From protein production to genome evolution in *Escherichia coli*

Susan Schlegel
List of Papers

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

Paper I
Tuning *Escherichia coli* for protein production.

Paper II
Optimizing membrane protein overexpression in the *Escherichia coli* strain Lemo21(DE3).

Paper III
Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels.

Paper IV - manuscript
Reconstructing the evolution of BL21 into C41 reveals rapid adaptability of a widely used *E. coli* strain.
**Schlegel S**, Genevaux P, Slotboom DJ, and de Gier J-W. *Manuscript*

* These authors contributed equally.
Reprints were made with permission from the publishers.
The author contributed to the following publications:


# Contents

## List of Papers

## List of Figures

## Introduction

1. *Escherichia coli* ........................................... 1
2. Different compartments of *Escherichia coli* ................. 2
   2.1 The cytoplasmic and the outer membrane .................. 2
   2.2 The periplasm ................................................ 3

## Proteins

3. Classification and terms used in this thesis ............... 5
4. The life cycle of a protein .................................. 5
   4.1 Synthesis ..................................................... 5
   4.2 Protein folding (and unfolding) .......................... 6
      4.2.1 Trigger factor .......................................... 7
      4.2.2 DnaKJ/GrpE ............................................. 7
      4.2.3 GroEL/ES ............................................... 8
      4.2.4 Additional folding modulators ....................... 8
   4.3 Protein degradation ......................................... 9

5. Biogenesis of membrane- and secretory proteins ............ 10
   5.1 Targeting to the membrane .................................. 10
      5.1.1 The targeting signal ................................... 10
      5.1.2 The co-translational pathway ......................... 12
      5.1.3 The post-translational pathway ....................... 13
   5.2 Membrane insertion and translocation ..................... 14
      5.2.1 The Sec-translocon ..................................... 14
      5.2.2 SecA ..................................................... 15
      5.2.3 YidC ..................................................... 16
      5.2.4 SecDFYajC, Syd, and YidD ............................ 17
   5.3 Quality control and degradation of membrane proteins ... 17
   5.4 Folding, quality control, and degradation in the periplasm 18
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The different compartments in <em>E. coli</em></td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Biogenesis of MPs and secretory proteins</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>The Sec-translocon</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>Disulfide bond formation in the periplasm</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>The <em>E. coli</em> heat shock response</td>
<td>23</td>
</tr>
<tr>
<td>6</td>
<td>The <em>lac</em> and <em>rha</em> promoters and BL21(DE3)</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>The Lemo21(DE3)-strain</td>
<td>52</td>
</tr>
<tr>
<td>8</td>
<td>The selection of C41(DE3) from BL21(DE3)</td>
<td>56</td>
</tr>
</tbody>
</table>
Introduction

1 Escherichia coli

*Escherichia coli* is a Gram-negative bacterium, which is part of the microbial flora of the intestine of warm-blooded organisms. *E. coli* is everything from the pathogenic, enterohemorrhagic *E. coli* (EHEC) to the harmless strains used in routine laboratory work. The laboratory strains used to date probably originate from two different ancestors that were isolated almost a hundred years ago [1]. *Bacillus coli* is the earliest traceable ancestor of the B-lineage of *E. coli*, which includes BL21 and derivatives. *E. coli* K-12 was isolated in 1922 and gave rise to the K-lineage, including, e.g., MG1655 and DH5α [1].

Laboratory *E. coli* strains have been subject to extensive manipulations throughout the years, and even strains that belong to the same lineage exhibit many differences [2; 3]. It is not surprising that the differences are even more prominent when strains are compared between the two lineages [3]. Therefore, even though the basic principles described in the following sections generally apply to *E. coli*, one should be cautious when directly comparing the results of studies if these were obtained using different strain backgrounds, even if strains are highly related.

*E. coli* is a popular organism for a variety of applications, ranging from studies aiming at the understanding of fundamental biological and biochemical processes to the production of biological substances like proteins and small organic compounds. During my Ph.D. studies I was particularly interested in the use of *E. coli* for protein production (papers I to III). The aim of my work was to identify factors hampering the production of especially membrane- but also secretory proteins and to use this knowledge to improve protein production. An important part of my work was the analysis of a mutant strain that is well suited for protein production. During the course of my studies it became apparent that the genome of this strain exhibited changes that were not solely important in the light of protein production, but in addition documented its adaptation to stresses that prevailed during its isolation. The results of this study are presented in paper IV.

The following chapters aim to give a general overview of topics that are
important for the work presented in papers I to IV. The first chapters will focus on aspects relevant to the biogenesis of proteins. Thereafter, I will give an overview of some topics that are relevant for protein production in *E. coli*. Finally, I will address some of the mechanisms that are involved in mutagenesis and, more generally, the ability of *E. coli* to evolve in response to different stresses.

2 Different compartments of *Escherichia coli*

The *E. coli* cell can be divided into four compartments: the cytoplasm, the cytoplasmic membrane, the periplasm, and the outer membrane (figure 1). The cytoplasm is the ‘center’ of the cell. It contains all genetic information and hosts a wide range of metabolic and biosynthetic processes. The cytoplasmic membrane, periplasm, and outer membrane together are referred to as the *E. coli* cell envelope [4]. The cell envelope is not only essential to contain the cytoplasm and maintain cell shape, but also provides a barrier that protects cells from an often harmful environment. This does by no means imply that the envelope isolates the cell from its surroundings. Membranes are naturally permeable to small, hydrophobic molecules. In addition, whilst hydrophobic molecules and larger structures cannot diffuse through membranes *per se*, the presence of proteins and protein complexes allows the controlled exchange of these substances with the environment [5].

The following sections will concentrate on the composition and some properties of the constituents of the *E. coli* cell envelope (for review see [4]). For the sake of simplicity, the different constituents will be addressed separately. However, it should be stressed that they are heavily intertwined and act together in many capacities.

2.1 The cytoplasmic and the outer membrane

Simplified, a membrane can be considered as a network of various lipids and proteins, in some instances with carbohydrate moieties attached to lipids. Membrane lipids are composed of a hydrophilic head-group and a rather hydrophobic tail. Within a membrane, lipids are oriented such that their head-groups face the aqueous environment, whereas the tails are separated from it. This gives the membrane its typical bilayer structure [5] (see figure 1). The lipid portion of the *E. coli* cytoplasmic membrane consists mainly of the glycerophospholipids phosphatidyl-ethanolamine (75%), phosphatidyl-glycerol (20%), and cardiolipin [6]. Remarkably, *E. coli* is able to maintain the composition and fluidic properties of the membrane rather constant at different temperatures by modifying the hydrophobic tails of the phospholipids [6–8].
The lipid portion of the inner leaflet of the outer membrane consists mainly of glycerol-phospholipids, whereas the outer leaflet is composed of glycolipids (lipopolysaccharides) [4, 6]. Thus, in contrast to the cytoplasmic membrane, the outer membrane is asymmetric. Proteins of both membranes can be (partially) embedded in the membrane, or attached to it by a lipid anchor or non-covalent interactions.

For a long time, the unanimously accepted model describing the lipid-protein-network of a membrane was the fluid-mosaic model [9]. According to this model the membrane is a ‘sea’ of lipids with proteins floating therein. Recent research suggests a much more complex organization, where both lipids and proteins are organized in complex and defined areas and influence each other’s shape and distribution [10, 11]. This organization can be influenced by structural components outside the plane of the membrane, e.g., the (bacterial) cytoskeleton and the peptidoglycan layer. In eukaryotic organisms, the presence of membrane rafts has gained a lot of attention [12]. These are local ‘assemblies’ of proteins in areas that are highly enriched in specific lipids. Similarly, also bacterial membranes contain structured domains [13].

2.2 The periplasm

The periplasm contains the peptidoglycan layer that shapes and strengthens the cell [14] (figure 1). This compartment is not only important from a structural point of view, but it is also rich in terms of biochemical processes. It contains proteins involved in the uptake and transport of nutrients, protein folding and degradation, peptidoglycan synthesis etc. In addition, the periplasm is much more directly exposed to environmental insults as compared to the cytoplasm. Recent research suggests that the viscosity of the periplasm differs only marginally from that of the cytoplasm [15].
Figure 1: The different compartments in *E. coli*. The *E. coli* cytoplasm is surrounded by two membranes; the cytoplasmic membrane and the outer membrane. In between the two membranes lies the periplasm, which contains the stabilizing peptidoglycan layer. All compartments contain a variety of proteins. The lipid-portion of the cytoplasmic membrane consists mainly of phospholipids that are oriented such that the membrane exhibits a characteristic bilayer structure. The outer membrane is asymmetric, with the inner leaflet composed of phospholipids whereas the outer leaflet is composed of lipopolysaccharides (LPS). LPS consist of a lipid A-moiety with oligosaccharides attached to it. The O-antigen is a repetitive glycan polymer that is characteristic for a certain *E. coli* strain and highly immunogenic. The *E. coli* strain BL21 and its derivatives, that were used throughout my studies, lack the O-antigen. For a detailed description see section 2 and subsections. Figure modified after [4].
Proteins

3 Classification and terms used in this thesis

At the most basic level, one can distinguish between soluble proteins and membrane proteins [16]. Membrane proteins can be discriminated, based on their association with the membrane, into integral (including lipoproteins) and peripheral membrane proteins. Integral membrane proteins cannot be extracted from the membrane by means that leave the lipid bilayer intact, whereas peripheral membrane proteins can be extracted by treatment with reagents that interrupt non-covalent interactions (e.g., urea, high salt). Integral membrane proteins can be further classified based on the secondary structure of their membrane-spanning, proteinaceous domain(s), into α-helical and β-barrel [16]. In *E. coli*, β-barrel membrane proteins are confined exclusively to the outer membrane [17]. Hereafter, outer membrane proteins will be referred to as OMPs. During my studies I was primarily interested in membrane-spanning proteins of the cytoplasmic membrane (hereafter referred to as MPs), and proteins that are destined for the periplasm and beyond (secretory proteins).

4 The life cycle of a protein

Independent of the nature of a protein, its life cycle can roughly be divided in synthesis, folding (including mis- and refolding), function, and degradation. Below I will discuss these processes with a special emphasis on cytoplasmic proteins. Components and mechanisms specific to membrane- and secretory proteins will be addressed in separate sections.

4.1 Synthesis

At the most basic level, all proteins are a chain of amino-acids. The sequence of the amino-acids in this chain is encoded by its corresponding gene. To make a protein, the information contained in a gene must first be transcribed
into messenger (m)RNA with the help of RNA polymerase (RNAP). Ribosomes then translate the information contained in the mRNA into an amino-acid chain. The single amino-acids are delivered to the ribosome by transfer (t)RNAs [18].

The *E. coli* ribosome consist of a 30S subunit (composed of 21 proteins and the 16S ribosomal (r) RNA), and a 50S subunit (composed of 33 proteins, 23S, and 5S rRNA). Together, the two subunits form the bacterial 70S ribosome. With respect to translation, three distinct sites can be discriminated, an A- (aminoacyl-tRNA entry), a P- (peptidyl-tRNA binding) and an E- (exit) site. Translation is initiated at a specific sequence in the mRNA, the ribosome binding site (RBS; also called Shine-Dalgarno-sequence) (reviewed by [19]). First, the 30S subunit is recruited to the mRNA such that the initiator codon of the mRNA (mostly AUG [20]) is positioned in the P-site. Subsequently, an initiatory tRNA (methionyl-tRNA) is recruited. The 50S subunit completes assembly of the ribosome. Newly arriving tRNAs, charged with an amino-acid, enter the ribosome at the A-site and the amino-acid is attached to the methionine via a peptide bond. Now, the ribosome moves one codon further along the mRNA and the discharged tRNA is released from the E-site. This elongation continues until a stop-codon is reached, which has no complementary tRNA [19]. The growing polypeptide chain exits the ribosome via the ribosome exit tunnel.

Under optimal conditions, *E. coli* ribosomes can join approximately 22 amino-acids per second [21]. Recent research suggests that the nature of the amino-acids in the growing polypeptide chain influences the speed at which polypeptide synthesis proceeds, *i.e.*, positively charged amino-acids are likely to slow down translation as they interact with negative charges in the ribosome exit tunnel [22].

After (or during) translation, the polypeptide chain folds into its correct conformation. The earliest steps towards the native conformation of a protein already occur in the ribosome exit tunnel. The dimensions of the tunnel (80-100 Å long and 10-20 Å wide, *e.g.*, [23, 24]) allow the nascent chain to adopt a certain degree of secondary or, possibly, tertiary structure (*e.g.*, [25–27]). However, the protein can only adopt its fully folded conformation after emergence from the ribosome exit tunnel.

### 4.2 Protein folding (and unfolding)

Anfinsen elegantly demonstrated that all information needed for the correct folding of a protein is, in principle, contained in its amino-acid sequence by showing that denatured ribonuclease A can refold into a fully functional state in solution [28]. Yet, in a cell the polypeptide chain encounters a very crowded...
environment and it is believed that, within this environment, many proteins require the assistance of molecular chaperones in order to reach their native fold. Chaperones are molecules that interact transiently with a polypeptide chain. They assist the protein in its search for the native conformation, but are not part of the final structure. Chaperones that have so far primarily been associated with the recovery/removal of aggregated or misfolded proteins will be discussed in section 6.

Chaperones have been identified in all domains of life and several are essential for viability. The majority of chaperones identified so far are proteins. Here, I will therefore focus on this class of chaperones. However, it should be kept in mind that molecules other than proteins, like lipids and RNA, have been shown to influence the final structure of a protein and, consequently, exhibit chaperone-like effects.

The three major chaperones/chaperone systems in the E. coli cytoplasm that are involved in the folding of newly synthesized proteins are trigger factor (TF), DnaKJ/GrpE, and GroEL/ES. Notably, these systems are able to promote substrate folding in a defined, sequential manner.

4.2.1 Trigger factor

TF is an ATP-independent chaperone which is highly abundant in the E. coli cytoplasm and displays chaperone- and peptidyl-prolyl-isomerase (PPIase) activity. TF binds, with similar affinity, to non-translating and translating ribosomes and meets the nascent polypeptide chain when it emerges from the ribosome exit tunnel. Thus, it has been suggested that the majority of newly synthesized polypeptide chains may encounter TF. Interestingly, compact polypeptide structures (presumably α-helices) present within the ribosome exit tunnel disfavor TF-binding to the ribosome. It should also be noted that a recent study suggests that TF only contacts the ribosome when approximately 100 amino-acids of the polypeptide chain have been synthesized.

TF interacts with unfolded or partially folded polypeptide chains and protects them from aggregation. Recent research demonstrates that TF is also able to unfold partially folded polypeptide chains. Notably, TF activity is not restricted to ribosome-bound polypeptide chains, but can also aid the formation of protein complexes.

4.2.2 DnaKJ/GrpE

DnaK is the E. coli homologue of the mammalian Hsp70. It does not operate on its own but requires the co-chaperone DnaJ (Hsp40) and, mostly, the nucleotide exchange factor GprE (reviewed, e.g., by). Starting point for one
cycle of substrate binding is ATP-bound DnaK with an open substrate binding pocket, a state that allows high substrate exchange rates. The substrate is delivered by the co-chaperone DnaJ [47,48]. When substrate-bound DnaJ docks onto DnaK, the ATPase activity of DnaK is stimulated. ATP-hydrolysis induces conformational changes that increase the affinity of DnaK to its substrate and lead to closing of the substrate binding domain. Subsequently, ADP-release, stimulated by GrpE, leads to release of the substrate which may now refold into its active conformation (e.g., 34). DnaK-binding-motifs are enriched in hydrophobic amino-acids that are buried in the final fold of the protein [49]. Recently, Calloni et al. identified about 700 polypeptide chains that are able to interact with DnaK in vivo and defined a central role for DnaKJ in the E. coli chaperone network [50].

4.2.3 GroEL/ES

In contrast to DnaKJ/GrpE and TF, the GroEL/ES chaperone system is essential for viability of E. coli [51]. The central part of the chaperone is composed of 14 GroEL subunits that form a 'double-doughnut'-shaped structure with two opposite folding chambers [52,53]. Each of these chambers is able to support the folding of substrates that are up to 57 kD in size [54]. Once a substrate has entered an un-occupied cavity, ATP binds to each subunit and the heptameric GroES-ring will 'close' the cavity [55]. Structural changes induced by ATP- and GroES-binding gradually turn the hydrophobic GroEL-interior, ideally suited to bind to exposed hydrophobic substrate residues, into a largely hydrophilic surface that promotes sequential substrate release and folding [56–58]. The main purpose of the GroEL/ES chamber is to provide an environment in which the polypeptide chain can fold undisturbed. However, it has been demonstrated that substrate binding to GroEL and, subsequently, ATP-binding, promotes local expansion or unfolding of the polypeptide chain [58,59].

The speed at which GroEL/ES operates is determined by the rate of ATP hydrolysis [60]. Once hydrolysis is complete, GroES and ADP dissociate from the complex and the protein is released. This process is triggered by substrate/ATP binding to the previously unoccupied chamber [61,62]. Notably, substrates may have to undergo multiple rounds of binding and release before they reach their final conformation. Numerous proteins, some of which are essential, have been shown to be completely or partially dependent on GroEL/ES for proper folding [63–65].

4.2.4 Additional folding modulators

It is worth mentioning that, in addition to the above mentioned chaperones, other components have been shown to play a role in the folding of polypeptide
chains, including components that have traditionally been linked to translation (e.g.,[66–68]). As indicated above, chaperones are not only involved in the de novo biogenesis of proteins, but are just as important for the recovery of misfolded proteins and the removal of undesirable protein aggregates. Owing to the complexity of the regulatory components involved, these processes will be addressed in section 6.

4.3 Protein degradation

The life of a protein ends with its degradation by proteases and, subsequently, by peptidases. Proteolysis is not only important to remove unwanted proteins (i.e., proteins that are terminally damaged/aggregated), but is also essential for the regulation of many cellular processes. Here, I will address the process of proteolysis and the key players involved in the *E. coli* cytoplasm. Just as cytoplasmic chaperones, these proteins are not only important under regular (i.e., non-stressed) conditions, but are also essential components of global responses that counteract stress. Global stress responses will be discussed in section 6.

In *E. coli*, five major proteolytic systems that act on proteins in the cytoplasm have been identified so far: ClpXP, ClpAP, HslUV, Lon, and FtsH. All of these belong to the AAA+ type family of ATPases. Members of this family have been found in all organisms, albeit to varying extent [69; 70]. They consist of at least one ATPase domain and a peptidase domain that reside either on two separate polypeptides (ClpXP, ClpAP, HslUV) or are contained in one single polypeptide chain (Lon, FtsH) [69; 71]. The membrane-bound protease FtsH is the only enzyme that is indispensable for viability and targets both cytoplasmic proteins and MPs [72–75].

All AAA+ proteases seem to form barrel-shaped complexes with the proteolytic activity located on the inside, even though the number of subunits needed to form the (active) barrel varies [76–81]. To prevent the unregulated hydrolysis of proteins, the access to the proteolytic chamber is restricted by a narrow entry pore. The ATPase domain, also referred to as unfoldase, is most commonly located on top of the barrel and unfolds the substrate protein in an ATP-dependent manner, even though alternative configurations have been described for Lon [81]. The unfoldases ClpX and HslU have also been shown to play an important role in regulating the access of a substrate to the proteolytic chamber by modulating the pore diameter [82; 83]. In the proteolytic chamber, the peptidase hydrolyses exposed peptide bonds. The resulting peptide fragments are released into the cytoplasm where they are further degraded.

It is intriguing that only few substrates have been identified for these proteases so far (e.g., [71; 74; 84]). Nonetheless, given their abundance it seems likely that they act on a wide variety of substrates.
5 Biogenesis of membrane- and secretory proteins

As for all *E. coli* proteins, also the biosynthesis of MPs and secretory proteins starts in the cytoplasm. In order to reach their final location, they need to be targeted to the membrane, where they either have to be integrated and folded (MPs) or translocated across into the periplasm (secretory proteins).

5.1 Targeting to the membrane

In *E. coli*, proteins can be targeted to the membrane in a co-translational or a post-translational fashion. The commonly accepted view is that co-translational targeting is mediated by the signal recognition particle (SRP). Most MPs, and a handful of secretory proteins, are thought to be targeted *via* this pathway [85–87]. The majority of secretory proteins is currently believed to be targeted post-translationally in a SecB/SecA dependent fashion (compare section 5.1.3).

The SRP-dependent, co-translational targeting pathway and the SecB/SecA-dependent, post-translational targeting pathway converge at the Sec-translocon, that mediates both the insertion of MPs into and the translocation of secretory proteins or periplasmic domains of MPs across the membrane. It should be noted that there is some evidence of overlap between the two pathways. Thus, a strict discrimination may turn out to be an over-simplification. However, independent of which targeting pathway is used, the targeted protein has to be maintained in a largely unfolded state in order to be compatible with translocation/insertion.

Another pathway that targets proteins in a post-translational fashion is the twin-arginine translocation (TAT)-pathway [85; 88]. In contrast to the Sec-dependent pathways described above, proteins following the TAT-pathway can be folded prior to translocation *via* the TAT-translocase. Most TAT-dependent proteins identified so far contain co-factors that need to be acquired in the cytoplasm, which necessitates their folding in this compartment [88]. Since the TAT-pathway was not part of this thesis, it will not be discussed further. For more information on this pathway see, e.g., [88].

5.1.1 The targeting signal

In order to be targeted to the membrane proteins must be equipped with a signal that is recognized by the different targeting components and funnels them into one of the targeting pathways (see section 5.1). The existence of such a signal was first demonstrated in eukaryotic systems by Gunter Blobel and his colleagues, who established that 'the information for segregation of a translation product (into microsomes from dog pancreas) is encoded in the mRNA' [91].
Figure 2: Biogenesis of MPs and secretory proteins. MPs are primarily targeted to the cytoplasmic membrane in a co-translational fashion, mediated by SRP and its membrane-bound SRP-receptor FtsY. At the membrane, the SRP-bound RNC-complex is transferred to the protein-conducting Sec-translocon, which mediates the insertion of TMs into the cytoplasmic membrane (CM) and the translocation of soluble loops across. The translocation of large periplasmic loops requires energy provided by the ATPase SecA. YidC may assist the Sec-translocon in the insertion process, but can also act as an integrase on its own. MPs can be targeted to YidC via the SRP-pathway or a hitherto unknown mechanism. It has also been proposed that mRNAs encoding MPs can be targeted directly to the membrane. Non-MP encoding mRNAs are possibly discriminated from MP encoding mRNAs due to an enrichment of uracils in regions encoding hydrophobic amino-acids [89]. The targeting of secretory proteins to the Sec-translocon occurs primarily via the post-translational pathway, which involves SecB, SecA and, possibly, TF (see sections 4.2.1 and 5.1.3). At the Sec-translocon, the nascent chain is translocated into the periplasm, which requires energy provided by SecA and the proton-motive force. The targeting signal is removed by leader peptidase (Lep). The SecDFYajC-complex may have a role in assisting the biogenesis of MPs and secretory proteins. OM = outer membrane. Figure modified after [90].
Now it is evident that for most MPs the first transmembrane domain (TM), also referred to as signal anchor sequence, serves as targeting signal that initiates SRP-dependent, co-translational targeting upon emergence from the ribosome exit tunnel (see, e.g., [85; 90] and figure 2). TMs are enriched in hydrophobic amino-acids and usually devoid of polar and charged residues, which would be highly unfavourable in the hydrophobic core of the membrane. The signal anchor sequence is not cleaved but remains part of the MP.

Secretory proteins are typically equipped with a cleavable, N-terminal signal sequence, consisting of a positively charged N-terminus, a hydrophobic core (the H-segment) that has a high propensity to form an α-helix, and a C-terminus that is enriched in small, polar amino-acids [92]. At the very C-terminus, a signal peptidase recognition motif is localized that allows removal of the signal sequence upon translocation across the membrane. Generally, a highly hydrophobic N-terminal targeting signal is likely to mediate co-translational targeting, whereas a less hydrophobic signal funnels proteins into the post-translational pathway, even though this discrimination is not absolute [86; 93]. The signal sequence of the periplasmic protein DsbA, which was employed for the studies presented in paper III, targets proteins in an SRP-dependent manner [87]. Notably, some MPs are, just as secretory proteins, equipped with a cleavable targeting signal (e.g., [94; 95]).

Different studies indicate that the nature and properties of the protein determine the choice of the targeting pathway: A fast-folding protein may preferably be targeted in a co-translational fashion to avoid premature folding, whereas slower folding proteins can be kept translocation-competent even upon post-translational targeting [86; 87; 96]. In this respect it is also worth mentioning that the signal sequence itself can be important to keep the protein in a translocation-competent state (e.g., [97]).

5.1.2 The co-translational pathway

Co-translational targeting of ribosome-nascent chain complexes (RNCs) relies on the SRP-targeting pathway. This pathway comprises the SRP and its receptor FtsY [98], both of which are GTPases. The *E. coli* SRP consists of a protein subunit, the fifty-four-homologue Ffh (named after its mammalian homologue 54), and the 4.5S RNA. The genes encoding SRP cannot be deleted, but SRP-levels can be reduced to a very high extent without notably affecting cell viability [99] [100].

SRP associates with non-translating ribosomes with an affinity of 50-100 nM [101] [102]. The presence of a polypeptide chain within the ribosome exit tunnel results in an increase in affinity to approximately 1 nM. The ribosome-associated SRP scans the nascent chain for the presence of a targeting signal...
as it emerges from the ribosome. If an appropriate signal (i.e., a signal-anchor sequence or a sufficiently hydrophobic signal sequence) is recognized SRP binds to it, which stabilizes the RNC-SRP-complex [102]. Emergence of a non-SRP substrate leads to loss in SRP-affinity and dissociation from the ribosome [101]. Recent research suggest that this initial scanning process provides the first of several quality control steps to ensure that only true SRP-substrates are targeted via the SRP-pathway [103].

At the membrane, the SRP-RNC complex interacts with FtsY, a process that requires both SRP and FtsY to be in a GTP-bound state. An alternative theory suggests SRP-independent ribosome targeting to the membrane via nascent FtsY [104,105]. Studies performed by Braig et al. have demonstrated the existence of 'empty', membrane bound SRP-FtsY complexes (i.e., complexes that are not in contact with a RNC), supporting the idea of SRP-independent targeting of the RNCs [106]. Whichever the case, it is clear that the majority of FtsY is bound to the cytoplasmic membrane [98] and that SRP bound to the correct RNC has an increased affinity to FtsY [101].

Interaction of SRP-FtsY initiates GTP-hydrolysis and induces structural rearrangements that allow efficient transfer of the RNC to the Sec-translocon [107]. Upon GTP hydrolysis, SRP and FtsY are released, freeing them for a new round of targeting (e.g., [85]). The RNC stays attached to the Sec-translocon and translocation proceeds. It should be noted that SRP also has been shown to target MPs to the insertase/translocase YidC [108,110] (see section 5.2.3).

5.1.3 The post-translational pathway

Most proteins that are to be secreted across the cytoplasmic membrane utilize the post-translational pathway [85,90]. Importantly, 'post-translational' does not strictly mean that the translation process is complete, but that a major portion of the polypeptide chain has been synthesized [111].

One of the major challenges of this pathway is to maintain the polypeptide chain in a translocation-competent state, i.e., unfolded. The first chaperone to accomplish this task may be TF (see section 4.2.1). Subsequently, the cytoplasmic chaperone SecB may bind to various stretches of an emerging polypeptide chain to prevent premature folding and aggregation [112]. The latter notion has supported the term of 'SecB-dependent, post-translational targeting'. Notably, SecB does not bind to the signal sequence itself [112], and proteins may also be targeted to the Sec-translocon without involving SecB (e.g., [113]). Work by Lui et al. suggests that, besides chaperones, the signal sequence of the secretory protein itself may have a role in maintaining the translocation competence of post-translationally targeted proteins [97].
The prevailing view is that the (SecB-bound) nascent chain is subsequently transferred to the ATPase SecA associated with the membrane. SecA interacts with the Sec-translocon and mediates the transfer of the polypeptide chain to the Sec-translocon, where it is translocated across the membrane (e.g., [113, 114]).

An increasing body of evidence suggests that the classical view on post-translational targeting may require refinement (e.g., [115, 116]). It has been observed that SecA binds to translating ribosomes, with an affinity that increases in the presence of native substrate. Together with the notion that SecA interacts with signal sequences of non-SRP-substrates, these data promote the idea that SecA interacts co-translationally with substrates prior to funneling them into the post-translational pathway [116]. Nonetheless, for the remainder of this thesis, I will adhere to the classical term of ’post-translational targeting pathway’.

5.2 Membrane insertion and translocation

5.2.1 The Sec-translocon

The Sec-translocon is a protein conduction channel that mediates the biogenesis of both MPs and secretory proteins. The core unit of the Sec-translocon consists of the integral MPs SecY (α), SecE (γ) and SecG (β) in a 1:1:1 ratio [117]. SecY and SecE are highly conserved across all kingdoms and, consistently, have been shown to be essential, whereas SecG is dispensable [85]. SecY levels are reduced in SecE-depleted cells due to degradation of un-complexed SecY by the protease FtsH [118, 119].

The crystal structure of the Sec-complex from Methanococcus jannaschii was solved in 2004 [117] (compare figure 3). The ten TM helices of Secα (SecY) that make up the channel are organized in two halves, consisting of five helices each. Secγ (SecE) makes contact with helices from both halves and thus appears to ‘clamp’ the sites together. Overall, the channel displays an hour-glass shape with a wide opening to the cytoplasmic side and a narrow constriction towards the middle. A plug closes the channel in the resting state and must be displaced to allow translocation of a polypeptide [117, 120]. It has been suggested that TMs or signal sequences partition laterally into the lipid bilayer via a ‘gate’ in the translocon [117, 121]. The idea of a lateral gate is, e.g., supported by studies that the TM of FtsQ can be cross-linked simultaneously to SecY and to lipids [122].

Research by Cymer and von Heijne suggests that for MPs with multiple TMs, in addition to hydrophobic interactions with lipids, also hydrophilic interactions with preceding TMs contribute to the partitioning. This could aid membrane insertion as well as folding of these proteins [123].
Figure 3: The Sec-translocon. A The two halves of SecY (blue and red) are 'clamped' together by SecE (gray). B TM2b, 3, 7, and 8 have been suggested to make up the lateral gate, that allows partitioning of a TM or a signal sequence into the lipid bilayer. A model TM and its potential lateral movement are indicated (magenta). The plug is depicted in green. Reprinted with permission from [117].

As mentioned initially, also secretory proteins employ the Sec-translocon. Once the signal peptidase recognition motif emerges on the periplasmic side of the membrane or upon completion of translocation, the *E. coli* signal peptidase Lep removes the signal peptide. This allows the release of the translocated protein into the periplasm [124]. The optimal cleavage site for Lep has been found to be AXA, where X is a small, non-polar amino-acid [124]. In the periplasm, the polypeptide chains of soluble proteins fold into their final conformation. OMPs are further targeted to the outer membrane, where they integrate and fold. The remaining signal peptides are degraded by signal peptidases, *e.g.*, RseP [125].

Besides the SecYEG-core, assessor components have been identified that (may) aid the Sec-translocon in the biogenesis of MPs and secretory proteins. These comprise of SecA, YidC, the SecDFYajC-complex, Syd, and YidD.

5.2.2 SecA

The ATPase SecA is a peripheral component of the Sec-translocon, which is essential for viability. In addition to mediating the transfer of post-translationally targeted proteins to the Sec-translocon (see section 5.1.3), SecA fuels the translocation process of secretory proteins by multiple rounds of ATP-binding and hydrolysis [113, 126]. Translocation also requires the proton-motive force [127].

As opposed to the translocation of secretory proteins, the translocation of periplasmic domains of MPs is generally supported by the 'pushing' ribosome. Energy provided by SecA is required if large soluble domains need to be translocated across the membrane [87, 128, 129], and for the translocation
of the C-termini of single spanning TMs [129]. Notably, if the translocation capacity of the Sec-translocon is insufficient, *E. coli* 'responds' by increasing SecA-levels [130].

### 5.2.3 YidC

YidC is an essential MP, and, just as the Sec-translocon, highly conserved between different organisms (Oxa1 in mitochondria, Alb3 in chloroplasts) [131]. In *E. coli*, YidC is involved in membrane insertion as well as folding of MPs and can assist the assembly of MP complexes (see below). In agreement with YidC being essential, its depletion severely compromises the composition of the cell envelope. This results in cell envelope stress, as shown by both transcriptome- and proteome analysis [132–135].

Amongst the proteins that at least partially require YidC for proper insertion into the cytoplasmic membrane are *F*<sub>o</sub>*a* [136], *CyoA* [137] [138], *NuoK* [139], and *ProW* [131]. It has been suggested that YidC receives TMs of MPs at the lateral gate and aids their insertion into the membrane [131, 140–142]. This notion was recently strengthened by a study demonstrating that YidC can be cross-linked to the helices that make up the lateral gate of the Sec-translocon *in vivo* [143].

MPs whose insertion has been shown to be solely YidC dependent are *F*<sub>o</sub>*c* [144] and the MscL-channel [109]. Recent *in vitro* experiments suggest that even multi-spanning MPs can be efficiently inserted by YidC as long as they lack long, periplasmic loops (*i.e.*, do not require the action of SecA) [110]. As mentioned above, targeting of nascent chains to YidC can occur *via* SRP or a hitherto unknown mechanism. *In vitro* experiments indicate that YidC may even function as post-translational insertase [145].

The proteins LacY and MalF do not require YidC for insertion, but rather for folding and stability in the membrane [146–147], and YidC is required for the assembly of the maltose transport complex [146]. It should be noted that YidC has also been shown to interact with the *E. coli* autotransporter Hbp, and may thus have a role in the biogenesis of secretory proteins (*e.g.*, [148]).

In paper I and II, YidC is used as a model protein to study MP production. Given the role of YidC in MP biogenesis, one may ask whether the results obtained are influenced by a possible activity of the produced YidC. However, it has been reported that non-functional versions of YidC elicit essentially the same responses when produced in *E. coli* as compared to the functional protein [149] and, consequently, YidC may be used as a model protein.
5.2.4 SecDFYajC, Syd, and YidD

Compared to the above described components, SecDFYajC, Syd, and YidD are far less well studied. It has been suggested that the SecDFYajC complex, which can interact with SecY [150] and YidC [140], provides the link between SecYEG and YidC [151]. However, YidC can contact SecY also in a SecDFYajC depletion background [143]. It has also been proposed that the SecDF-complex has a role in proton-motive force dependent protein translocation across the membrane [152], and that SecD facilitates the release of secretory proteins into the periplasm upon translocation [153].

The membrane-associated protein Syd interacts with SecY and destabilizes the SecYEG complex in vitro, and has been suggested to have a regulatory role in protein export [154; 155]. The protein YidD is not essential for cell viability, but deletion of the gene encoding YidD reduced the levels of the YidC-dependent proteins CyoA and F_o [156].

The network of factors involved in MP insertion and folding might be even more complex than already appreciated. It should be noted that the folding of cytoplasmic or periplasmic loops may be assisted by soluble chaperones. Since most MPs in E. coli function as part of homo- or hetero-oligomeric complexes, proteins must not only integrate into the membrane but, subsequently, associate with the respective interaction partners. The assembly of individual MPs into complexes has not been part of my Ph.D.-studies and will therefore not be discussed.

5.3 Quality control and degradation of membrane proteins

Unfortunately, not much is known about quality control mechanisms that act on MPs in E. coli. One of the best-studied proteins implicated in the quality control of MPs is the AAA+ protease FtsH described in section 4.3). FtsH is anchored to the membrane via two N-terminal TMs with the catalytically active domain located in the cytoplasm [73; 74; 79]. FtsH degrades MPs after extracting them from the membrane [157 [158]. As described in section 4.3) substrates can only enter the proteolytic chamber of FtsH after they have been unfolded, a process that requires ATP. Interestingly, it was shown that GFP equipped with a signal targeting it for degradation by FtsH, or a signal recognized by both FtsH and ClpXP, cannot be degraded by FtsH as long as the GFP moiety is folded [159]. In contrast ClpXP readily degraded folded GFP equipped with the respective degradation signal. This led to the suggestion that FtsH has only a rather weak unfolding activity and will not be able to act on targets that contain a stably folded domain [159]. This notion agrees well with the role of FtsH in the degradation of presumably unstable MPs like uncomplexed SecY [119] or F_o [160]. Two accessory proteins, HflK and
HflC, modulate the activity of FtsH [161]. Cross-linking experiments support a concerted action of FtsH, HflKC, and YidC in MP quality control [162].

So far, two other proteases have been shown to act on MPs in *E. coli*, HtpX and GlpG. Both proteases are far less well studied than FtsH. Like FtsH, HtpX is anchored to the membrane via two TMs, with the active site residing in the cytoplasm [163]. HtpX has been shown to degrade SecY [164]. Since HtpX does not possess an intrinsic ATPase domain it appears unlikely that the protein by itself can extract substrates from the membrane prior to degradation.

The polytopic membrane protein GlpG is a member of the rhomboid family of proteases [165]. As such, it is able to hydrolyse substrates in the plane of the membrane [124; 166]. GlpG has been shown to cleave truncated versions of the MP MdfA [167], and a TM of LacY incorporated in a model MP [165].

It is likely that, besides factors that are integral to the cytoplasmic membrane, also cytoplasmic and periplasmic chaperones/proteases play a role in MP quality control. Further research is needed to develop our understanding of this complex process.

5.4 Folding, quality control, and degradation in the periplasm

The folding of periplasmic proteins involves mechanisms quite similar to the folding of cytoplasmic proteins, with two major differences: Whereas the cytoplasm is a reducing environment, the periplasm is oxidizing and supports the formation of structural disulfide bonds; and factors acting in the periplasm operate in an ATP-independent manner as this compound is exclusively localized to the cytoplasm. The periplasm contains a variety of folding catalysts, chaperones and proteases. As outlined in the next sections, these functions are in some instances contained within the same polypeptide chain.

5.4.1 The disulfide bond formation system

The native fold of several periplasmic proteins contains intra-molecular disulfide bonds. For a recent review on disulfide bond formation see, *e.g.*, [168; 169].

The disulfide-bond oxidoreductase DsbA is the initial catalyst for disulfide bond formation (see figure 4). DsbA is a soluble, periplasmic protein which contains one disulfide bond. It oxidizes free cysteines of a substrate polypeptide, which results in the formation of a disulfide bond in the substrate. Reduced DsbA is reactivated with the help of the MP DsbB. Electrons are transferred from DsbA via DsbB to ubi- or menaquinone. If the initial oxidation results in the formation of ‘wrong’ disulfide bonds, proteins cannot reach their native conformation; they are misfolded. Such proteins can be sensed by the
Disulfide bond formation in the periplasm. DsbA catalyzes the first step in the formation of disulfide bonds, the oxidation of substrate cysteines. Electrons are then transferred via DsbB and ubiquinone (Q) to a terminal electron acceptor (e.g., oxygen). Wrongly formed disulfide bonds can be reduced via DsbC and DsbD. The electrons that are required for the reduction stem from reducing equivalents in the cytoplasm and are transferred to DsbD via thioredoxin (TrxA). For a detailed description of the Dsb-system refer to section 5.4.1. Modified after [168].

Similar to DsbC, the DsbG protein has been demonstrated to be involved in disulfide bond isomerization. However, it was proposed that substrates of DsbG are (partially) folded rather than misfolded [170]. Recent research suggests that DsbG primarily prevents the oxidation of single cysteine residues that are important for protein function in concert with DsbC [171].

5.4.2 Periplasmic chaperones and folding catalysts

In addition to the Dsb system, the periplasm harbours various other components involved in protein folding. Here, I will briefly describe those that are best characterized, or that have been associated with the production of secretory proteins, namely SurA, Skp, FkpA, PpiD, and Spy (compare section 12.3). Importantly, even proteins that so far have mainly been linked to the production of OMPs in vivo have been employed for the production of soluble proteins in

---

Figure 4: Disulfide bond formation in the periplasm. DsbA catalyzes the first step in the formation of disulfide bonds, the oxidation of substrate cysteines. Electrons are then transferred via DsbB and ubiquinone (Q) to a terminal electron acceptor (e.g., oxygen). Wrongly formed disulfide bonds can be reduced via DsbC and DsbD. The electrons that are required for the reduction stem from reducing equivalents in the cytoplasm and are transferred to DsbD via thioredoxin (TrxA). For a detailed description of the Dsb-system refer to section 5.4.1. Modified after [168].
the periplasm.

SurA is a chaperone with PPIase activity, which has been suggested to assist the folding of OMPs and the assembly of OMP-complexes [172] [173]. A proteomics analysis showed that only eight of the 64 identified OMPs were negatively affected in a SurA deletion/depletion mutant. Out of these, only LptD could be identified as true SurA substrate [174].

The chaperone Skp is thought to interact with parts of the polypeptide chains of OMPs to keep them in a state compatible with subsequent insertion/assembly events [175]. In addition, Skp may act as a periplasmic targeting factor for OMPs [176]. Substrate proteins may interact in a sequential manner with Skp (or SurA) and the chaperone/protease DegP [177] (see section 5.4.3). A recent proteomic study suggests that SurA is the major factor for the folding of OMPs, but can mostly be compensated for by Skp [178]. Skp may also have a role in preventing the aggregation of periplasmic proteins [179].

FkpA is a periplasmic chaperone and PPIase [180] [181]. Data on the in vivo-activity of FkpA on endogenous targets are unfortunately scarce. FkpA prevents misfolding of an unstable variant of the maltose binding protein [180], and was recently shown to interact with the passenger domain of an autotransporter from E. coli O157:H7 [182].

PpiD is a MP whose catalytic activity is localized in the periplasm [183]. It has been suggested to exhibit both chaperone and PPIase activity [184], even though the latter is controversial [183] [185]. PpiD was previously thought to be primarily involved in the folding of OMPs, but may even have a more general role by supporting the folding of newly translocated proteins [184] [186].

Spy is a periplasmic protein that only recently has been shown to exhibit chaperone activity in vitro [187]. Its levels are much increased upon the production of an unstable, soluble protein in the periplasm. Spy appears to act on soluble proteins rather than OMPs [187], and thus may prove a valuable tool for the production of secretory proteins (see section 12).

5.4.3 Periplasmic proteases

To date, 20 different proteases have been identified in the periplasm of E. coli [188]. Here, I will limit myself to the High temperature requirement A (HtrA) proteases DegP, DegQ, and DegS, which are well studied and have been shown to be involved in quality control. HtrA proteases are serine proteases and have a minimum of one peptide binding domain. As the name suggests, this class of proteases plays an important role during the prokaryotic heat shock response (see section 6.1).

DegP has both chaperone- and protease activity [189] [191]. The protease activity is important under conditions that induce misfolding of proteins, e.g.,
elevated temperatures \[191\]. In the absence of substrate (unfolded proteins), DegP exists as a dimer of trimers, with the proteolytically active sites at the inside and blocked by the opposite trimer. Binding of substrate protein appears to induce and stabilize higher oligomeric assemblies up to 30mers that are proteolytically active \[192; 193\]. The minimal unit of DegP capable of cleaving a substrate has been shown to be a trimer \[194\].

Based on structural data and the notion that the protease DegP only acts on unfolded polypeptide chains, Krojer et al. \[192\] postulate that proteins that are able to re-fold within the proteolytic cavity escape proteolysis, whereas those that fail to do so are cleaved. Thus, the chaperone activity would be attributed to the ability of DegP to provide a protective environment. Assembly and activation of protease activity requires the presence of distinct recognition motifs in the substrate protein \[193\].

DegQ is another member of the Deg-family of proteases. Its gene was identified in 1996, and it was proposed that DegQ acts as a protease in the periplasm \[195\]. In contrast to \textit{degP}, expression of \textit{degQ} is not induced at elevated temperatures \[195\]. In the presence of substrate, DegQ is converted from trimers \[196\] or hexamers \[197; 198\] to 12- and 24mers that efficiently degrade substrate proteins or allow their folding \[198\]. DegQ and DegP may have partially overlapping functions \[195\], which appears plausible given their similarities in mechanism and architecture. Just as DegP, DegQ can function as protease and chaperone \[197\]. Research by Bai et al. suggests that DegQ displays higher chaperone and less protease activity, whereas the opposite is the case for DegP \[196\].

The essential protease DegS \[195\] is the third member of this family. DegS forms trimers, but has so far not been shown to assemble into higher oligomeric structures and is, in contrast to DegP and DegQ, attached to the cytoplasmic membrane \[124\]. In the absence of substrate (\textit{i.e.}, misfolded OMPs), the DegS active side is blocked by an internal loop \[199\]. Substrate binding to the peptide binding domain releases this blockage and allows DegS to become proteolytically active \[200\]. DegS is part of the envelope stress response system and, as such, plays an important role in maintaining the integrity of the cell envelope \[201\] (see section 6.2).

6 Stress responses and regulators

Cells have to be able to react quickly and coordinated to various kinds of stress. An adequate stress response requires the concerted action of different cellular factors. To achieve this, global regulators of gene expression, like \(\sigma\) factors, are indispensable. \(\sigma\) factors are exchangeable subunits of the \textit{E. coli} RNA polymerase (P) core enzyme that determine its promoter specificity and, thus,
can direct the expression of specific genes [202]. Also, two-component regulators, like the CpxAR system, have been shown to enhance expression of genes involved in stress responses. Genes whose transcription is regulated in concert by the same transcription factor or transcriptional regulator are referred to as 'regulon'.

Of particular interest to my thesis are stress responses that are induced upon increased protein misfolding and/or aggregation, which is a common side-effect when attempting to produce proteins in E. coli, as demonstrated, e.g., in paper I for the production of MPs. These are the $\sigma^{32}$ or $\sigma^H$ ($\text{rpoH}$) mediated heat shock response, and the $\sigma^{24}$ or $\sigma^E$ ($\text{rpoE}$) and CpxAR mediated envelope stress responses. It should be noted that there is extensive overlap between the target genes of these stress responses. Also, the transcription of many of the target genes, like those encoding the chaperones DnaK and GroEL/ES and the protease Lon, is performed by the RNAP-$\sigma^{70}$ holoenzyme, since the proteins they encode have important house-keeping functions [203].

6.1 The heat shock response

As the name implies, this response has initially been linked to the (transient) exposure of cells to high temperature, which leads to increased protein misfolding and aggregation. $\sigma^{32}$, the central regulator of this response, coordinates the expression of a multitude of genes encoding proteases, chaperones, and accessory proteins, many of which are essential even under non-stressed conditions [203–205]. Amongst these are the afore-mentioned chaperones DnaKJ/GrpE and GroEL/ES, and the proteases FtsH, Lon, and HtpX.

Also the AAA+ protein ClpB and the small heat shock proteins IbpA and B are part of the $\sigma^{32}$ regulon. ClpB has been recognized for its ability to resolve protein aggregates together with the DnaKJ/GrpE-system [206]. ClpB is thought to act by threading the substrate through its central cavity in an ATP-dependent manner, thereby mediating disaggregation and substrate unfolding [207]. This is in agreement with the mechanism previously described for the unfoldase components of the AAA+ proteases (see section 4.3). IbpA and B assist chaperones in the disassembly of protein aggregates and have been shown to act in concert with ClpB and DnaKJ/GrpE [208]. Intriguingly, approximately 25% of the genes of the $\sigma^{32}$ regulon encode MPs [205]. To understand the role of all these genes and their respective proteins in the recovery of E. coli from heat shock, further research is needed.

In order to prevent futile gene expression and stress responses, $\sigma^{32}$ levels are strictly regulated (see figure 5). $\text{rpoH}$ is constitutively transcribed and, even though regulation on a transcriptional level has been described (e.g., [209]), regulation appears to occur primarily after transcription. The $\sigma^{32}$ mRNA has
Figure 5: The *E. coli* heat shock response. Depicted are some of the components/factors involved in regulation of $\sigma^{32}$ and in the $\sigma^{32}$-mediated heat shock response. Levels of $\sigma^{32}$ are regulated on a transcriptional, translational, and post-translational level. $\sigma^{32}$ regulates the expression of numerous genes, amongst them those encoding the protease FtsH, the chaperones DnaKJ, GroEL/ES, and ClpB, and the small heat shock proteins IbpA/B. Misfolded proteins can be degraded by proteases (like FtsH) or unfolded and refolded with the assistance of chaperones, like DnaKJ/GrpE or GroEL/ES. IbpA/B, ClpB, and DnaKJ/GrpE can form a functional triad to resolve protein aggregates. For more information see section 6.1.

been shown to form secondary structures that limit access of the ribosome to the translation initiation site at permissive temperatures. An increase in temperature results in an increase in translation rate due to the resolution of these structures (*e.g.*, [210]). On a post-translational level, DnaKJ/GrpE and GroEL/ES have been shown to regulate $\sigma^{32}$ activity. Different models have been proposed to explain the post-translational regulation by these chaperones, amongst them the sequestration of $\sigma^{32}$ by chaperones that are not occupied with substrate (*i.e.*, mis- or unfolded proteins) or chaperone-mediated degradation of $\sigma^{32}$ [204, 211–213]. Proteases like FtsH and HslUV degrade $\sigma^{32}$, thereby reducing its levels [214, 216]. Thus, $\sigma^{32}$ and the proteins encoded by genes in the $\sigma^{32}$ regulon are intimately connected.

### 6.2 Envelope stress responses

Envelope stress responses are induced by factors that negatively affect the integrity of the cell envelope. The $\sigma^{24}$ and CpxAR mediated responses are prob-
ably the best characterized ones in *E. coli*. Both systems respond to the presence of misfolded or mislocalized proteins in the cell envelope (*e.g.*, [217]).

6.2.1 The σ^{24} envelope stress response

The first evidence that σ^{24} is involved in regulating responses to envelope stress came from an experiment where an OMP was produced at high levels. This led to upregulation of σ^{24}, probably due to the accumulation of OMPs in the periplasm [218]. Just as for σ^{32}, the levels and activity of σ^{24} are strictly regulated [201]. In unstressed cells, σ^{24} is tethered to the cytoplasmic membrane *via* the anti-σ factor RseA (regulator of sigma E) and thus not active. The presence of misfolded proteins in the *E. coli* periplasm triggers a cascade that involves several proteolytic events [201]. First, the protease DegS cleaves RseA on the periplasmic side of the membrane. RseA is further cleaved by RseP, which results in release of the cytoplasmic, σ^{24} bound domain into the cytoplasm. Degradation of the cytoplasmic domain of RseA by the ClpXP protease leads to release of σ^{24}, which now can associate with the RNAP core enzyme and change its promoter specificity.

The σ^{24} regulon consists of a variety of proteins and small RNAs (*e.g.*, [219, 220]). Amongst the proteins whose genes are part of the σ^{24} regulon is the periplasmic protease and chaperone DegP, the chaperones FkpA, Skp, and SurA, and the disulfide bond isomerase DsbC, but also other components that play a role in multiple cellular pathways and are not exclusively linked to protein folding and quality control in the cell envelope [220–223]. Also, *rpoH* is part of the σ^{24} regulon, as is the *rpoE* operon itself [222, 224–226].

6.2.2 The CpxAR response

The CpxA/CpxR two-component-system is another important instance that guards the integrity of the cell envelope and responds to the presence of unfolded (misfolded) proteins [222, 227]. CpxA is a sensor histidine kinase in the cytoplasmic membrane, which is inhibited by the periplasmic protein CpxP. When misfolded proteins accumulate in the periplasm, CpxP binds to them and can no longer inhibit CpxA. Subsequently, CpxA is autophosphorylated at a conserved histidine residue in its cytoplasmic domain. This phosphate is transferred to the response regulator CpxR, which then acts as a transcriptional activator [228, 229]. CpxAR enhance the transcription of numerous genes, including those encoding DegP, the periplasmic disulfide isomerase DsbA, and σ^{32}. For an overview on CpxAR regulated genes see [230]. CpxAR acts also on its own promoter as well as on the promoter of CpxP [228, 231]. CpxAR can even function as negative regulator of transcription, as seen, *e.g.*, for *rpoE* [230].
7 Energy demands of protein biogenesis

The biogenesis of proteins is not only complex, but it is also energetically demanding. Even if only the most basic requirements – the steps involved in synthesizing a polypeptide chain – are considered, protein biogenesis requires charging tRNA with the correct amino-acid, initiating translation, elongating the polypeptide chain, and, finally, translation termination\textsuperscript{[18]}. It has been estimated that the biosynthesis of a single polypeptide chain requires the energy from the hydrolysis of four energy-rich phosphate bonds (ATP or GTP) per amino-acid\textsuperscript{[232]}. This does not take into consideration the energy that might be required for folding and refolding, potential targeting and translocation events, or degradation. Intriguingly, a recent study suggests that re-folding a protein may be much more favourable than degrading and re-synthesizing it. For DnaK-mediated refolding, five molecules of ATP (that drive unfolding) were sufficient to allow an artificially misfolded substrate (firefly luciferase) to adopt a functional conformation\textsuperscript{[233]}. Nevertheless, the production of a protein requires a lot of energy. This is an aspect that is particularly important in the light of the next chapter - the production of proteins in \textit{E. coli}. 
Protein production

The main aim of my Ph.D. studies was to understand what hampers the production of MPs and secretory proteins in *E. coli*, and to use this knowledge to improve the production of these two classes of proteins. But why do proteins need to be produced and why do we use *E. coli* for that purpose? Proteins are essential for all organisms. To understand how these important molecules work, their structure and function needs to be studied, which most often necessitates their isolation. However, the natural abundance of most proteins is too low to obtain sufficient quantities for structural and functional studies directly from the host organism. Also, isolation from natural sources is often not possible. Therefore proteins need to be produced. *E. coli* is a convenient organism for the production of proteins – it is easy to handle, fast-growing and relatively cheap to cultivate. In addition, it is genetically accessible and a wide range of genetic tools are available. For the vast majority of my studies I used the *E. coli* strain BL21(DE3) or derivatives thereof, which will be described in the next section.

8 The protein production strain BL21(DE3)

BL21(DE3) and its derivatives are particularly suited for the purpose of protein production as they lack two major proteases, the outer membrane protease OmpT and the cytoplasmic protease Lon. Both proteases have been shown to interfere with protein production and isolation (e.g., [234][235]).

BL21(DE3) was developed by Studier and colleagues for the production of soluble proteins [236]. ‘DE3’ denotes a prophage which has been integrated into the genome of BL21. The prophage sequence contains the gene of the T7 RNAP governed by the *lacUV5* promoter. The gene encoding the protein of interest is provided on a plasmid under the control of a T7-based promoter. The addition of inducer (lactose or its non-hydrolysable analogue isopropyl-β-D-thiogalactoside (IPTG)) initiates transcription of the T7 RNAP-gene. Subsequently, the T7 RNAP transcribes the target gene (see figure 6).

The *lacUV5* promoter [237] is derived from the well-known *lac* promoter and, consequently, shares the same basic features (see figure 6). *lac* promoters
consist of the -35/-10 region, three operator sites, and a binding site for the cAMP-receptor-protein CRP. As mentioned previously, transcription of genes that are under the control of a lac promoter is induced by lactose or its analogues, like IPTG. In the absence of inducer, the lac inhibitor (LacI) binds to up to two operator sites which results in a loop-like DNA-structure and prevents RNAP-binding \[238; 239\]. The addition of inducer releases LacI from the operator sites, which allows RNAP to transcribe any gene placed downstream of the lac promoter/operator region \[238\]. The lacUV5 promoter exhibits four single point mutations as compared to the wild-type promoter. These are located in the promoter (-10) and operator (O1) region as well as in the CRP-binding site. The mutations are thought to increase efficiency of DNAP binding and to make the promoter less sensitive to glucose catabolite repression \[237; 240\]. Thus, the lacUV5 promoter represents a stronger variant of the lacWT promoter.

The combination of the strong lacUV5 promoter and the T7 RNAP, which transcribes eight times faster as compared to E. coli RNAP \[241\] and targets only the promoter in front of the target gene, makes BL21(DE3) a powerful tool for the production of proteins. However, this system often suffers from leakiness (i.e., background transcription of the target gene prior to induction), due to imperfect repression of the lacUV5 promoter. Since even small amounts of T7 RNAP can be troublesome when the expression of a given gene is toxic to E. coli, additional levels of control are often provided. For the experiments that are presented in paper I-III, the target gene was placed under the control of a T7lac-hybrid promoter containing the lacO1 operator that can be bound by LacI. To further improve repression, an additional copy of lacI was provided on the expression vector. Note that the pOGCP-plasmid used by \[242\] and in paper IV employs a T7 promoter and does not harbour lacI. Background levels of T7 RNAP can also be inhibited by its natural inhibitor, T7 lysozyme, supplied from a separate plasmid, e.g., pLysS and pLysE \[243\]. In paper I, we describe the development of Lemo21(DE3). Lemo21(DE3) is a BL21(DE3) derivative that carries the gene encoding T7 lysozyme under control of a rhamnose-inducible promoter. In contrast to pLysS and pLysE, the amounts of T7 lysozyme in Lemo21(DE3) can be precisely set by the addition of rhamnose. In the next section, I will therefore briefly discuss the rhamnose promoter, rhaBAD.

9 The rhaBAD promoter

The use of the rhaBAD promoter system for protein production was introduced by Giacalone et al. \[244\]. In the presence of L-rhamnose, the transcriptional regulator RhaR activates transcription of RhaS and RhaR \[245; 246\]. RhaS
Figure 6: A Comparison of the lacWT- and the lacUV5 promoter. The binding sites for CRP-cAMP and the RNAP (-35/-10) and the transcription start site (+1) are indicated. * denotes sequence differences between the two promoters. B and C Schematic representation of the E. coli lac and rha operons in their induced state. B The addition of lactose results in dissociation of LacI from the operator (O) sites, allowing RNAP to transcribe the genes encoded by the lac operon. C Addition of rhamnose induced expression of rhaS (and rhaR). RhaS activates the transcription of genes encoded by the rhaBAD operon. Binding sites for regulatory elements are depicted as dotted lines. Red lines indicate the -35/-10 regions. Both the lac and the rhaBAD promoter are subject to CRP-cAMP-mediated catabolite repression D Schematic representation of the basic elements involved in protein production in BL21(DE3). For a detailed description see sections 8 and 9.
binds to the rhaBAD and rhaT promoters, controlling expression of genes involved in rhamnose metabolism and uptake, respectively (see figure 6). Excess amounts of RhaS down-regulate the expression of rhaS and rhaR [246]. Thus, RhaS can act both as transcriptional activator and repressor. Just as the lac operon, the rha operon is susceptible to glucose catabolite repression. Efficient regulation of the rhaBAD promoter system appears to rely on an intricate interplay of CRP-cAMP, RhaS, and RhaR [245, 246]. One major advantage of this system is that the addition of increasing amounts of inducer leads to a gradual increase in the level of transcription, whereas lac promoters are hardly titratable but appear to operate in an on/off mechanism [244].

10 The production of membrane proteins

The hydrophobic nature of MPs makes them rather difficult targets for production. MPs can be produced in inclusion bodies in the E. coli cytoplasm, from which they can be extracted and, subsequently, refolded. Even though this option seems straightforward, reports about MPs that have been recovered successfully from inclusion bodies are the exception rather than the rule [247]. Alternatively, MPs can be produced such that they integrate into the membrane (in a lipid-soluble form), from which they can be isolated. There are numerous examples of the successful extraction of MPs using detergents. Therefore, production of MPs in the membrane appears to be the method of choice. However, the production of MPs in the membrane may be toxic to the host (see section 10.1), which reduces protein production yields. In addition, it has been suggested that the available space in the E. coli membrane can be yield-limiting (e.g., [247]).

There can also be restrictions linked to the target gene or the protein itself, for example instability of the mRNA encoding the MP, or instability of the MP itself in the production host or during the isolation process. Heterologous MPs may lack a targeting/insertion signal that can be recognized by the E. coli MP biogenesis machinery, and the lipid composition of the E. coli membrane may differ significantly from the one of the organism whose MP should be produced, which can affect levels of a target MP. These examples illustrate that the production of MPs in E. coli may suffer from various drawbacks. In the next sections, I will give a brief overview over the topics most relevant for my studies; the consequences of MP production and measures taken to improve MP yields. For a comprehensive review on improving the production of proteins in general (including MPs), see, e.g., [248].
10.1 Consequences of membrane protein production

The production of MP can have consequences that are negative for the production host that may be attributed to intrinsic properties of the produced protein itself (e.g., [249]) or to side-effects of the production process. Several studies have addressed the latter point in great detail. Wagner et al. observed that the high-level production of homologous MPs has a severe negative impact on *E. coli* [149]. Inclusion bodies, containing amongst others the target MP and the precursor form of several secretory proteins, formed in the cytoplasm and sequestered cytoplasmic chaperones like DnaKJ and GroEL/ES as well as proteases. Accumulation levels of various MPs and MP-complexes were greatly reduced, in particular those of proteins of the respiratory chain. The latter resulted in an *E. coli* metabolism rewired toward energy conversion via the less efficient acetate-Pta-pathway, which is a particular concern in the light of the previously discussed energy requirements for the synthesis of proteins (section 7). In addition, the production of MPs induced the heat shock response. Based on these results it was proposed that the high-level production of MPs may saturate the capacity of the protein biogenesis machinery at the level of the Sec-translocon [149]. A later study noted that the production of heterologous proteins, as exemplified by the human KDEL-receptor, results in surprisingly similar effects, even though the production levels of the target protein in this study were markedly lower [250].

The production of the Cannabinoid receptor CB1, a G-protein coupled receptor (GPCR), or the *E. coli* AAA+ protease FtsH led to an increased activity from some – but not all – promoters of genes belonging to the $\sigma^{32}$ and $\sigma^{24}$ regulon [251]. Thus, the production of MPs resulted in an increased expression of stress-regulated genes, which is in agreement with results obtained previously [149, 250]. Interestingly, the mechanisms underlying this induction appeared to differ depending on the target MP [251].

Gubellini et al. investigated the consequences of the production of soluble proteins and MPs [252]. The authors concluded that there appears to be no general obstacle for the production of MPs, but rather that the success or failure of such an undertaking is determined by the target protein itself, which contradicts the main conclusion from previous studies [149, 251]. Notably, there is also some overlap between the different studies in the sense that MP production induces certain stress responses that can be linked to either a cytoplasmic heat shock response or a compromised cell envelope. Also, both Wagner et al. and Gubellini et al. observed stress responses upon the production of soluble proteins, albeit these were less pronounced [149, 252].

Unfortunately, it is impossible to directly compare the studies described above due to fundamental differences in the experimental setup (e.g., strain
background, expression systems, and production conditions). In either case these studies demonstrate that MP production can be detrimental to the host.

11 Improving membrane protein production

Different approaches have been employed to improve MP production yields, including the modification of the culture or induction conditions, varying strain backgrounds, plasmids, and promoter systems, isolating or engineering strains, and engineering the gene encoding the MP of interest. In the next sections I will present a selection of the most recent advances. The use of cell-free expression systems will be discussed in section 14.

11.1 Modifying culture conditions

The most straightforward way to improve the production of MPs in *E. coli* is to modify the culture conditions, *e.g.*, temperature, composition of growth medium, or the feeding method (batch versus fed-batch).

Temperature appears to be one of the prime points of optimization. Often, cultures are shifted to temperatures well below 37°C upon, or prior to, addition of inducer to increase protein yields (*e.g.*, [253–256]). Lowering the temperature during the production process may also improve the amount of membrane integrated material rather than increasing the total amount of protein produced [257]. This observation is consistent with the idea that decreasing the cultivation temperature reduces the burden of MP production on the cell by decelerating transcription and/or translation. Increased levels of certain chaperones at low temperatures due to the *E. coli* cold shock response may also be beneficial [256; 258].

Further, the composition of the growth medium can have a profound influence on the production process by, *e.g.*, supporting high density growth and/or modulating the expression kinetics/levels of the gene encoding the MP (*e.g.*, [256; 259]). Interestingly, Weiss and Grisshammer noted that addition of glucose to a culture producing the human Adenosine A<sub>2α</sub> receptor increased not only cell density but, at the same time, the number of functional receptors per cell [260]. Even though an explanation for this phenomenon was not provided it is tempting to speculate that the presence of glucose in the culture medium may have reduced the initial rate of transcription from the *lac* promoter based expression vector employed in this study.
11.2 Using mutant strains

It is probably safe to say that amongst the most successful approaches to improve the production of difficult targets in *E. coli* is the isolation of C41(DE3) and C43(DE3) from BL21(DE3) [242], given the widespread use and superiority of these strains (e.g., [242, 261]). These strains are commonly referred to as the Walker strains.

C41(DE3) is a direct derivative of BL21(DE3). This strain was isolated based on its ability to tolerate the production of the bovine oxoglutarate malate carrier protein (OGCP) (figure 8). Briefly, BL21(DE3) was transformed with a plasmid harbouring *ogcp* under control of a T7 promoter. The expression of *ogcp* following induction with IPTG is toxic to BL21(DE3) ([242] and paper IV). Expression of *ogcp* was induced with IPTG in liquid culture and cells surviving this induction were selected for on an agar plate containing IPTG. A colony that still produced OGCP was cured from the plasmid and, upon retransformation, again analyzed for its ability to produce OGCP. This procedure resulted in the isolation of C41(DE3). C43(DE3) was isolated from C41(DE3) for its ability to produce the *E. coli* *F*$_o$F$_1$ATPase subunit b.

C41(DE3) and C43(DE3) are able to produce a variety of difficult to produce proteins (both membrane and soluble) to high levels [242, 261]. Given their improved ability to produce otherwise toxic proteins due to unknown mutations, the Walker strains were ideally suited to investigate which factors can improve or hamper protein production. In agreement with previous suggestions [242], we were able to show that the main mutations responsible for the improved performance of the Walker strains weaken the *lac*UV5 promoter that controls expression of the T7 RNAP. This notion led to the development of Lemo21(DE3) (see section 8 and paper I).

The screen performed by Miroux and Walker was designed to improve yields of toxic proteins and the occurrence of mutations was not actively facilitated. In contrast, a later study sought to isolate *E. coli* strains that are generally improved in their ability to produce MPs, independent of if the production was toxic or merely inefficient. The accumulation of mutations was facilitated by chemical mutagenesis or by using a strain harbouring a proofreading-deficient version of DNAP III. The screen yielded several promising mutants, even though the nature of the mutation(s) remains unknown. However, it was speculated that, for one of the mutants, a decrease in plasmid copy number and the resulting decreased load on the protein biogenesis machinery of the host might be partially responsible for the increase in MP production levels [262]. This is in agreement with the observations presented in paper I.
11.3 Co-production of biogenesis factors

The co-production of factors that have been linked to the biogenesis of MPs appears to be another promising venue to improve MP levels in *E. coli*.

The levels of membrane-integrated CorA, an *E. coli* magnesium transporter, increased markedly when DnaK and DnaJ were co-produced [257]. Intriguingly, CorA could be produced at high levels even in the absence of additional chaperones, but the vast majority accumulated in inclusion bodies instead of being integrated into the membrane. Co-production of the chaperone system GroEL/ES only modestly increased the amount of membrane-integrated CorA, whereas co-production of SRP, SecA, and SecB had a negative effect on CorA-levels. The authors argue that, given the large N-terminal extension of CorA (235 amino-acids) prior to the first TM it is likely that this transporter is targeted to the membrane in a post- rather than a co-translational fashion. Thus, increased levels of the DnaKJ chaperone system might be required to keep CorA in a translocation-competent state [257].

A later study analyzed how the production of human GPCRs, amongst them the aforementioned CB1, is affected upon co-production of Ffh, the 4.5S RNA of SRP, FtsY, SecYE, the chaperones DnaKJ, SecB, GroEL/ES, and TF and FtsH [255]. Co-production of FtsH resulted in the most pronounced increase in levels of two GPCRs. However, the increase in levels did not correspond to an increase in active receptor, since only a modest increase in activity was detected for one target. The authors suggest that the positive effect on the target protein levels requires the protease activity of FtsH and that FtsH co-production may alter the lipid composition of the cell [255]. In a later study it was proposed that the co-production of FtsH ‘primes’ the cell for the production of the GPCRs, thereby improving their levels [251].

Finally, co-production of the established MP integrase and foldase YidC increased the amount of membrane-integrated target protein in a protein-dependent manner [263].

It is unfortunately not possible to predict if MP production levels can be improved by the co-production of biogenesis factors and which factors are the most promising ones. In addition, it should be kept in mind that, just as it has been shown for the production of soluble proteins [264], the co-production of biogenesis factors can also have negative consequences or be without effect as indicated above. Considering the finely regulated MP biogenesis machinery, care should be taken when selectively increasing the amounts of one of the components involved as to not upset this balanced system.
11.4 Screening for factors improving membrane protein production

In addition to the co-production of factors that have an established role in the biogenesis of MPs, screens have been conducted to identify genes that, upon co-expression, improve production levels of MPs in *E. coli*. *ybaB* (encoding a putative DNA-binding protein) was identified as one such gene. Its co-expression led to a pronounced increase in the production levels of several MPs [265]. For one of the targets, insertion into the membrane was demonstrated. *ybaB* is part of the $\sigma^{24}$ regulon, which responds to the presence of mis-folded proteins [217] (see section 6.2.1). Thus, it is possible that YbaB may be involved in MP quality control [265]. Unfortunately, it is not clear if the production of the MPs in this study caused any kind of folding stress.

Also co-expression of *nagD* (encoding a ribonucleotide phosphatase), a truncated version of *nlpD* (encoding a predicted lipoprotein), and the *ptsN-yhbJ-npr* gene cluster (encoding proteins of the nitrogen phosphotransferase system) increased production levels of properly folded GPCRs [254]. It was shown that the production of the target GPCRs alone did – at most – give rise to a very mild stress response, which is in agreement with the idea that there is no general obstacle to the production of MPs [252]. In contrast, co-production of NagD or NlpD (but not PtsN,YhbJ and Npr) induced expression of some, but not all, genes that are part of the $\sigma^{32}$ or envelope stress regulon ($\sigma^{24}$/CpxAR). The genes identified in this study, amongst them *ftsH* and *degP*, overlap partially with the ones identified by Xu and Link [251]. In agreement with the suggestions from this study, it was speculated that one of the ways by which the co-production of NagD and NlpD improves the yields of properly folded GPCRs is to prime the cells for the production of MPs [254].

11.5 Screening for factors that hamper membrane protein production

In addition to factors that improve the production of MPs, it is conceivable that there are also factors that interfere with the successful production of a given MP. Several screens have been conducted to identify genes encoding such factors. For example, Skretas and Georgiou found that a disruption of the gene encoding the DnaK co-chaperone DnaJ markedly increased production levels of the GPCR CB1 [266]. The authors propose that, either, strong DnaJ-binding to the nascent chain of CB1 interferes with SRP-binding and thus correct targeting, leading to reduced levels or, alternatively, that CB1 is recognized as a non-native target and degraded via DnaKJ. Disruption of two other genes, *nhar* (encoding a transcription factor) and *dinG* (encoding a DNA-damage inducible helicase), was also beneficial, even though the improvement was less pronounced. Notably, the increase in CB1 levels coincided with an increase in biomass. This indicates that disruption of the aforementioned genes at least
partially alleviated the toxic effects of the production process [266].

Studies by Nannenga and Baneyx show that the deletion of the gene encoding TF improved levels of functional, membrane-integrated protein for one *E. coli* MP and two archeal bacteriorhodopsins [263]. The effects were attributed to a more efficient binding of the targeting factor SRP to the polypeptide chain in the absence of TF. Interestingly, SRP-co-production did not affect levels of the *E. coli* MP or – in the case of the rhodopsins – even decreased MP levels in the membrane. The explanation favoured by the authors is that a disproportional increase in targeting factors shifts the production bottleneck from targeting to membrane insertion due to a limited Sec-translocon capacity [263].

It has also been suggested that high levels of acetate may hamper the production of MPs in *E. coli*, since MP protein production could be improved by using *E. coli*-K derived mutant strains that were impaired in their ability to take up glucose, and showed a reduced production of acetate [267]. Interestingly, BL21(DE3), that was included in this study and is known to accumulate lower levels of acetate [268], could compete with the mutant strains in terms of MP levels.

11.6 Modifying the gene encoding the target protein

As indicated above, limitations for the production of a given MP may be linked to the protein itself or the sequence of the gene encoding it. Modifying the gene encoding the target MP may help to overcome these limitations. The next sections will provide a brief overview of those modifications that have been used to improve MP production levels.

11.6.1 Fusion partners and tags

It should be noted that the presence of a fusion partner or a small affinity tag may serve purposes other than enhancing the levels of a given target protein. Small affinity tags or fusion partners are employed to facilitate detection and purification of a MP or to probe its topology and, thus, correct insertion into the membrane (see below). Fusion partners have also been used in screens aimed at the isolation of strains that are improved in their ability to produce MPs to rapidly identify promising mutants [262] and for the identification of factors hampering or improving MP production [254; 255; 265; 266].

There are several aspects that one ought to consider before deciding to equip a MP with a tag or fusion partner, irrespective of the purpose. Fusion partners are frequently attached to the N- or the C-terminus of a MP. Since the information initiating targeting and insertion is located in the N-terminus it must be ascertained that modifying it does not interfere with these steps. The activity of a fusion partner might be specific to a certain compartment and thus
allow detection of protein levels only when localized correctly. It has also been shown that certain tags can interfere with the ability of a protein to obtain its native fold or hamper activity \(\text{(e.g., [269])}\). Below, I will give a brief overview of the most frequently employed fusion partners, with a specific focus on those that have been used for the purpose of improving protein production.

**N-terminal fusions**

Maltose binding protein (MBP), fused to the N-terminus of a target MP, has been successfully used to improve MP production levels. This approach should be particularly promising when trying to produce heterologous MPs that do not possess targeting signals recognized by the \(E. coli\) targeting factors. Full-length MBP has first been employed by Grisshammer and coworkers, who could improve production levels of the rat neurotensin receptor (NTR) 140-fold \([253]\). In addition to a role in targeting, MBP may also enhance the solubility of a protein, as demonstrated for soluble proteins \([270]\). Notably, MBP utilizes the post-translational targeting pathway.

In contrast, the so-called P8CBD-tag utilizes the co-translational SRP-pathway \([271]\). This tag combines the M13 phage major coat protein with one TM of \(E. coli\) Lep to ensure efficient membrane targeting and insertion of the fusion protein and was successfully used for the production of functional YidC from \(Thermatoga maritimum\) in \(E. coli\).

Mistic, a small, highly hydrophobic protein from \(B. subtilis\), facilitated membrane integration of heterologous MPs in \(E. coli\) when fused to their N-terminus \([272]\). The ability of Mistic to associate tightly with the \(E. coli\) cytoplasmic membrane was shown to be crucial for its role in supporting the production of membrane-integrated proteins, a notion that was reinforced by a later study \([273]\). Nonetheless, it should be noted that the presence of Mistic may also interfere with the activity of its fusion partner, thus necessitating the removal of the tag prior to activity measurements \([269]\).

Recently, bacteriorhodopsin has been used as an N-terminal fusion partner to improve yields of two \(E. coli\) MPs \([274]\). The spectral properties of bacteriorhodopsin allow to closely following the production process. In addition, bacteriorhodopsin may facilitate targeting of the fusion partner to the membrane.

Finally, \(\beta\)-lactamase (Bla) has been utilized to probe the correct topology of MPs in screens aiming to identify components that enhance the levels of GPCRs \([254, 265]\). In these screens, Bla was equipped with a DsbA-signal sequence. Bla confers resistance to ampicillin (amp) when translocated across the bacterial membrane. It should be noted that Bla itself has not been shown to improve MP levels so far. However, based on the results obtained with MBP,
it is conceivable that also Bla itself can aid the production of MPs.

C-terminal fusions

GFP fused to the C-terminus of MPs has, e.g., been used to determine the topology of *E. coli* MPs (pioneered by [275]), to assess levels and proper membrane integration of fusion proteins when attempting to improve their production (e.g., [266]), and to monitor the quantity and quality of proteins during production and isolation (e.g., [276]). Due to its fluorescent properties the GFP-moiety can easily be monitored in whole cells. Since properly folded, fluorescent GFP does not denature when exposed to low amounts of detergent (e.g., SDS) at physiological temperatures, the integrity of MP-GFP-fusions can be visualized in an SDS-gel. Geertsma *et al.* described the use of GFP as an indicator for membrane insertion using SDS-PAGE/Immunoblotting [277], an application that has been crucial for the results presented in paper II. One of the major drawbacks of GFP is that most variants, with the notable exception of super-folder (SF)GFP [278], do not fold properly in the *E. coli* periplasm. Therefore, the use of GFP has originally been restricted to MPs whose C-terminus is located in the cytoplasm. This limitation can be overcome by introducing an additional TM at the C-terminus of MPs with a C\textsubscript{out} topology [279].

Also, thioredoxin has been utilized as a C-terminal fusion partner to enhanced yields of MBP-NTR-fusions. It was suggested that this enhancement most likely is due to a stabilizing effect of thioredoxin on the fusion protein, since thioredoxin displays a very compact fold and cannot easily be degraded [280].

11.6.2 Modifying the coding sequence

Modifying the sequence of the gene encoding a target protein is an established tool to improve the production of (heterologous) MPs. The coding sequence can be modified such that it leaves the amino-acid sequence unaltered (synonymous base exchange) or that it leads to exchange of amino-acids (non-synonymous base exchange). Also deletions of parts of the coding sequence are possible. Here, I will briefly describe the most basic approaches, as this specific aspect has not been a major part of my studies.

One way to improve final protein yields is to create protein versions that exhibit enhanced thermostability. 'Enhanced thermostability’ describes an increased resistance of a protein towards thermal denaturation, which can be crucial for their successful purification especially if proteins are unstable at ambient temperatures. This demonstrates that modifying the coding sequence not necessarily leads to improved MP production levels in *E. coli*, but may
also improve yields due to enhanced stability upon extraction from the production host. In 2000, Zhou and Bowie demonstrated that it is possible to create mutant versions of a MP with improved thermostability using *E. coli* diacylglycerol kinase as a target [281]. Several publications report the use of alanine- or alanine-leucine scanning to improve the thermostability of MPs (e.g., [282, 283]). In this method, the amino-acids of a target protein are substituted one by one by alanine (or leucine), and the different versions are probed for enhanced thermostability. However, it is not described how the substitutions affected the levels of the improved variants upon production in *E. coli*.

The gene encoding rat NTR was mutated using error-prone PCR, in order to create variants of the MP that were produced at higher levels in *E. coli* and displayed increased stability in the detergent-solubilized state [284]. It was hypothesized that the major cause for the improved production levels in *E. coli* were improvements in co-translational folding and membrane insertion. Schlinkmann *et al.* conducted an exhaustive screen where they tested the combinatorial effects of various mutations on the production of NTR1 variants [285]. This resulted in protein variants that could be produced at high levels and were stable in various detergents. Production levels could be further improved by expressing the genes encoding these NTR1-variants from a high-copy number plasmid. This suggests that these variants decreased the demands of the high-level production of MPs on the cells MP biogenesis machinery, supposedly due to superior insertion- and folding characteristics [285]. Interestingly, this indicates that high-level production of MPs must not essentially saturate the MP biogenesis machinery of *E. coli* if the MP of interest exhibits favourable features, and may help to explain why the expression of some, but not all, genes encoding a MP must be regulated in order to achieve high yields and circumvent toxicity towards *E. coli* (see, e.g., paper I). It should be noted that the NTR1 receptors and variants thereof used by [284] and [285] were fused to MBP at the N-terminus and to thioredoxin at the C-terminus.

In addition to changing the coding sequence *per se*, MP yields can be improved by completely removing unfavourable regions, for example those that reduce protein stability (e.g., [260]). It is also possible that heterologous proteins possess extensive N-terminal extensions that interfere with targeting in the production host, as shown for MPs produced in *Lactococcus lactis* [286].

Despite the documented success, there are drawbacks when modifying the coding sequence of a gene to improve yields of a protein. The introduced modifications may affect the function of the target proteins, which bears the risk of analyzing a protein variant whose properties differ from the ones of the wild-type. It is not always known how a certain mutation contributes to protein production or stability. Especially when multiple mutations are present that are not evaluated individually one cannot be sure that all mutations are beneficial.
Finally, it should be kept in mind that manipulating the sequence of a gene can have unpredictable consequences, ranging from mRNA stability to protein folding issues (for a recent review see [287]).

11.6.3 Codon-optimization

One question that has not been addressed in the previous section is whether one should use a gene that has been codon-optimized for expression in E. coli or the un-optimized cDNA sequence when attempting to produce a (heterologous) MP. Several studies described during the previous sections report the use of codon-optimized genes. But how necessary/useful is codon optimization? It has been shown that already the selective alteration of the codon directly following the start codon in a gene can improve levels of homologous proteins [288]. This indicates that, even for heterologous targets, optimizing the entire coding sequence may not be necessary (or even counterproductive). Independent of the alterations introduced, it should be kept in mind that, as noted in the previous section, changing the sequence of a gene may interfere with downstream processes. For example, changes that affect the translation speed can interfere with protein folding and thus may negatively influence the stability and function of a protein [287]. Particularly in the light of studies describing the saturation of the E. coli MP biogenesis machinery at too high expression levels of the gene encoding a MP, the general usefulness of increasing translation efficiency by codon-optimization might be questioned. Thus, codon optimization requires careful consideration and does not necessarily improve MP levels but may just as well have the opposite effect.

12 The production of secretory proteins

Secretory proteins are proteins that are equipped with a signal sequence directing them to the cytoplasmatic membrane and beyond. For the sake of clarity, from here on the term ‘secretory proteins’ will be exclusively used for soluble proteins that depend on the Sec-translocon for their translocation across the cytoplasmic membrane. For other secretion pathways see, e.g., [289; 290].

There are several major reasons to produce soluble proteins in the periplasm rather than in the cytoplasm. First, the oxidizing periplasm promotes the formation of disulfide bonds (see section 5.4.1) which many proteins, like antibodies or fragments thereof, require in order to become functional. In contrast, the cytoplasm of E. coli is highly reducing and disulfide bond formation is disfavoured. Second, the periplasmic fraction can be released selectively without releasing the contents of the cytoplasm. This should facilitate protein purification given the lower amount of ‘contaminating’ proteins in the released frac-
tion. Also, it is thought that the proteolytic activity in the periplasm is lower as compared to the one in the cytoplasm (e.g., [289] [290], but see section 12.4).

Secretory proteins do not necessarily remain in the periplasm but can be released into the culture medium, from which they can be recovered (e.g., [289] [291] [292]). Since *E. coli* does not naturally secrete large amounts of proteins into the medium, the use of mutants that show an increased permeability of the outer membrane has proven beneficial [291]. It has also been demonstrated that the capability of *E. coli* to secrete proteins into the medium can be influenced by the cultivation conditions [292] (see section 12.5).

There is a wealth of information available on the production of secretory proteins. In the next sections I will give a succinct overview of factors that were most relevant to my studies.

### 12.1 Choosing a targeting signal

As indicated earlier, secretory proteins can be targeted to the Sec-translocon via the co-translational, SRP-dependent pathway or the post-translational pathway (see section 5.1). The signal sequence is the prime determinant for which pathway is selected. Thus, one of the first considerations when deciding to produce a (heterologous) protein in the *E. coli* periplasm is to select a signal sequence.

Intuitively, a signal sequence targeting proteins to the membrane in a post-translational manner might be a logical choice, given that the majority of secretory proteins in *E. coli* are thought to utilize this pathway (e.g., [86]). However, proteins that tend to fold rapidly and stably in the cytoplasm may not be compatible with post-translational export via the Sec-translocon [86] [87] [96]. Such proteins ought to be equipped with a signal sequence funneling them into the SRP-dependent, co-translational pathway. It should be noted that signal sequences promote export to varying extends even if they are destined for the same pathway; a feature that may be determined by the protein that is to be exported (e.g., [96]).

In some instances, signal peptides from heterologous organisms have been shown to promote efficient export of secretory proteins in *E. coli* (e.g., [293]). Instead of a signal sequence, also an entire secretory protein fused to the N-terminus of a target protein has been used to obtain high production levels of soluble proteins in the periplasm [294].

Unfortunately, only few studies are available that present a systematic comparison of different signal sequences for the production of secretory proteins [86] [96]. In either case, it does not seem to be possible to pinpoint the ’optimal’ signal sequence for a given target.
12.2 Transcription and translation levels

Several studies show that the levels of transcription and translation can influence secretory protein production. Kadokura et al. investigated the effects of high-level PhoA production on the protein secretion capacity of *E. coli*. High-level production led to reduced accumulation levels of the mature forms of PhoA, OmpA, and β-lactamase. *In vitro* translocation assays using inverted membrane vesicles indicated that this was due to a defect in secretion capacity at the level of the membrane. It should be noted that, for unknown reasons, an increase in OmpA-levels was observed upon high-level production of PhoA [295].

Simmons and Yansuro designed a series of vectors with translation initiation regions of varying strength and investigated the production of heterologous and homologous secretory proteins from these vectors in *E. coli*. They demonstrated that ’the secretion of heterologous proteins in *E. coli* (into the periplasm) can be significantly enhanced by manipulating the level of translation’ [296]. The amount of regulation necessary was protein dependent. Interestingly, in contrast to data presented by Kadokura et al. [295], native *E. coli* proteins were efficiently secreted even at high translation levels [296]. The results of these studies are in agreement with previous observations, namely, that defects in growth and secretion in a secA mutant strain could be recovered by reducing the rate of protein synthesis [297], and illustrate nicely that the rate of protein synthesis should be balanced with the protein biogenesis/secretion machinery of *E. coli* in order to avoid adverse effects.

The transcription of a gene encoding a (secretory) protein can be modulated by, e.g., varying the copy number of the expression vector or the promoter controlling target-gene expression (e.g., [298; 299]). In paper I of this thesis, we present how transcription can be tuned by regulating the activity of the polymerase transcribing the target gene. This approach was successfully employed for the production of secretory proteins (see paper III). Finally, also the choice of the culture medium has been suggested as a way to balance protein synthesis with the secretion capacity of *E. coli* (e.g., [293; 298]).

12.3 Co-producing folding catalysts and targeting components

Just as observed for MPs, secretory protein production may benefit from the co-production of chaperones and folding catalysts (see section 5.4). Multiple studies have demonstrated the potential of these proteins to enhance the production of secretory proteins. The examples chosen below present only a limited selection of these studies. Not surprisingly, disulfide-bonded proteins have been shown to benefit from the co-production of components of the Dsb-system (see section 5.4.1). For example, co-production of DsbA, B, C, and D
markedly improved the amount of soluble, mature horseradish peroxidase in
the periplasm [300]. The production of DsbC along with human tissue-type
plasminogen activator (Pa) resulted in a pronounced increase in the yields of
active protein [301]. Human PA is an extraordinarily difficult target to produce
in the active form as it possesses 17 disulfide bonds. Thus, improvement upon
co-production of a disulfide bond isomerase appears logical. There are vari-
ous other examples where DsbC was the key contributor to improving protein
production levels (see, e.g., [302]).

Apart from components of the Dsb-system, also other periplasmic proteins
can be beneficial for the production of secretory proteins when co-produced,
like FkpA [303], Skp [304] and a combination of FkpA and SurA [305]. Based
on the notion that chaperones and folding catalysts can aid the production of
secretory proteins, a series of vectors (designated pTum), encoding different
combinations of these proteins, was developed that may aid the production of
secretory proteins [305].

Interestingly, it was reported that even the co-secretion of the cytoplasmic
co-chaperone DnaJ, equipped with an OmpA signal sequence, markedly im-
proved levels of a PA-variant, pro-insulin, and a scFv upon secretion into the
periplasm [306].

Finally, co-producing the cytosolic targeting component SRP gave rise to
higher levels of leech carboxypeptidase inhibitor in the medium when the pro-
tein was equipped with the SRP-dependent DsbA-signal sequence [307]. The
same study reports that the absence of the cytosolic chaperone TF was benefi-
cial for the production. The authors suggest that the absence of TF facilitates
SRP binding to the nascent chain and, consequently, targeting of the secretory
protein to the membrane. This contrasts results obtained for the production of
membrane proteins, where SRP co-production had a negative effect [263].

12.4 Protease-deficient strains

In some cases periplasmic proteases negatively affected production levels by
degrading the target protein (e.g., [294; 308]). This bottleneck can be over-
come by the use of protease-deficient strains [308; 309]. However, it should
be kept in mind that protease deficiency in combination with the production of
a secretory protein may lead to toxic effects, as was shown for DegP [310].

12.5 Cultivation conditions and additives

The composition of the culture medium has been shown to influence the pro-
duction levels of secretory proteins. Mergulhão et al. noted that secretion into
the periplasm was less efficient in LB-medium as compared to M9-minimal
medium [293; 298]. As indicated above, the improved secretion in minimal
medium has been suggested to originate from a controlled cell-growth rate, at which the expression level of the gene is in balance with the cells secretion capacity. Interestingly, Ukkonen et al. demonstrated that the choice of medium and the cultivation conditions (here the oxygen transfer) not only influences yields of secretory proteins in general but also the release of the protein from the periplasm into the medium (e.g., [292]). In this study, highest yields could be obtained by a restricted feed of glucose in an otherwise rich medium. The high yields were attributed to an overall increase in biomass formation, rather than to an increase in target protein levels per cell.

Finally, the addition of low molecular mass additives, like L-arginine and sorbitol, has proven beneficial for the production of secretory proteins (e.g., [306; 311]). For disulfide-bonded proteins, the addition of reduced glutathione (or another redox-active component) may increase protein production levels (e.g., [302; 306]).

13 Disulfide bond formation in the cytoplasm

As an alternative to the production of soluble proteins in the periplasm, strains have been engineered that allow and support the formation of disulfide bonds in the cytoplasm [312; 314]. The two most popular ones are 'Origami' and 'SHuffle'. Both strains lack glutathione reductase and thioredoxin reductase, which results in an oxidizing cytoplasm. In addition, SHuffle possesses a cytoplasmically localized variant of the disulfide bond isomerase DsbC to support resolution of wrongly formed disulfide bonds. Even though these strains are valuable tools for the production of disulfide bond containing proteins especially in cases where secretion across the membrane cannot be achieved, or production in the periplasm is hampered by other factors, they still may suffer from some of the drawbacks of protein production in the cytoplasm, e.g., cumbersome purification of the target protein from whole-cell extracts.

14 Protein production using cell-free systems

Given the impressive progress that has been made with cell-free expression systems, especially with respect to MP production, it seems appropriate to include a brief overview of these systems. Proteins can be produced in cell-free systems based on lysates from different sources, including E. coli, wheat-germs, rabbit reticulocytes, and yeast (e.g., [315–317]). Trials have even been set up with systems based on HeLa-cells, a cell-line of human origin (e.g., [318]). As alternative to cell lysates, systems have been established that are based upon purified components [317; 319]. Cell-free systems come in two flavours; ei-
ther as translation-only systems where the mRNA is provided, or as coupled transcription/translation systems, that include the transcription of the gene of interest.

The production of proteins in cell-free systems offers multiple advantages compared to their production in whole cells \[316\] \[317\]. First and foremost, yield-limiting toxicity of the production process is circumvented. Cell-free systems are open systems that allow addition of components at virtually any time during a proceeding reaction, including the addition of non-natural amino-acids or low molecular mass additives that may stabilize the protein during the production process \[320\]. Systems based on purified components are reduced in complexity with respect to the number of different proteins are present, which should facilitate the purification of the target protein from the reaction mixture. If purified components are used, proteolytic degradation is circumvented. For the production of MPs in a soluble form, cell-free systems can be supplemented with detergents, liposomes, nanodiscs, or other lipid-based or lipid-mimicking structures that prevent aggregate formation \[316\] \[321\]. Even proteins that require disulfide bonds have been successfully produced in cell-free systems \[322\].

It should be kept in mind that the use of cell-free systems does not completely circumvent problems experienced with in-cell protein production. For example, cell-free systems that are used for the heterologous production of proteins will still suffer from drawbacks with respect to species-specific modifications like glycosylation \[316\]. Nonetheless, cell-free systems ought to be considered a serious alternative to cell-based protein production.

15 Concluding remarks on protein production

The above mentioned studies and many others demonstrate that the amount of protein produced in *E. coli* (independent of the protein’s final destination) can be influenced by, *e.g.*, the strain background, cultivation conditions, induction regimes, the co-production of chaperones and folding catalysts, and fusion partners. This makes it difficult, if not impossible, to design *a priori* a strategy that guarantees the highest yields of a given target protein. Consequently, the optimal production conditions must still be largely determined by exhaustive protein production screens. Alternatively, developing an understanding of what hampers the production of different classes of proteins can provide a basis for the rational design of the production process, including the engineering of strains (*e.g.*, paper I).
Making the connection between protein production and genome evolution

Assuming that it is not obvious how such seemingly different topics are connected it appears appropriate to provide a short explanation. As indicated during the previous chapters, the starting point for my thesis was the characterization of the protein production strain C41(DE3), with the aim of identifying mutations that make this strain so suitable for the production of many proteins. During the course of this analysis it became apparent that C41(DE3) had not only acquired mutations accounting for its improved protein production characteristics, but, in addition, mutations that occurred in response to the cultivation conditions used during its isolation. Given that the isolation conditions are well documented [242] we reasoned that C41(DE3) provides us with the unique opportunity to gain insights into the evolution of a widely used strain in response to various stresses under (controlled) laboratory conditions.
17 Mutations and evolution

One of the most famous sentences attributed to the biologist Monod is “Any-thing found to be true of **E. coli** must also be true of elephants”. This statement may be used to explain why **E. coli** still is a favourite model organism to investigate various phenomena, amongst them the mechanisms underlying evolution (e.g., [323]). It is widely accepted that ‘mutation is the most important driving force behind evolution’ [323]. ’Mutation’ is every sequence change occurring in the basic genetic material of an organism (i.e., the chromosome and extra-chromosomal elements). Mutations can arise for example due to environmental insults that directly alter the structure of the DNA (like exposure to UV-light or chemicals) followed by incorrect repair, by mistakes during genome replication, recombination and the activity of mobile DNA-elements (like insertion (IS) elements or transposons). Mutations occur to varying extents, ranging from the exchange of single bases to the deletion, insertion, or rearrangement of large fragments of DNA. Whether a certain mutation is beneficial, neutral, or detrimental is determined by the character and site of the mutation and by the prevailing environmental conditions (e.g., [324]).

**E. coli** exhibits a rather low intrinsic mutation rate of approximately $10^{-9}$ to $10^{-11}$ per base per replication, most likely to minimize the risk of potentially detrimental outcomes [325]. This can be attributed to the existence of DNA-repair mechanisms and the high fidelity of **E. coli** DNAP III; the prime polymerase for DNA replication [324][325]. There are four more DNAPs in **E. coli**; DNAP I, II, IV, and V. The activity of the different DNAPs during DNA repair or basic replication has been linked to the occurrence of mutations. In the next section, I will therefore briefly discuss the different **E. coli** DNAPs and their potential impact on mutation rates.
18 DNA polymerases and their effect on the mutation rate

As mentioned in the previous section, DNAP III is the prime polymerase for DNA replication in E. coli, and with 17 subunits by far the most complex one [325]. DNAP III possesses an intrinsic 3’-5’ exonuclease activity that permits immediate correction of an incorrectly inserted base, accounting for its high accuracy. DNAP I possesses both 3’-5’ and 5’-3’ exonuclease activity. Amongst other functions it removes the RNA-primers that are required for replication of the lagging strand, and closes the resulting gaps. It can also participate in DNA-repair [325]. DNAP II resembles DNAP III in the sense that it also possesses 3’-5’ exonuclease activity. It has been suggested that DNAP II can substitute for DNAP III during replication, thus acting as ‘back-up’ for this essential polymerase. In addition, DNAP II also appears to have an important role during repair of different DNA-damages [325].

Neither DNAP IV nor DNAP V has any reported exonuclease activity and, consequently, both are prone to introduce mutations when replicating a DNA-sequence. Nonetheless, these polymerases have an important role in replication when the integrity of the DNA is compromised, since they can synthesize through DNA-damages that normally cause polymerase stalling [324; 325]. Experimental evidence suggests that DNAP IV activity preferably results in frame-shift mutations, whereas DNAP V activity gives rise to base exchanges. DNAP II, IV, and V are part of the SOS-regulon. The expression of genes included in this regulon is induced upon extensive DNA-damage, as occurs for example during exposure to UV-light [324].

19 Increasing the mutation rate in response to stress

There is increasing evidence that under conditions that are particularly challenging or even harmful to E. coli, mechanisms are induced that increase the basic mutation rate. This phenomenon has been described as 'stress-induced' mutagenesis [323]. For example, Cairns and Foster observed that a strain with impaired ability to utilize lactose due to a frame-shift in a lacI-lacZ fusion (Lac−) gave rise to mutant cells that were able to efficiently grow on medium with lactose as sole carbon and energy source (Lac+) upon prolonged incubation [326]. A large fraction of the cells exhibited reverting frame-shift mutations, which depended primarily on the induction of the error-prone DNAP IV, and, probably, the presence of double-strand breaks [323; 326; 328].

Another phenomenon, termed ‘growth advantage during stationary phase’ (GASP), describes an increased mutation rate in stationary phase cells. Yeiser et al. [329] demonstrated that the number of viable cells in cultures that were starved for a prolonged period of time decreased markedly in the absence of
DNAP II, IV, and V. Real time (RT) PCR performed after five days of cultivation showed increased expression levels of the genes encoding DNAP II and IV. Expression of recA, which is induced during SOS-response, was not altered [329]. These results are intriguing as they suggest that replication by error-prone polymerases is important under conditions that cells are likely to experience in nature, like here, the deprivation of nutrients. In this respect it should be noted that even under laboratory conditions cells may be exposed to periods of starvation and, consequently, experience GASP. Part of the selection procedure of C41(DE3) involved a prolonged (seven-day) incubation in liquid culture without external supply of nutrients (see paper IV). Thus, it is tempting to speculate that parts of the mutations that occurred in the genome of this strain can be linked to GASP.

20 IS-element mediated mutagenesis

Insertion sequences or IS-elements are mobile, genetic elements that have been identified in most organisms. IS-elements are small as they most commonly carry only the information necessary for their own transposition [330]. IS-elements have been found in all *E. coli* genomes analyzed thus far, but their types and distribution vary. For example, in BL21(DE3) there are 62 IS-elements, whereas the genome of MG1655 only harbours 54. In addition, there is little overlap in the numbers of a certain type of IS-elements and in their distribution between these strains [2].

IS-elements can give rise to mutations simply by moving within the genome, since they are able to re-locate by ‘jumping’ out of one position and into another or to move via replication intermediates [331]. It should be noted that simple excision not linked to a transposition event has also been observed (e.g., [332]). IS-elements can even cause the deletion of entire regions of a genome. This was for example observed in cells of the *E. coli* B strain REL606, which partially or entirely lost the genes required for ribose uptake and catabolism (i.e., the rbs operon) when grown for a long time in medium with glucose as sole carbon and energy source [333].

21 Concluding remarks on evolution and mutagenesis

The above sections cannot cover the number of experiments that have been performed to study mutagenesis and evolution using *E. coli*. Nonetheless, they demonstrate that there is a striking variability with respect to the underlying mechanisms. Notably, many of these mechanisms appear to be conserved between different organisms. Given the generality of many of the underlying
processes it seems that, in agreement with Monod's statement, even studying a 'simple' bacterium like *E. coli* can improve our understanding of such complex processes like evolution.
Summaries of papers I-IV

Paper I – Tuning *Escherichia coli* for membrane protein production.

Background

The Walker strains C41(DE3) and C43(DE3) were isolated from BL21(DE3) and are able to produce many difficult proteins, including membrane proteins, at high levels. As in BL21(DE3), the expression of a given target gene is driven by T7 RNAP, whose gene is located on the chromosome. The target gene is introduced on an expression vector under control of a T7-based promoter. Expression of the gene encoding T7 RNAP is governed by the *lac*UV5 promoter, which is a stronger variant of the *lac*WT promoter (see figure 6). Addition of IPTG (or lactose) induces expression of the T7 RNAP-gene, and, subsequently, T7 RNAP transcribes the target gene. To understand why the Walker strains perform better as compared to BL21(DE3) for the production of many proteins, in particular MPs, we expressed a gene encoding a difficult target, the membrane protein YidC C-terminally fused to GFP, in BL21(DE3) pLysS, C41(DE3) and C43(DE3); and compared YidC-GFP production levels and the cellular responses to the production. Note that BL21(DE3) pLysS was chosen instead of BL21(DE3) to reduce the toxic effects that arise from background expression of *yidC-gfp* prior to induction.

Results

In BL21(DE3) pLysS, YidC-GFP production levels were low. The production of YidC-GFP in BL21(DE3) pLysS i) severely hampered growth, ii) led to inclusion body formation, and iii), increased levels of proteins whose genes are part of the heat shock regulon (compare [149]). In the Walker strains, these effects still occurred but to a lesser extent, and YidC-GFP levels were markedly higher. Interestingly, the onset of YidC-GFP production was much faster in BL21(DE3) pLysS as compared to the Walker strains. This difference could be attributed to mutations in the above described *lac*UV5 promoter: Three out
of the four bases that differentiate lacUV5 from lacWT were reverted to wild-type, resulting in a weakened promoter and, subsequently, lower T7 RNAP levels. But why should lower T7 RNAP levels give rise to higher levels of YidC-GFP? Previous studies demonstrated that the high-level production of membrane proteins can saturate the capacity of the Sec-translocon, which impedes biomass formation \[149\]. A likely effect of lowering yidC-gfp expression levels is therefore a reduced load on the membrane protein biogenesis and insertion machinery (in particular the Sec-translocon). This reduces the toxic effects of membrane protein production, resulting in more biomass formation and, consequently, higher final protein production levels.

To strengthen this conclusion, a strain was constructed where the activity, rather than the levels, of the T7 RNAP can be controlled via its natural inhibitor, T7 lysozyme (see figure 7). To this end, the gene encoding T7 lysozyme was placed under control of the well-titratable rhaBAD promoter on a plasmid designated pLemo. The plasmid was introduced into BL21(DE3) and the resulting strain was termed Lemo21(DE3), the idea being that the more rhamnose is added, the more T7 lysozyme is produced that can inhibit T7 RNAP, which should gradually reduce expression levels of yidC-gfp. Indeed, increasing amounts of rhamnose increased biomass formation and reduced toxic effects in Lemo21(DE3) cells producing YidC-GFP. In addition, various other proteins could be produced in Lemo21(DE3) at levels approaching, or even exceeding, those in the Walker strains. For proteins whose production was toxic to BL21(DE3) but not to the Walker strains, higher amounts of rhamnose were required to improve production in Lemo21(DE3). Notably,
some proteins could be produced well in BL21(DE3) or in Lemo21(DE3) in the absence or presence of low amounts of rhamnose, indicating that saturation of the Sec-translocon does not always occur upon MP-production.

Conclusions

The characterization of the Walker strains C41(DE3) and C43(DE3) resulted in the engineering of a BL21(DE3)-based protein production strain, Lemo21(DE3) which is a valuable tool for the production of (membrane) proteins.

Paper II - Optimizing membrane protein overexpression in the *Escherichia coli* strain Lemo21(DE3).

Background

In Paper I we have described the development of a BL21(DE3) based strain, called Lemo21(DE3). In Lemo21(DE3) expression levels of a given target gene can be precisely controlled by adding varying amount of rhamnose to the culture medium. Here, we have studied how the production yields of a given membrane protein are optimized in Lemo21(DE3), and investigated the effects of cultivation conditions (e.g., temperature) to probe the versatility of Lemo21(DE3) as a tool for membrane protein production.

Results

Increasing amounts of rhamnose resulted in increasing amounts of T7 lysozyme which, in line with the results from paper I, alleviated the toxic effects of the production of YidC-GFP and GltP-GFP in Lemo21(DE3). This is demonstrated by, e.g., improved cell growth and decreasing levels of the inclusion body binding proteins IbpA/B. Accordingly, levels of the MP-GFP fusions increased.

At sub-optimal rhamnose concentrations, the majority of the MP-GFP fusions existed in a non-fluorescent, aggregated form, which is in agreement with the increased levels of IbpA/B under these conditions. Increasing concentrations of rhamnose did not simply increase the production level of the two MP-GFP-fusions but rather shifted the ratio of not properly membrane-inserted MPs: properly inserted MPs towards the latter. Under conditions where the production of YidC-GFP and GltP-GFP was toxic to Lemo21(DE3), there was a high pressure to evade MP production, as demonstrated by the appearance of non-producing sub-populations of cells in the absence of rhamnose. In contrast, almost all cells in a culture produced YidC-GFP when the toxicity was
alleviated.

Our results show that YidC-GFP production levels reached at 37°C were markedly lower as compared to those reached at 30°C. Lowering cultivation temperature further (to 20°C) did not increase production levels by much under the conditions tested (batch cultivation). Notably, the concentration of rhamnose needed to obtain optimal MP levels decreased with decreasing temperatures. This agrees with the idea that a decrease in temperature reduces the burden of MP-production on the *E. coli* MP-biogenesis machinery. Finally, optimizing the expression levels of *gltP-gfp* also increased the production levels of functional GltP-GFP; and the MPs NhaA and MhpI produced in Lemo21(DE3) were suitable for crystallization.

Conclusions

Our results indicate that, in Lemo21(DE3), tuning the expression levels of a gene encoding a MP of interest optimizes MP production levels by modulating the ratio of not properly membrane-inserted MP to properly inserted MP. This is in agreement with a scenario where high-level expression of genes encoding a membrane protein can lead to the saturation of the Sec-translocon, resulting in accumulation of aggregates in the cytoplasm. Further, Lemo21(DE3) can be used to optimize the production of MPs for functional and structural studies.

Paper III – Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels.

Background

Here, we explored how suitable Lemo21(DE3) is for the production of secretory proteins. To this end, we used two proteins, SFGFP and the scFv BL1, equipped with a DsbA-derived signal sequence. DsbA is targeted to the periplasm via the co-translational, SRP-dependent pathway. SFGFP is a fast-folding version of GFP that, in contrast to most GFP-variants, can fold into its fluorescent form in the periplasm, which allows to quickly monitor protein levels. The scFv BL1 was chosen for its ability to bind to β-galactosidase, which provides a straightforward activity assay.

Results

In BL21(DE3) or in Lemo21(DE3) at sub-optimal rhamnose concentrations, the production of secretory SFGFP and the scFv BL1 resulted in reduced cell
growth and increased levels of IbpA/B. The toxicity accompanying the production process could be reduced by adding increasing concentrations of rhamnose. For secretory SFGFP, already low levels of rhamnose were sufficient to significantly reduce toxicity and increase levels of SFGFP in the periplasm. For the secretory scFv BL1 we show that, i) increasing concentrations of rhamnose resulted in an increase of the mature form of the protein while the amount of precursor decreases, ii) no precursor could be detected at optimal rhamnose concentrations, iii) the mature form was localized in the periplasm, and, iv) at least a portion of the protein produced at the optimal rhamnose concentration was able to bind to β-galactosidase and, consequently, active. Reduced BL1 did not bind to β-galactosidase. Thus, under the conditions used in this study the activity of this scFv required the presence of disulfide bonds. Size-exclusion chromatography indicated homogeneity of the scFv BL1 upon purification.

Finally, we noted that the production of the two secretory proteins at high gene expression levels (i.e., at sub-optimal rhamnose concentrations) not only led to increasing levels of IbpA/B but also resulted in the accumulation of the precursor form of OmpA and reduced levels of the mature forms of the periplasmic proteins MBP and DegP, all of which are targeted to the Sec-translocon in a post-translational manner. This suggests that the cell is impaired in its ability to translocate proteins across the cytoplasmic membrane, which supports the model of insufficient Sec-translocon capacity.

Conclusions

Lemo21(DE3) is a valuable tool for the production of secretory proteins. In this strain, expression levels of the genes encoding secretory proteins can be tuned such that saturation of the Sec-translocon capacity is minimized and production levels in the periplasm optimized. This is in keeping with previous studies showing that the production levels of periplasmic proteins can be improved by varying the level of translation (see section 12).

Paper IV – Reconstructing the evolution of BL21 into C41 reveals rapid adaptability of a widely used E. coli strain.

Background

Our initial proteomics characterization C41(DE3) (see paper I) suggested that the mutations in the lacUV5 promoter cannot represent the only difference to BL21(DE3), but that other mutations must have occurred during the isolation
To select for mutants of the well-known protein production strain BL21(DE3) that are able to produce ‘toxic’ proteins at high levels, the strain was transformed with a plasmid encoding the bovine oxoglutarate-malate-carrier protein (pOGCP). pOGCP carried an amp resistance marker. Expression of $ogcp$ is toxic to BL21(DE3) ([242] and paper IV). Expression of $ogcp$ was induced in liquid culture by the addition of IPTG and survivors of this induction were selected for on IPTG-containing agar-plates. From this plate, a colony producing OGCP was selected and cured from pOGCP by a seven-day incubation in rich medium in the absence of amp. The presence of IPTG should facilitate plasmid loss. After seven days, a colony occurred that had lost pOGCP and, upon re-transformation with the plasmid, retained the ability to produce OGCP at high levels.

of this strain. When the complete genome sequence of BL21(DE3) became available in 2009 [2], it became feasible to analyze the genome of C41(DE3).

Results

Mutations in four different regions in the genome of C41(DE3) relative to BL21(DE3) could be attributed to the isolation of the strain. Changes in the $lacUV5$ promoter region lead, as described above, to reduced levels of T7 RNAP upon induction with IPTG as compared to BL21(DE3). A non-synonymous point mutation in the gene encoding the sensor-histidine kinase YehU appears to dramatically increase its activity. YehU regulates expression of $yjiY$, a gene encoding a putative peptide transporter. The mutation in $yehU$ results in strongly elevated YjiY-levels.

BL21(DE3) is unable to grow on ribose only due to an IS-element in $rbsD$, the first gene of the ribose operon. In C41(DE3), this IS-element is absent, resulting in an intact $rbsD$ open reading frame (orf). This enables C41(DE3) to grow on ribose as sole carbon and energy source. Finally, a small deletion in the gene encoding the dicarboxylate sensing histidine kinase DcuS allows C41(DE3) to grow on $C_4$-dicarboxylates, like succinate, malate and aspartate, as sole carbon and energy source. Similar to $rbsD$, also $dcuS$ is a pseudogene in BL21(DE3), but its orf has been restored in C41(DE3) due to a small deletion.

The mutations weakening the $lacUV5$ promoter in C41(DE3) enabled this
strain to survive the production of OGCP. We reasoned that the other mutations may enable C41(DE3) to meet the elevated energy-, carbon-, and nitrogen-demands of OGCP-synthesis. Surprisingly, the mutations affecting yehU, rbsD, and dcuS did not alter OGCP production levels. Mimicking the selection process (compare figure 8) revealed instead that at least the mutations affecting rbsD and dcuS could be linked to the second step in selection process, i.e., the prolonged incubation of BL21(DE3) in a closed system with no external supply of nutrients, rather than to the production of OGCP.

The mutations in the lacUV5 promoter were intriguing for two reasons: First, we found that already prior to induction of ogcp expression (see figure 8), approximately 0.9% of the cells that contained pOGCP (i.e., were resistant to amp) were able to form colonies in the presence of IPTG, a number that increased to 77% 3 h after induction. This indicates rapid adaptation to protein production stress. Second, out of the cells that were resistant to IPTG and amp, almost all contained the same, weakening mutations in the lacUV5 promoter, i.e., had reverted to C41(DE3) in this position. Further experiments demonstrated that the mutations in the lacUV5 promoter already could be selected for as early as 5 h after transforming BL21(DE3) with pOGCP. The occurrence of the mutations depended on the presence of the lacWT promoter in BL21(DE3).

Conclusions

In this study we show that the mutations in the genome of C41(DE3) occurred in a sequential manner during the selection of the strain. The lacUV5 promoter reversion seems to be a prerequisite for BL21(DE3) to survive background levels and the production of OGCP and can occur within a remarkably short period of time. The mutations affecting rbsD, dcuS, and yehU were likely selected for during the subsequent curing process, and allow C41(DE3) to utilize an extended array of carbon-, energy-, and possibly nitrogen sources as compared to BL21(DE3). Importantly, the dependency of the mutations in the lacUV5 promoter on the lacWT promoter suggests a recombination-dependent mechanism, that allows rapid adaptation to protein production stress.
Outlook

The aim of this thesis was to further our understanding of what can hamper the production of MPs and secretory proteins in *E. coli* and to find ways to improve protein production levels, based on an in-depth analysis of two mutant strains that are able to produce many 'difficult' proteins. This analysis resulted in the development of Lemo21(DE3), a strain that is very well suited for the production of MPs and secretory proteins.

If the saturation of the Sec-translocon capacity is limiting to the production of MPs and secretory proteins, Lemo21(DE3) is a highly valuable tool since it allows to alleviate this bottleneck. However, based on the results presented in papers I to III, and extensive research done by others, we realize that there are other bottlenecks. Thus, Lemo21(DE3) may be seen as a starting point for the development of strains that are even better suited for the production of MPs and secretory proteins. It would be interesting to see if (functional) levels of these proteins can be further improved by the controlled co-production of targeting components, chaperones, and folding catalysts as described in sections 11 and 12. In addition, protease-deficient strains may be considered. However, particular care should be taken with such approaches as to not unbalance protein homeostasis in *E. coli*. In this respect it should be noted that attempts to improve levels of the secretory scFv BL1 (paper III) by using a TF deletion mutant as described before [307] were not successful (unpublished results).

As it has been suggested that both the available space in the membrane and in the periplasm may limit the levels of proteins that can be produced in these compartments, measures may be considered that increase this space. Membrane space could be improved by, e.g., inducing the formation of additional membranes or the formation of cytoplasmic membrane vesicles (e.g., [334, 335]). The periplasmic space may be 'increased' indirectly by supporting further secretion of the protein of interest into the culture medium. Careful design of culture medium and the cultivation process, mutant strains with increased outer membrane permeability, or the co-production of proteins inducing or facilitating release into the medium may be employed to that end (e.g., [289, 291, 292, 336]).

Heterologous MPs, especially eukaryotic ones, appear to be particularly challenging with respect to production, as exemplified by the relatively low accumulation levels of TspA, TspB, and NTR in paper I even at the optimal concentration of rhamnose in Lemo21(DE3) (see also section 11). The reasons for this are not clear yet and require further research. Features like mRNA structure and stability may be considered, as well as the requirement of specific folding assistants, differences in membrane composition, and enhanced susceptibility to proteases in the heterologous production host etc. Unfortu-
nately, the modulation of several of these factors require the specific engineering of the target rather than allowing for a generic approach. An often pursued option to improve production of heterologous (membrane) proteins is to codon-optimize the gene of interest. However, as outlined before, this approach may suffer from various drawbacks (see section 11.6.3). In our hands, codon-optimization has thus far not been successful, except for a minor improvement in protein production levels for one target (paper II and results not shown). As an alternative to codon-optimization, the use of strains supplying additional copies of rare codon tRNAs, as Rosetta(DE3), might be considered.

Finally, despite the benefits of *E. coli* for the production of proteins, for some targets it may be necessary to select a different production host. Preferably, hosts should be considered that are more closely related to the organism whose protein ought to be produced than *E. coli*, in particular if post-translational modifications, like glycosylation, are required. Notably, MPs have been successfully produced in non-*E. coli*-based systems (e.g., [337–340]), and eukaryotic systems are commonly used for the production of antibodies (secretory proteins) (e.g., [341]).

A very successful venue to improve the production of proteins in *E. coli* is the isolation of mutant strains, e.g., C41(DE3) [242]. In paper IV we have demonstrated that the genotype of C41(DE3) reflects the stresses and conditions prevailing during its isolation. The rate at which the mutations in the lacUV5 promoter occurred, and at which they came to dominate the culture, has consequences for the use of BL21(DE3) for the purpose of protein production and for physiological studies, in particular when proteins are produced that are ‘toxic’ to *E. coli*. The levels of target proteins may be both negatively or positively affected by mutations in the lacUV5 promoter, since a reduced gene expression rate has been shown to be beneficial for the production of certain proteins. However, if a study aims at investigating the physiological responses of BL21(DE3) to the production of a given protein, the presence of a mixed population with respect to the lacUV5 promoter certainly complicates the interpretation of the obtained results. Importantly, our observations suggest that the presence of multiple copies of a gene within the chromosome may allow *E. coli* and other bacteria to adapt to certain conditions more rapidly than commonly anticipated. It will be interesting to investigate the precise mechanisms underlying the identified mutations.

Taken together, I was able to improve protein production levels of MPs and secretory proteins in *E. coli* and my work furthers our understanding of how *E. coli* evolves in response to stress.
Zusammenfassung


Um herauszufinden, wie die Herstellung von Membranproteinen verbessert werden kann, habe ich während meiner Doktorarbeit einen Escherichia coli - Mutantenstamm untersucht, der aufgrund von unbekannten Veränderungen im Erbmaterial
al besonders gut für die Herstellung von Membranproteinen geeignet ist. Die aus-
schlaggebenden Mutationen verringern die Syntheserate des herzustellenden (Mem-
bran)proteins, was die Überlastung der Membran-Integrationsstellen verhindert. Ba-
sierend auf dieser Beobachtung wurde ein Escherichia coli - Stamm konstruiert, in
dem die Syntheserate der herzustellenden Proteine je nach Bedarf reguliert werden
kann. Dieser Stamm eignet sich sowohl für die Herstellung von Membranproteinen,
as auch von wasserlöslichen Proteinen. Schlussendlich zeigte die Untersuchung des
Mutantenstammes, das Escherichia coli über Mechanismen verfügt, die eine effekti-
ve Veränderung des genetischen Materials innerhalb weniger Stunden erlauben. Diese
gestatten Escherichia coli und (vermutlich) anderen Bakterien, sich innerhalb kürzes-
ter Zeit an Situationen anzupassen, die das Überleben der Zelle gefährden, wie sie
zum Beispiel durch die Gabe von Antibiotika entstehen.
Sammanfattning


Under min doktorandtid har jag främst jobbat med att förbättra framställningsprocessen för membranprotein i *Escherichia coli* genom att undersöka mutantstammar som lämpar sig bättre till proteinframställning än den ursprungliga stammen. Mina resultat visar att de mutationer som bidrar mest till förbättringen minskar syntesfrikheten av just det protein som ska framställas. Det förhindrar att membran-integrationsställen överlastas. Utifrån dessa resultat konstruerade vi en *Escherichia coli*-stam som tillåter att syntes-friheten anpassas för varje protein. Stammen lämpar sig inte bara för framställning av membranproteiner, utan även för vattenlösliga proteiner. Dessutom har vi sett att det bara tar några timmar för *Escherichia coli* att föränd-
ra arvsmassan. Det betyder att kolibakterier, och förmodligen även andra bakterier, kan anpassa sig snabbt och effektivt till situationer som hotar cellens överlevnad, en situation som till exempel uppstår när man äter antibiotika.
First and foremost I’d like to thank my supervisor Jan-Willem. I really appreciate your support and advice through all these years. Not many supervisors would have granted their Ph.D. students the amount of freedom you did. And even though there were some days where I believe we both did not want to see too much of each other, I think the result was worth the effort!

My co-supervisor Gunnar: I find it amazing that someone who has as much to do as you have still always has an open ear and door for everyone at DBB, no matter where you are or what time it is. Thank you for your support!

Stefan, thank you for all the work you do for the Ph.D. students at the department. I don’t want to imagine how many 1-year reports, half-time seminars, TIBs-exams and theses you have looked at, but it is highly appreciated!

All the present and past members of the Jan-Willem de Gier group. It is hard to find the right words for how wonderful it was to get to know you. Louise, we only overlapped by a couple of weeks, but I am still grateful for being able to stay at your place. Sam, there are no words to describe your energy! Thank you for your enthusiasm, for introducing me to the art of pouring and analyzing 2D-Gels, for your generous supply of sweets and the delicious white-chocolate ’kladdkaka’ Less is more ;). Dr. David, my joyful bench-neighbour: What would I have done without your singing and jokes, not to mention your flawless German? Thank you for always staying so calm and for all your support. Mirjam, I don’t know many people who are able to make science look so elegant and easy. Your are a great scientist, mother, and friend, and your support through all these years both in- and outside the lab is beyond price. Dimitra, thank you for a lot of laughs in the lab and for many, many delicious greek sweets! Anna and Thomas, thank you for always being so calm and for not taking me to seriously all the time :). With you in the lab it never gets boring! I believe you must have suffered quite a bit listening to all my stories about how great Sonja is... Edurne, the best student anyone could ask for! Thank you for all your help and support, and for staying long nights to finish a purification or a gel. I am so happy that you decided to stick around for a bit longer. Our
students Jonas, Carolin, João a.k.a ’J’, Lukasz, Daphne, and Uma, thanks for contributing to the nice atmosphere in the lab.

There are some people who helped me a lot during my Ph.D-studies: Pierre, thank you for welcoming me to your lab in France, for introducing me to P1 transduction, and for your advice on our manuscript. A big THANK YOU also to your group - I had a great time in Toulouse! Rob, for all your advice on the last paper. I highly appreciate it and it is truly fun to see your interest and enthusiasm. Dan S and Mike for practical and theoretical advice on how to do FSEC. Bill, Petra, and Stina for your help with microscopy. Gianlucca, for your patience when explaining Mass Spectrometry. Peter L, thank you for your help with the oxygraph. All my collaborators throughout these years that I have not mentioned yet, especially Yimmy (Mass-Spec), John (FACS), and Dirk-Jan and Joen for fruitful discussions and the entire Luirink-group for a very nice stay in Amsterdam. David D, for all ideas and work you put into paper II. Sara, you are not 'in here' yet but working with you is always inspiring.

All the wonderful and fun people from the ’neighbouring’ labs and offices, and from the four corridors that at one point belonged to DBB/CBR: I started making a list of all the people that should be in here, but I realized after six lines with names that you are simply too many :) Thank you all for making DBB a nice environment to work in, for many talks, laughs, and for the generous supply of chemicals!

Åke, you truly were an encyclopedia in biochemistry. No matter how strange a topic was you would always know where to find the source!

Ann, Maria, Lollo, Lotta, Anki, Anita, Alex, Bogos, Håkan, Peter, Bengt, Torbjörn, and Haidi: You deserve a medal for putting up with all the small and big problems at the department. Thank you for keeping the place running behind the scenes!

The ’Settlers’: Per, David, Mirjam, Christian, Joanna, Sven and Aron. Even though we play a lot less frequently lately it was so much fun! Enjoying a nice meal together, the ’peaceful’ atmosphere, Per’s vein ..... I miss this!

I also got to know many great people outside of work, who made it very easy for me to feel at home in Sweden. Hendrik, Alicia, and Immo: we share a unique experience, and I am really glad that we can look back at this now and laugh about it. To the Lappis-gången Christin, Can, Jonathan, Melinda, and Yasen: Thank you for making my first student-corridor experience so nice. I
Anja: När fyller du år?? Well, this was also a rather unique experience!

Mutti und Sven: Auch, wenn wir nicht mehr unbedingt nebeneinander wohnen, weiß ich, dass ich mich immer auf Euch verlassen kann. Sven, vielleicht sehen wir uns ja bald öfter!

References


http://bass.bio.uci.edu/hudel/bs99a/lecture25/lecture6-2.html. 25


Raran-Kurussi S and Waugh DS. The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding protein but their folding is either spontaneous or chaperone-mediated. *PLoS One*, 7(11):e49589, 2012.


Ukkonen K, Veijola J, Vasala A, and Neubauer P. 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(1):73, 2013. 41


Salema V and Fernández LA. High yield purification of nanobodies from the periplasm of *E. coli* as fusions with the maltose binding protein. *Protein Expr Purif*, 91(1):42–8, 2013. 41


