP exhibits a fluorescent signal and the fluorescence intensity should correlate with the accumulation levels of YidC-GFP, monitoring the fluorescence of mutants allowed us to screen for strains with high production levels of membrane inserted YidC-GFP. Using this screen, Mt56(DE3) was isolated. It was demonstrated that the defining mutation in Mt56(DE3) results in an amino acid substitution in the T7 RNAP and it was shown that the accumulation level of the T7 RNAP in Mt56(DE3) was, in contrast to C41(DE3) and C43(DE3), similar to the one found in BL21(DE3). The mutant T7 RNAP has a considerable lower binding affinity for the T7 promoter that controls target gene expression. In turn, this lowers the target gene expression intensity and presumably harmonizes the expression rate of the recombinant gene with the capacity of the membrane protein biogenesis machinery, thereby increasing membrane protein production yields.
and biomass formation. Notably, Mt56(DE3) does not only produce more YidC-GFP than BL21(DE3), but also outcompetes the C41(DE3) and C43(DE3) strains for most membrane proteins tested. With this project we have demonstrated that E. coli can be evolved to produce membrane proteins to higher levels than the established membrane protein production strains and it seems unlikely that we have already reached the limits of the E. coli-based production of membrane proteins.

Besides directing the evolution of E. coli towards the improved production of recombinant membrane proteins, different strategies to engineer this production host have been used.

5.3.2 Engineering E. coli

The well-characterized biology of E. coli makes it possible to try to rationally engineer this bacterium for the enhanced production of membrane proteins. Chen et al. pioneered this engineering approach to improve the production of the E. coli magnesium transporter CorA by co-producing different membrane protein biogenesis factors and cytoplasmic chaperones [185]. Interestingly, the co-production of SRP and SecA did not increase yields of CorA, indicating that neither targeting of CorA to the Sec-translocon nor the translocation of periplasmic loops limit the production of CorA. In contrast, the co-production of the cytoplasmic DnaKJ-GrpE chaperone system reduced protein aggregation in the cytoplasm and improved yields of membrane inserted CorA about four times. Potentially, the folding of the large cytoplasmic domain of CorA is the rate limiting step in its biogenesis and the DnaKJ-GrpE chaperone system is crucial for the proper folding of this domain. Furthermore, the improved folding of
CorA may result in decreased aggregate formation, thereby improving the fitness of the cell.

Others have tried to promote the targeting of recombinant membrane proteins by deleting the tf gene encoding for the major E. coli chaperone TF, because TF competes with the membrane protein targeting factor SRP for binding to the translating ribosome [186]. Indeed the deletion of tf considerably increased the yields of the E. coli kinase ZraS and two archaeal rhodopsins. However, also here the co-production of SRP did not increase membrane protein production yields indicating that targeting of these membrane proteins to the Sec-translocon did not limit their production. On the other hand, it was shown that, if TF is absent, the DnaKJ-GrpE system interacts with TF substrates and promotes their proper folding [187]. Therefore, a deletion of tf may not improve the targeting of a recombinant membrane protein, but allows the DnaKJ-GrpE chaperone system to assist the folding of cytoplasmic domains of the membrane proteins produced. In the same study, it was shown that the co-production of YidC somewhat increased the correct folding of the two tested rhodopsins and this effect was attributed to the membrane protein chaperone function of YidC [188].

To increase the yields of the GPCRs CB1 and the bradykinin receptor 2 (BR2), Link et al. co-produced different membrane protein biogenesis factors as well as cytoplasmic chaperones [189]. Surprisingly, only the co-production of the integral membrane protease FtsH increased yields of the tested GPCRs markedly. Although it was not shown how the co-production of FtsH improved membrane protein production yields, it is tempting to speculate that FtsH clears misfolded membrane proteins that would otherwise compromise cell viability. Hence, it seems likely that the co-production of FtsH increased
Taken together, our limited knowledge of the biogenesis of membrane proteins makes it difficult to predict which gene should be co-expressed or deleted to improve the yield of a particular membrane protein. Therefore, Gialama et al. co-produced every *E. coli* protein together with the GPCR BR2 to identify factors that can increase membrane protein production yields [190]. Because the production of BR2 impairs biomass formation dramatically, they selected for co-produced proteins that resulted in an increased size of colonies formed on agar plates. Using this screening approach they showed that the co-production of DjlA, which is the membrane localized DnaJ like co-chaperone of DnaK, and RraA, which inhibits the degradation of mRNA, improves biomass formation when producing BR2 [191,192]. Notably, the co-production of DjlA as well as RraA enhanced not only the yield of membrane inserted BR2 per cell, but also improved the production yields of both pro- and other eukaryotic membrane proteins. Later it was shown that the beneficial effect of co-producing DjlA depends on its interaction with the chaperone DnaK [193]. Potentially, the co-production of DjlA increases the membrane localisation of DnaK that, in turn, facilitates the folding of cytoplasmic domains of recombinant membrane proteins. In the same study, it was also demonstrated that RraA improves membrane protein production yields through its interaction with RNase E, but the underlying mechanism was not elucidated [193].

Instead of manipulating levels of proteins involved in membrane protein biogenesis, others have tried to modulate the expression intensity of the gene encoding for the target membrane protein to improve membrane protein production yields. Often when producing a recombinant protein the target gene is expressed using a strong pro-
moter resulting in a high target gene expression intensity. However, for the production of membrane proteins it has been suggested that high target gene expression intensities can lead to the saturation of the membrane protein biogenesis machinery [194]. In the widely used protein production strain *E. coli* BL21(DE3) the expression intensity of the gene encoding the target membrane protein is particularly high. However, yields of membrane proteins produced in BL21(DE3) are usually low due to the aforementioned saturation of the membrane protein biogenesis machinery [194]. To modulate the gene expression intensity in BL21(DE3) and thereby avoid a saturation of the membrane protein biogenesis machinery, T7 lysozyme, which is a natural inhibitor of the T7 RNAP, can be co-produced from a constitutive promoter. T7 lysozyme decreases the T7 RNAP activity and in turn much less mRNA encoding for the recombinant protein is transcribed [195]. For a number of cases it has been shown that co-producing the T7 lysozyme results in considerably increased yields of membrane inserted protein [178,196]. Membrane proteins differ in *e.g.*, their number of transmembrane helices or periplasmic loops, and it is therefore likely that the optimal target gene expression intensity depends on the membrane protein produced. To regulate the expression intensity of a recombinant gene in *E. coli* BL21(DE3), the T7 lysozyme can also be co-produced using the well titratable rhamnose promoter [178]. In this Lemo21(DE3) system, the amount of L-rhamnose added to the culture medium correlates well with the amount of T7 lysozyme that is produced in a cell [196]. In turn, by varying the amount of L-rhamnose the expression intensity of a recombinant gene can be regulated and this allows to screen for an optimal expression intensity where the yield of any given protein is maximized [178,196,197]. Nowadays, Lemo21(DE3) is routinely used by many labs to produce recombinant proteins. As suggested, the
target gene expression intensity required to maximize production yields differs greatly from one membrane protein to another, presumably reflecting differences in their biogenesis. Notably, for a few target membrane proteins it has been shown that decreasing their gene expression intensity does not increase membrane protein production yields. Potentially, for these target membrane proteins a high target gene expression intensity does not lead to a saturation of the membrane protein biogenesis machinery.

Rather than indirectly regulating the expression intensity of a target gene, a recombinant protein can also be produced directly from the rhamnose promoter. Recently, it has been shown that the titratability of the rhamnose promoter is based on the consumption of the inducer L-rhamnose and that the protein production rate is independent of the L-rhamnose concentration [198]. To be able to truly modulate the protein production rate, Hjelm et al. introduced a frameshift in the gene that encodes for the L-rhamnose metabolising enzyme RhaB and deleted the gene encoding for the L-rhamnose transporter RhaT [198]. The absence of RhaB results in much lower concentrations of L-rhamnose that are required to fully induce the rhamnose promoter and the lack of RhaT appears to slow down the diffusion of L-rhamnose from the medium into the cell. This engineered rhamnose system made it possible to adjust the protein production rate of several membrane proteins, by varying the L-rhamnose concentration, thereby increasing production yields considerably [198]. Notably, as mentioned for the Lemo-setup, the optimal protein production rate where membrane protein yields are maximized differed from one protein to another.

Besides regulating the expression intensity of a recombinant gene, also the translation initiation rate of the recombinant gene can be altered to modulate the protein production rate [199]. Initiation of trans-
lation is known to be the rate limiting step in protein translation and it is affected by e.g., secondary structures formed by the mRNA and the GC content [200]. Mirzadeh et al. used a single step PCR-based mutagenesis method to create a library of different translation initiation sequences in conjunction to the gene encoding for the E. coli arabinose transporter AraH C-terminally fused to GFP [201]. By screening the library for highly fluorescent clones, they isolated different translation initiation sites that greatly improve the production of the AraH-GFP fusion. Using this workflow they could also improve the yields of several other membrane protein GFP-fusions. Interestingly, the optimal translation initiation sequence differed for different membrane proteins, indicating again that the optimal protein production rate depends highly on the recombinant protein produced.

Besides producing a recombinant membrane protein to levels as high as possible, extracting the desired protein from the cytoplasmic membrane is a major issue [202]. Usually, membrane protein extraction is done by solubilising the inner membrane with detergents. However, detergents are expensive and they can affect the folding of a membrane protein resulting in non-functional protein. To overcome this problem Mizrachi et al. developed a method where the recombinant membrane protein is produced in the cytoplasm of E. coli in small patches of native lipids [203]. To form these so-called nanodiscs, a variant of the human protein ApoAI is fused to the membrane protein of interest. ApoAI is an amphipathic protein that encloses the nanodiscs. The isolation of nanodiscs harbouring the recombinant membrane protein fused to ApoAI in a native-like membrane environment is much easier than extracting the produced membrane protein from the inner membrane.

Taken together, different engineering approaches have been used to increase the yields of different recombinant membrane proteins.
Nevertheless, the use of e.g., different *E. coli* strains, promoter systems and methods to monitor production levels makes it difficult to design a strategy that maximizes the production yields of a particular membrane protein.

5.4 Improving the production of proteins in the periplasm

Many recombinant proteins, like antibody fragments and hormones, are produced in the *E. coli* periplasm. These proteins require the formation of disulfide bonds to fold properly. Disulfide bond formation is unfavourable in the reducing environment of the cytoplasm, whereas it can be catalysed in the more oxidizing environment of the periplasm (see 4.2.1). Moreover, compared to the cytoplasm, the periplasm is a less complex protein mixture allowing an easier isolation of the produced protein from this compartment. Unfortunately, the production of recombinant proteins in the periplasm of *E. coli* often results in unsatisfactory yields.

However, the yield of a recombinant protein in the periplasm can be increased by e.g., improving the targeting of the recombinant protein to the Sec-translocon (see 5.4.1), aiding the folding of the recombinant protein in the periplasm by co-producing chaperones (see 5.4.2) and enhancing the stability of the recombinant protein by deleting genes that encode for periplasmic proteases (see 5.4.3).

5.4.1 **Improving the targeting of secretory proteins**

To reach the periplasm most secretory proteins are translocated across the inner membrane *via* the Sec-translocon (see 4.2.1). A recombinant protein can be N-terminally fused to a signal peptide that guides the recombinant protein either co- or post-translationally to the Sec-translocon (see 4.2.1) [67].
It has been shown that the signal peptide used can have a tremendous impact on the yield of a recombinant protein. For example, when testing for the efficient targeting of a phage coat protein fused to a to be secreted protein used in a phage display system, it was found that using co-translational signal peptides, compared to post-translational signal peptides, increased the levels of the target proteins on the surface of phage particles considerably [204]. However, a phage display depends on the membrane insertion of a phage coat protein, fused to the recombinant protein of interest, into the inner membrane. Because most *E. coli* membrane proteins are targeted co-translationally to the Sec-translocon, using co-translational signal peptides in combination with a phage display system may improve the membrane insertion of the phage coat protein, rather than enhancing protein secretion of the fusion partner. On the other hand, signal peptides that mediate post-translational targeting are used routinely to produce recombinant proteins in the *E. coli* periplasm [205].

Instead of using an *E. coli*-derived signal peptide to secrete a recombinant protein to the periplasm, signal peptides from other organisms can also be used to facilitate protein secretion in *E. coli*. The PelB signal peptide is probably the most prominent example of a heterologous signal peptide that is used to produce recombinant proteins in the *E. coli* periplasm. It originates from the pectate lyase B from *Erwinia carotovora* and already in 1987 PelB was shown to be secreted into the *E. coli* periplasm [206]. Although frequently used, it was never tested in a systematic manner if the PelB signal peptide mediates protein targeting more efficiently than other signal peptides.

More recently, it has been shown that the signal peptide from an organophosphorus hydrolase from *Flavobacterium* facilitates the transport of recombinant proteins to the *E. coli* periplasm [207]. Notably, this signal peptide can target the recombinant protein to the Sec-
and the Tat-translocon and this dual targeting may actually lower the risk of saturating the protein translocation capacity of the cell.

Cyclodextrin glucanotransferases (CGTases) from Gram-positive bacteria are routinely produced in the *E. coli* periplasm and their native signal peptides target these proteins to the Sec-translocon [208–210]. Furthermore, a CGTase signal peptide has been engineered with the aim to secrete a recombinant protein more efficiently into the extracellular medium. Introducing a helix breaking glycine residue in the *h*-region of a CGTase signal peptide increased the extracellular accumulation of the produced CGTase [211]. The authors did speculate that the introduced glycine promotes the degradation of the signal peptide after its cleavage from the CGTase. Because the accumulation of a signal peptide would disturb the inner membrane, its rapid degradation presumably increases the cell fitness. However, how increasing the fitness of the cell leads to an increased release of the produced CGTase into the extracellular medium remains unclear.

Efforts have been made to engineer the *E. coli* L-asparaginase II signal peptide to increase the yields of CGTases. First, the net charge of the *n*-region of this signal peptide was increased by mutating the three residues that follow the N-terminal methionine (EFF) to arginine residues [212]. This prevented the degradation of the produced CGTase in the periplasm. It seems likely that the mutated signal peptide does not target the produced CGTase to the Sec-translocon, but instead guides it to the Tat-translocon. Potentially, the translocation of the CGTase in an unfolded state by the Sec-translocon makes the recombinant protein more susceptible to degradation and degradation is diminished if the folded CGTase is translocated into the periplasm by the Tat-translocon. In a recent study, the same group tried to increase the extracellular production of a CGTase by engineering the modified L-asparaginase II signal peptide [213]. They found that mu-
tations at specific positions within the h-region of the signal peptide increased the extracellular accumulation of the recombinant CGTase up to 3-fold. However, this increased extracellular accumulation of the recombinant CGTase was probably due to dramatically increased cell lysis.

Others have tried to find an optimal signal peptide for the production of recombinant secretory proteins by using a “consensus sequence” based on several bacterial signal peptides [214]. Fusing this synthetic signal peptide to the granulocyte-macrophage colony-stimulating factor increased the yield of soluble recombinant protein per cell compared to the use of the E. coli OmpA or the PelB signal peptides. However, for the production of other recombinant proteins, using this synthetic signal peptide did not increase production yields when compared to the use of the other aforementioned signal peptides [215].

Taken together, signal peptides that are derived from E. coli secretory proteins or originate from other bacteria as well as engineered signal peptides can guide recombinant proteins to the E. coli periplasm. Unfortunately, the lack of mechanistic insight into why a certain signal peptide guides a certain recombinant protein more efficiently to the Sec-translocon makes it difficult, if not impossible, to rationally engineer signal peptides to improve protein production yields in the periplasm.

Therefore, we aimed to gain a more mechanistic understanding of the impact of the signal peptide used on protein production yields and host physiology. To this end, we studied the consequences of producing the model single-chain variable antibody fragment (scFv) BL1 genetically fused to two commonly used E. coli signal peptides; the DsbA signal peptide mediating co-translational targeting and the OmpA signal peptide mediating post-translational targeting to the
Sec-translocon (Paper II). Production of the scFv BL1 in BL21(DE3) had, irrespectively of the targeting pathway, a dramatic impact on the proteome composition of the cells resulting in e.g., protein misfolding/aggregation in the cytoplasm. The co-translational targeting of the scFv BL1 resulted in a severe saturation of the Sec-translocon capacity, which was accompanied by low biomass formation and low yields of the scFv BL1 in the periplasm. In contrast, the post-translational targeting of the scFv BL1 resulted not in a saturation of the Sec-translocon capacity and, compared to the production of DsbA-BL1, biomass formation as well as the yield of the secreted scFv BL1 were increased considerably. Furthermore, we found that upon the production of OmpA-BL1 the accumulation levels of chaperones and proteases in the cytoplasm were highly increased. Potentially, this helps to clear misfolded/aggregated proteins more efficiently from the cytoplasm, thereby improving the fitness of the cell.

Taken together, these findings indicate that the targeting pathway of a recombinant secretory protein, defined by its signal peptide, can have a tremendous impact on the physiology of the cell, thereby affecting biomass formation and protein production yields. With this study we have made a start to gain a more mechanistic understanding of the production of recombinant proteins in the periplasm of *E. coli*. Nevertheless, it would be interesting to expand this knowledge by studying the production of other recombinant proteins fused to other signal peptides in detail.

5.4.2 Increasing the yield and quality of secreted proteins

To improve the yield/quality of a secreted protein, different periplasmic chaperones have been co-produced together with the recombinant protein of interest. This strategy is based on the idea that
the recombinant protein titrates out chaperones in the periplasm leading to protein misfolding/aggregation, which can impair production yields and biomass formation.

It has been shown that the plasmid-based co-production of the periplasmic chaperones DsbA, DsbC, SurA and FkpA from their native promoters increases the yields of the soluble human plasma retinol-binding protein produced in the periplasm [216]. On the other hand, the co-production of these chaperones did not increase the yields of the extracellular carbohydrate recognition domain of the dendritic cell membrane receptor DC-SIGN per cell, but instead improved biomass formation considerably. Unfortunately, it remains to be elucidated why the co-production of these chaperones increased yields of a recombinant protein in one case, but only improved biomass formation if a different recombinant protein was produced. Furthermore, it was not tested if the observed effects are due to the co-production of a single chaperone or if a certain combination of chaperones was required.

Recently, it was shown that the co-production of DsbA or DsbC can improve the correct assembly of the antibody fragment (Fab) 3F3, but the beneficial effect of either DsbA or DsbC was dependent on the culture conditions [217]. Because DsbA oxidizes disulfide bonds and DsbC reduces them, these results indicate that the culture conditions can be important for the redox state in the periplasm.

Another study demonstrated that the co-production of either Skp or FkpA increases the yield of active scFv 3A21, but the co-production of DsbC had no positive impact on the folding of this scFv [218]. Notably, the co-production of DsbC together with Skp or FkpA had no positive effect on the production yields of the scFv 3A21, indicating that the beneficial effect of co-producing one chaperone can be neutralized by the co-production of another chaperone.
A different strategy to increase the yield of recombinant proteins in the periplasm is based on the assumption that high protein production rates saturate the Sec-translocon capacity, thereby impairing protein translocation. Therefore, the production rate of the recombinant protein has to be harmonized with the protein translocation capacity of the cell. One way to regulate the protein production rate is to control the expression levels of the recombinant gene encoding the protein of interest by using the *E. coli* Lemo21(DE3) strain. Lemo21(DE3) was initially engineered to improve the production of membrane proteins (see 5.3.2). However, Schlegel *et al.* used Lemo21(DE3) to harmonize the gene expression intensity of two recombinant secretory proteins, a variant of GFP and the scFv BL1 both targeted cotranslationally to the Sec-translocon by the DsbA signal peptide, with the protein translocation capacity of the cell [197]. By using Lemo21(DE3), the yield of active protein in the periplasm and biomass formation were increased considerably for both targets.

As part of my thesis, we aimed to characterize the consequences of optimizing the production of the scFv BL1 by using the Lemo-setup on the physiology of the cell. We found that, upon the non-optimized production of the scFv BL1, the proteome composition was changed dramatically compared to the proteome of cells harbouring the “empty” expression plasmid. In cells producing the scFv BL1 e.g., the accumulation levels of many proteins involved in protein synthesis were increased, translocation of secretory proteins was impaired resulting in protein misfolding/aggregation in the cytoplasm and an inefficient energy metabolism resulted in the decreased synthesis of ATP. Ultimately, the non-optimized production of the scFv BL1 leads to poor biomass formation and low protein production yields in the periplasm. Notably, cells that produced the secretory scFv BL1 under optimized conditions showed a proteome composition that was, be-
sides the scFv BL1, identical to the one of cells harbouring the “empty” expression plasmid. This indicates that optimizing the production of the secretory scFv BL1 using the Lemo-setup alleviates stress and therefore the yield of the recombinant protein in the periplasm is increased and biomass formation is improved. Although just shown for the production of one recombinant secretory protein, this implies that the physiology of the cell is not affected per se if the protein production rate of a secretory protein is harmonized with capacity of the Sec-translocon. Moreover, we anticipate that the production of proteins without or with low stress can be used as a basis to further engineer *E. coli* to *e.g.*, maximize the yields of recombinant proteins that are produced in the *E. coli* periplasm.

Besides regulating the expression intensity of a gene encoding a recombinant protein, modulating the translation initiation rate was also used to harmonize the production rate of a recombinant protein with the protein translocation capacity of the cell [219]. Interestingly, the translation initiation rate where protein production yields were highest was dependent on the recombinant protein produced.

Taken together, engineering *E. coli* to improve the levels and quality of a recombinant secretory protein by co-producing periplasmic chaperones appears to be an approach with unpredictable outcome. On the other hand, harmonizing the protein production rate with the translocation capacity of the cell by modulating either the transcription rate of the target gene or its translation initiation rate appears to be a more generic strategy to enhance the yield of a recombinant protein in the periplasm of *E. coli*. 
5.4.3 Increasing the stability of secreted proteins

The proteolytic degradation of a recombinant protein in the periplasm can impair its yield considerably and so far, more than 20 periplasmic proteases are annotated in the *E. coli* genome [144]. To avoid extensive degradation of a recombinant protein in the periplasm, genes encoding for these proteases can be deleted from the *E. coli* chromosome.

In 1988, Strauch and Beckwith devised a genetic screen to select for *E. coli* mutants with decreased proteolytic activity in the periplasm [146]. They fused the periplasmic protein alkaline phosphatase (AP) to the membrane protein Tsr in a way so that Tsr anchors AP at the periplasmic site of the inner membrane. In strains with a high proteolytic activity, AP will be cleaved from Tsr and diffuse into the extracellular medium through an outer membrane that is more permeable due to a deficiency of the lipoprotein Lpp. On agar plates containing the AP substrate 5-bromo-4-chloro-3-indolyl phosphate, the amount of AP released into the extracellular medium was monitored by comparing the size of the blue halos around the colonies. This setup allowed selecting for strains with a decreased proteolytic activity in the periplasm, because these mutants would secrete less AP into the medium resulting in a lower AP activity around the colony. Using this screen, the periplasmic protease DegP was identified and it was proposed that the absence of DegP could improve the yields of recombinant proteins in the periplasm.

Indeed later, *E. coli* strains were constructed lacking DegP and/or the periplasmic proteases Ppr and Tsp and the outer membrane protease OmpT, because at that time these were the only known secreted proteases in *E. coli* [220]. The authors claimed that by using these strains yields of different recombinant proteins could be increased. Recently, the strain deficient in all four proteases was used
to produce three different single domain antibodies, so-called nanobodies, because when the proteases were present extensive degradation of these nanobodies was observed [221]. Others have shown that Tsp degrades the light chain of a bivalent antibody fragment and that yields of the fully assembled antibody fragment in the periplasm were considerably increased if the proteases Tsp and DegP are absent [222].

Taken together, although only a few examples have been reported, deleting genes encoding for periplasmic proteases may be a promising approach to increase the yields of recombinant proteins in the periplasm. However, since periplasmic proteases are involved in the quality control of proteins (see 4.2.2), the absence of several proteases in the periplasm is likely to compromise the fitness of the cell. Nevertheless, by characterizing all known periplasmic proteases i.e., their substrate recognition motifs, it should be at some point possible to predict which protease may degrades a particular recombinant protein, thus allowing the use of specific protease deficient strains.

5.5 Display of recombinant proteins on the cell surface

Displaying recombinant proteins on the surface of *E. coli* has been used since the mid-1980s and it can be desirable because the genetic information encoding the displayed protein is maintained in the cell [223]. Therefore, protein display enables not only to create and screen a library of protein variants for a certain characteristic, but it also facilitates the identification of the desired protein variants by allowing the direct isolation of their respective genes [224]. Furthermore, protein display can be used to decorate the cell surface of *E. coli* with antigens, thereby allowing the development of vaccines (Paper IV).
To display a target protein on the surface of *E. coli* its genetic information is fused to a gene encoding for a carrier protein that contains all the information to transport the target protein across the inner and the outer membrane and finally anchors the target protein to the cell surface. Homologous and heterologous carrier proteins have been used to display recombinant proteins on the surface of *E. coli* [224]. *E. coli* OMPs are an obvious choice to serve as carrier proteins and different OMPs have been used successfully to display recombinant proteins on the cell surface [225–227]. However, if OMPs are used as carrier proteins the recombinant protein is often engineered into an extracellular loop, which limits the size of a recombinant protein that can be displayed. Alternatively, the recombinant protein can be fused to the extracellular C-terminus of the carrier OMP, thereby enabling the display of recombinant proteins larger than 50 kDa [228].

Besides OMPs, flagellae, fimbriae, pili as well as some lipoproteins have been used as carriers [224]. Especially the frequently used ice nucleation protein InaK from the plant pathogen *Pseudomonas syringae* appears to be a versatile display carrier [229]. The N-terminus of InaK is anchored to the outer membrane by an inositol moiety and the recombinant protein can be fused to the surface exposed C-terminus of InaK. Using this display system it was possible to present a 119 kDa large cytochrome that contained several cofactors on the surface of *E. coli* [230]. InaK has a signal peptide that is likely recognized by the Tat-translocon, which would allow the folded cytochrome containing the cofactors to cross the inner membrane. However, it remains elusive how such a large, folded protein is translocated across the outer membrane.

Also autotransporters (ATs) can be used to display recombinant proteins on the surface of *E. coli*. ATs consist of an N-terminal signal peptide that targets the AT to the Sec-translocon, a passenger do-
main that can be modified or partially replaced by a recombinant protein(s) and a C-terminal β-barrel translocator domain that is integrated into the outer membrane and facilitates the export of the passenger domain to the cell surface [231]. ATs can be produced at high levels on the cell surface and a variety of recombinant proteins, e.g., antigens and enzymes, have been presented on the cell surface using ATs as carriers [232,233]. Notably, the passenger domain of a subset of ATs can be cleaved from their C-terminal domain, thereby releasing it and any fusion partner into the culture medium [234].

As part of my thesis, we have used a variant of the AT hemoglobin protease (Hbp) to display different tuberculosis antigens on the surface of *E. coli* (Paper IV). To this end, we replaced different parts of the passenger domain of Hbp with up to three different *Mycobacterium tuberculosis* antigens, which were successfully displayed on the cell surface. Furthermore, Hbp was modified in such a way that the passenger domain is not cleaved from its C-terminal domain.

5.6 Export of recombinant proteins into the culture medium

The secretion of a recombinant protein into the culture medium facilitates its isolation, because *E. coli* exports hardly any endogenous proteins. Furthermore, the purification of a recombinant protein from the culture medium makes the elaborate process of collecting and disrupting cells unnecessary.

One of the few proteins that are secreted into the culture medium by *E. coli* is the outer membrane porin OmpF [235]. Therefore, OmpF was fused to β-endorphin and the fusion protein could be produced in high amounts in the culture medium [235]. Unfortunately, it was not addressed how OmpF, which is actually inserted into the outer membrane, can accumulate in the culture medium.
Later the same group found that the small osmotically-inducible protein OsmY is efficiently exported upon its recombinant production [236]. Furthermore, OsmY fused to either α-amylase or human leptin was also exported into the culture medium without causing extensive cell lysis. However, the mechanism of OsmY mediated protein export remains to be elucidated.

Others found that the small soluble protein of unknown function YebF accumulates in the culture medium and that YebF can be used as a fusion partner to efficiently export α-amylase and human interleukin-2 [237]. Translocation of YebF across the inner membrane is facilitated by the Sec-translocon and it was suggested that the OMPs OmpF and OmpC mediate the export of YebF into the extracellular space [238].

An alternative strategy to release recombinant proteins from the peri-plasm into the culture medium is to permabilise the outer membrane. This can be accomplished by e.g., adding compounds that destabilise the outer membrane like Triton X100, deleting proteins that stabilise the outer membrane and co-producing proteins that make the outer membrane more permeable [239].

Notably, recombinant proteins that are produced in the periplasm are found frequently in the culture medium and often it has been speculated that recombinant proteins somehow “leak” through the outer membrane [210,239–241]. However, in Paper II we show that a recombinantly produced secreted protein can be released into the culture medium by the formation of outer membrane vesicles (OMVs). OMVs are spherical particles with a diameter between 20 nm and 500 nm that bud off the outer membrane and OMV formation is known to be an important feature of many Gram-negative bacteria [242]. Moreover, the formation of OMVs has been described as a stress response and it seems likely that the stress caused by the production of
recombinant proteins stimulates the formation of OMVs. Hence, it is tempting to speculate that the recombinant secretory proteins, which have been found in the culture medium, are exported via OMVs.

As part of my thesis we used a hypervesiculating E. coli strain to release OMVs, coated with different tuberculosis antigens, into the culture medium (Paper IV). To display the antigens on the cell surface we used the AT Hbp as a carrier protein (see 5.5). It has been suggested that OMV-based vaccines may trigger a stronger immune response than the purified antigen, thereby making the use of adjuvants unnecessary.

Taken together, different strategies can be used to export a recombinant protein into the culture medium. Considering the enormous industrial interest in producing recombinant proteins in the culture medium, it is surprising that only little is known about the different mechanisms of protein export. Therefore, studying the different mechanisms of protein export will hopefully pave the way to further engineer carrier proteins and strains that facilitate the efficient production of recombinant proteins in the culture medium.
Conclusions and future perspectives

*E. coli* has been widely and successfully used to produce recombinant proteins. However, the production of proteins in the *E. coli* cell envelope *i.e.*, the production of membrane and secretory proteins, often leads to unsatisfactory yields. Therefore, the objectives of this doctoral thesis were (i) to investigate what can limit the production of recombinant proteins in the cell envelope of *E. coli* and (ii) to explore how this information can be used to design strategies to improve the yield of recombinant proteins in this compartment.

In the first study, we directed the evolution of *E. coli* BL21(DE3) to improve its production of integral membrane proteins (*Paper I*). Previously, it has been shown that producing recombinant proteins in BL21(DE3) often results in the accumulation of C41(DE3)/C43(DE3)-like mutations, which decrease the synthesis of the T7 RNAP driving the expression of recombinant genes (see 5.3.1). However, it has been shown that the production of the integral membrane protein YidC does not result in the accumulation of the aforementioned mutations. Therefore, we decided to use the efficient production of YidC as a selection criterion to isolate new BL21(DE3)-derivatives with improved membrane protein production characteristics. Using this approach we isolated Mt56(DE3), a strain that outcompetes the standard membrane protein production strains C41(DE3) and C43(DE3) for most membrane proteins tested. We have shown that in Mt56(DE3) the expression intensity of the recombinant gene is strongly reduced not due to the decreased synthesis of the T7 RNAP.
but due to a mutation in the gene encoding for the T7 RNAP that lowers its affinity to the T7 promoter. Presumably, this difference in the mechanism of reducing the expression intensity of the recombinant gene increases membrane protein production yields for a considerable number of membrane proteins even further. In conclusion, this study shows that directing the evolution of *E. coli* can be used to isolate improved membrane protein production strains. Here we focused only on the production of prokaryotic membrane proteins, but yields of eukaryotic membrane proteins produced in *E. coli* are particularly low. Therefore, it would be interesting to see if also *E. coli* strains can be isolated that produce high yields of eukaryotic membrane proteins.

In the second and third study we investigated what can limit the production of a recombinant secretory protein in *E. coli* BL21(DE3). It is observed frequently that changing the signal peptide that guides a secretory protein to the Sec-translocon can have a tremendous impact on biomass formation and protein production yields in the periplasm, but so far this phenomenon has not been studied in a systematic manner. Therefore, we aimed to characterize the impact of using different signal peptides that target the recombinantly produced scFv BL1 in different ways to the Sec-translocon on cell physiology and protein production yields (*Paper II*). To this end, the scFv BL1 was genetically fused to two different signal peptides; the well-characterized DsbA signal peptide that mediates co-translational targeting to the Sec-translocon and the frequently used OmpA signal peptide that directs the scFv BL1 post-translationally to the Sec-translocon. We found that the co-translational targeting of the scFv BL1 has a tremendous impact on the proteome composition of the cell and *e.g.*, saturates the capacity of the Sec-translocon, resulting in heavily impaired biomass formation and low protein production yields. Although also the post-translational targeting of the scFv BL1 affected
the proteome composition of the production host tremendously, it did not saturate the capacity of the Sec-translocon resulting in hardly affected biomass formation and increased protein production yields. This indicates that targeting a recombinant secretory protein posttranslationally to the Sec-translocon can be used to improve protein production yields in the periplasm of *E. coli*. However, it needs to be investigated if these observations are dependent on both, the produced recombinant secretory protein and the specific signal peptides used to mediate its targeting to the Sec-translocon.

Besides changing the signal peptide that targets a recombinant secretory protein to the Sec-translocon, it has been shown that the BL21(DE3)-based Lemo-setup can be used to harmonize the production rate of co-translationally targeted recombinant proteins with the capacity of the Sec-translocon, thereby not/hardly affecting biomass formation and increasing protein production yields. However, the consequences of optimizing the production of a recombinant protein in the periplasm using the Lemo-setup have not been studied in detail. Therefore, we aimed to characterize the effects of optimizing the production of the co-translationally targeted scFv BL1 on the cell physiology ([Paper III](#)). Surprisingly, the proteome of cells that produced the scFv BL1 under optimized conditions was - besides the produced target protein - identical to the proteome of cells that produced no recombinant protein. This indicates that a recombinant protein can be produced in the periplasm without affecting the physiology of *E. coli*. We imagine that this optimized production condition can be used as a starting point for further engineering, *e.g.*, increasing the stability of a recombinant protein by inactivating periplasmic proteases. Optimizing the production of the scFv BL1 was also accompanied by the OMV-mediated release of properly folded BL1 into the culture medium. Because it appears that only low amounts of OMVs were formed, en-
hancing the vesiculation could be a promising approach to increase the yields of a secreted protein in the culture medium, thereby facilitating its isolation.

In the last project, the AT Hbp was used to coat *E. coli* OMVs with different tuberculosis antigens in order to create the basis for a recombinant vaccine (Paper IV). I contributed to this study by demonstrating that the isolated OMVs formed a homogeneous population. This is important for the isolation of these vesicles using filtration methods, which are routinely used in industrial setups. Hopefully, future studies will explore the versatility of this vesicle-based vaccine platform further.

Taken together, I have studied what limits the production of recombinant proteins in the cell envelope and improved the production yields of membrane and secretory proteins. These research efforts (i) resulted in the isolation of Mt56(DE3), a new *E. coli* strain for the production of integral membrane proteins, (ii) shed light on the impact of the targeting pathway used to direct a recombinant protein to the Sec-translocon, (iii) showed that an optimized production of a recombinant secretory protein does not compromise the physiology of the production host and (iv) demonstrated that *E. coli* can produce a homogeneous population of antigen-coated OMVs. Hopefully, this doctoral thesis will pave the way to further improve the production of recombinant proteins in the *E. coli* cell envelope.
Populärvetenskaplig sammanfattning

Membran separerar celler från sin omgivning. Alla celler innehåller DNA, vilket innehåller koden med den information som krävs för att konstruera alla de små maskiner, såkallade proteiner, som gör att celler kan leva. Proteiner som inte fungerar som de ska är ofta orsak till sjukdomar.

Därför är det viktigt att studera proteiner så att vi kan öka vår förståelse av hur livet fungerar på en molekylär nivå och på så sätt förbättra vår förståelse av sjukdomar.

Eftersom att det är svårt att studera proteiner i den komplexa miljön som råder i cellen behöver vi rena fram det proteinet vi vill studera. För att få tillräckligt höga mängder av proteinet kan vi producera det med hjälp av bakterier. För att lyckas med det, att skapa en proteinfabrik av en bakterie, behöver vi introducera just den delen av DNA:t som kodar för det protein vi vill producera in i bakteriens DNA. Den bakterie som oftast används som proteinfabrik är *Escherichia coli*. Alla proteiner går emellertid inte att producera i tillräckliga mängder i *E. coli*. Dessa är t.ex. de proteiner som måste produceras i cellens ytterhölje, d.v.s proteiner som måste sättas in i - eller transporteras igenom - cellens membran.

Syftet med mina studier är att (i) undersöka vilka fakorer som begränsar produktionen av proteiner i cellens ytterhölje i *E. coli* och (ii) att utforska om - och hur - dessa begränsande faktorer kan överkommas.
Under den första delen av mina studier isolerade och karaktäriserade vi en ny *E. coli* stam som kan producera höga mängder av sådana proteiner som sitter i membranet. Eftersom dessa proteiner är involverade i många viktiga processer i cellen och deras dysfunktion ofta leder till sjukdom är det viktigt att studera dem. Vi hoppas därför att vår nya stam kan användas för att överkomma flaskhalsen i produktionen av denna typ av proteiner.

Under den andra delen av mina studier undersökte vi produktionen av proteiner som behöver transporteras igenom membranet. Vi upptäckte då att sättet som dessa proteiner är guidade på kan ha en väldigt stor effekt på produktionsnivåer och även på vilken typ av stress som cellerna utsätts för när de producerar dessa proteiner. De här observationerna är viktiga för att kunna förstå vad det är som begränsar produktionsnivåerna och för att kunna designa nya strategier att öka produktionsnivåerna av denna typ av proteiner.

I nästa studie upptäckte vi att om produktionen av ett icke-*E. coli* protein, som transporteras igenom membranet, optimeras på ett visst sätt, upplever cellen inte någon stress alls. Den här observationen är överraskande eftersom man antagit att produktion av alla proteiner i *E. coli* i någon grad stör fysiologin i cellen och därav orsakar stress.

Den sista delen av mina studier handlade om att försöka utveckla ett vaccin mot tuberkulos. Först producerade vi olika tuberkulosantigener på ytan av *E. coli*. Efter det modifierade vi *E. coli* så att den släpper ifrån sig små sfärer, såkallade vesiklar, som är beklädda med dessa antigener. Dessa vesiklar kan sedan användas som ett potentiellt mer effektivt vaccin.
Acknowledgements

First of all, I want to thank my supervisor Jan-Willem for giving me the chance to do my PhD in Stockholm, it has been a great time! Thank you also for sharing so much "insight information". I truly learned a lot during this years and I am grateful for this experience.

A special thanks goes to Gunnar for being my co-supervisor and to Stefan and Pia for doing such a great job in managing the PhD program at DBB.

A big thanks goes to my lab mates - I would not have survived without you! Susan for helping me to get a good start in the lab, introducing me to “Pärk” and for sharing your enthusiasm even at 7 am; Anna for suffering and celebrating together, I hope the latter continues :) Zhe, it was a great experience to visit china and an honour to be the serious guy at your wedding! Alex, it was a pleasure to be your lab mate, your roommate (next time get the damn gong!) and your "sambo". Anna - the Spanish girl- make sure that you don’t become too Swedish! Henry, thanks for passionate discussions and I hope that one day you will get your Mercedes! David, it was great to see how a one-man company works. And thanks to my students, Daphne, João and Erik, for their contributions to the projects.
Thanks to all current and former members of the von Heijne and the Daley lab for fantastic quality time at and outside of DBB, especially: Nurzian - see you on the court or on a playground, Rickard, Johannes, Florian, Pilar, Renuka- let’s get the hell out of science, Grant, Theresa, Jörg, Claudio - I am always up for a rematch- and Pat - when are you coming for badminton again?

Rob, thank you for great input on the projects. Chenge, Johan, Dan and Hao, we should definitely have more BBQs!

Anne und Manfred, thanks for fun game nights and a great trip to Gothenburg!

To our collaborators:
Jimmy, thank you for your patience in the mass spec project. Dirk-Jan Slotboom and his team, thanks for having me in Groningen, I had a great time! Herman, thank you for a long and ongoing collaboration.

Finally, I want to thank my little family, Nani and Dante, for reminding me every day what is most important in life.
References


Struct Mol Biol 18, 345–351.


42 Tang Y-C, Chang H-C, Roeben A, Wischnewski D, Wischnewski N,


UmuD and UmuD’ subunits of UmuD/D’ mediate tethering and
substrate processing by the ClpXP protease. *Proc Natl Acad Sci U S A*
100, 13219–13224.

response regulator RssB controls stability of the σ5 subunit of RNA

58 Mika F & Hengge R (2005) A two-component phosphotransfer network
involving ArcB, ArcA, and RssB coordinates synthesis and proteolysis
of σ5 (RpoS) in *E. coli*. *Genes Dev* 19, 2770–2781.

use distinct modes of binding to inhibit the RssB-dependent turnover of
RpoS (σ5) by ClpXP. *Front Mol Biosci* 2, 15.

recognition signals ensure LexA destruction after DNA damage. *Genes
Dev* 17, 1084–1089.

of the *Escherichia coli* Transcriptional Regulator ZntR. *J Bacteriol* 189,
3017–3025.

contacts direct high-priority recognition of the tetrameric transposase-

63 Keiler KC, Waller PRH & Sauer RT (1996) Role of a Peptide Tagging
System in Degradation of Proteins Synthesized from Damaged

Substrate specificity of the DnaK chaperone determined by screening

65 Orfanoudaki G & Economou A (2014) Proteome-wide Subcellular
Topologies of *E. coli* Polypeptides Database (STEPdb). *Mol Cell
Proteomics* 13, 3674–3687.


67 Valent QA, Scotti PA, High S, de Gier J-WL, von Heijne G, Lentzen G,
Wintermeyer W, Oudega B & Luirink J (1998) The *Escherichia coli*
SRP and SecB targeting pathways converge at the translocon. *EMBO
J* 17, 2504–2512.

68 Randall LL (1983) Translocation of Domains of Nascent Periplasmic
Proteins Across the Cytoplasmic Membrane Is Independent of

*Biochim Biophys Acta* 1808, 851–865.

70 Samuelson JC, Chen M, Jiang F, Möller I, Wiedmann M, Kuhn A, Phillips
GJ & Dalbey RE (2000) YidC mediates membrane protein insertion in


111 Zwizinski C & Wickner W (1980) Purification and Characterization of


125 Kadokura H & Beckwith J (2002) Four cysteines of the membrane
protein DsbB act in concert to oxidize its substrate DsbA. *EMBO J* 21, 2354–2363.


Biotechnol Bioeng 102, 357–367.


240 Ukkonen K, Veijola J, Vasala A & Neubauer P (2013) Effect of culture medium, host strain and oxygen transfer on recombinant Fab antibody
fragment yield and leakage to medium in shaken *E. coli* cultures. *Microb Cell Fact* **12**, 73.

