

Doctoral Thesis in Biochemistry, Stockholm University,  
Sweden

Optimizing membrane and secretory protein production in  
Gram-negative bacteria

Anna Hjelm

Information about the cover: Crystal structure of the Sec-translocon  
and electron micrographs of *Escherichia coli*.

©Anna Hjelm, Stockholm University 2015

ISBN 978-91-7649-284-0

Printed in Sweden by Holmbergs, Malmö 2015

Distributor: Department of Biochemistry and Biophysics



# Optimizing membrane and secretory protein production in Gram-negative bacteria

Anna Hjelm

# List of papers

- I. Schlegel S, Löfbom J, Lee C, **Hjelm A**, Klepsch M, Strous M, Drew D, Slotboom D-J and de Gier J-W. (2012) Optimizing membrane protein overexpression in the *Escherichia coli* strain Lemo21(DE3). *Journal of Molecular Biology* 423, no 4: 648-59.
- II. Lee C, Kang H J, **Hjelm A**, Quereshi A A, Nji E, Choudhury H, Beis K, de Gier J-W and Drew D. (2014) MemStar: A one shot *Escherichia coli*-based approach for high-level bacterial membrane protein production. *FEBS Letters* 588, no 20: 3761-9.
- III. Zhang Z, Kuipers G, Niemiec L, Baumgarten T, Slotboom D-J, de Gier J-W and **Hjelm A**. (2015) High-level production of membrane proteins in *E. coli* BL21(DE3) by omitting the inducer IPTG. *Microbial Cell Factories* 142, no 14.
- IV. **Hjelm A**, Rujas E, Zhang Z, Karyolaimos A, Vikström D and de Gier J-W. Engineering *E. coli* for rhamnose P<sub>BAD</sub> promoter-based production of membrane and secretory proteins. *Manuscript in preparation*.
- V. **Hjelm A**, Söderström B, Vikström D, Jong W, Luirink J and de Gier J-W. (2015) Autotransporter-based antigen display in bacterial ghosts. *Applied and Environmental Microbiology* 81, no 2: 726-35.

# Additional publications

- VI. **Hjelm A**, Schlegel S, Baumgarten T, Klepsch M, Vickström D, Drew D and de Gier J-W. (2013) Optimizing *E. coli*-based membrane protein production using Lemo21(DE3) and GFP-Fusions. *Methods in Molecular Biology* 1033, 381-400
- VII. Schlegel S, **Hjelm A**, Baumgarten T, Vickström D & de Gier J-W. (2014) Bacterial-based membrane protein production. *Biochimica et Biophysica Acta* 1843, no 8: 1739-49.

# Contents

List of papers.....	iv
Additional publications .....	v
Introduction .....	1
Gram-negative bacteria.....	3
1 The cytoplasm.....	3
2 The cytoplasmic membrane .....	3
3 The periplasm .....	5
4 The outer membrane.....	5
Proteins and their biogenesis in <i>E. coli</i> .....	7
5 Protein folding in the cytoplasm .....	9
5.1 Trigger Factor .....	10
5.2 DnaKJ/GrpE .....	11
5.3 GroEL/ES .....	11
6 Protein targeting to the inner membrane .....	12
6.1 The co-translational insertion/translocation pathway.....	14
6.2 The post-translational translocation pathway(s).....	14
7 Destination: inner membrane & periplasm .....	15
7.1 The Sec-translocon.....	15
7.2 Auxiliary Sec-translocon subunits .....	17
7.3 Insertion and topology of inner membrane proteins .....	19
7.4 Folding and processing of periplasmic proteins .....	20
8 Destination: outer membrane & extracellular milieu .....	22
Protein homeostasis in <i>E. coli</i> .....	24
9 Protein degradation.....	24
10 Stress responses .....	26

Recombinant protein production .....	28
11 Recombinant protein production bottlenecks.....	28
12 Protein production hosts .....	29
12.1 <i>E. coli</i> .....	30
13 Optimization strategies.....	31
13.1 Expression vector design.....	31
13.1.1 Choice of promoter system.....	31
13.1.1.1 The <i>lac</i> promoter .....	32
13.1.1.2 The <i>rhaBAD</i> promoter .....	33
13.1.1.3 Other promoters used for recombinant protein production .....	35
13.1.2 Additional elements of an expression vector.....	35
13.2 Optimization of the protein production host.....	37
13.2.1 The BL21(DE3)/pT7 protein production platform .....	37
13.2.2 Isolation of improved <i>E. coli</i> protein production hosts.....	38
13.2.2.1 Isolation of the Walker-strains .....	38
13.2.2.2 Isolation of other <i>E. coli</i> protein production hosts .....	39
13.2.3 Developing <i>E. coli</i> hosts <i>via</i> targeted engineering strategies .....	42
13.2.3.1 Tuning membrane protein production rates in Lemo21(DE3) ....	42
13.2.3.2 Generation of knockout strains.....	43
13.2.3.3 Co-production of protein biogenesis factors .....	44
13.2.3.4 Disulfide bond formation in the <i>E. coli</i> cytoplasm .....	45
13.3 Modifying the target protein and the gene encoding it.....	46
13.3.1 Protein engineering and codon optimization .....	46
13.3.2 Fusion partners and tags .....	48
13.3.2.1 The GFP pipeline .....	49
13.3.3 The use of different signal sequences for secretory proteins .....	51
13.4 Culture conditions .....	51
Autotransporter-mediated surface display .....	53
14 <i>Salmonella typhimurium</i> .....	53
15 Autotransporters .....	54
15.1 The hemoglobin protease-based display platform.....	55
16 Bacterial ghosts and lysis protein E.....	57
Summaries of papers I-V and Outlook.....	58
Populärvetenskaplig sammanfattning	
Acknowledgements	
References	



# Introduction

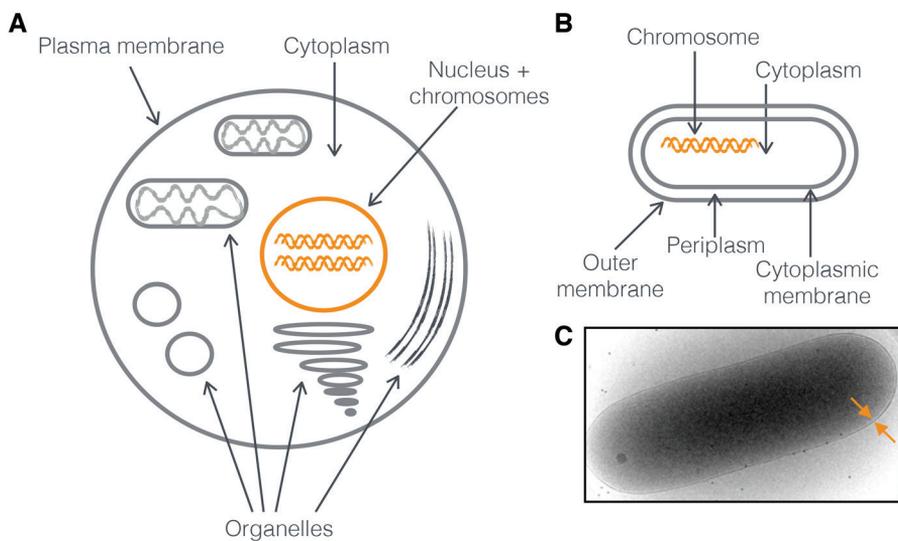
Proteins perform a variety of essential functions in the cell, and they come in many different shapes and sizes. To study the function and structure of proteins, they usually have to be produced in a recombinant fashion using a suitable host organism. The protein production host can be either prokaryotic, *e.g.* *Escherichia coli* (*E. coli*), or eukaryotic, *e.g.* yeast, insect cells and mammalian cells (Fig. 1). The choice of host depends on the origin of the protein to be produced and on the purpose of its production. The Gram-negative bacterium *E. coli*, used in my PhD studies, is not only the most common prokaryotic model organism but also the most common prokaryotic protein production host.

Roughly spoken, there are two kinds of proteins: membrane-associated proteins and soluble proteins. The soluble proteins reside in aqueous environments such as the bacterial cytoplasm, whereas membrane proteins reside in membranes, which constitute the boundaries between cells, or organelles, and their surroundings (Fig. 1). Membrane proteins are involved in important processes such as nutrient import, export of toxic compounds, stimuli responses and other types of communication with the surrounding milieu. Due to the importance of these processes, it is not surprising that membrane proteins play key roles in health and disease. It is thus of great importance to study the function and structure of this class of proteins. However, the natural abundance of most membrane proteins is low and their hydrophobic nature makes them difficult to handle. Moreover, their production is usually toxic to the production host, which leads to low production yields. Therefore, there is a need to develop improved strategies for the production of membrane proteins.

Also soluble proteins can be difficult to produce. Proteins containing disulfide bonds for example, like hormones and antibody fragments, cannot adopt their native three-dimensional fold in reducing environments such as the bacterial cytoplasm. A common approach to solve this problem is to produce these proteins in the oxidizing environment of the bacterial periplasm, which is the compartment between the two membranes surrounding Gram-negative bacteria. The production of proteins in this compartment is often hampered by similar

obstacles as the production of membrane proteins and does therefore benefit from the same optimization strategies.

In the following sections I will give introductions to the anatomy of Gram-negative bacteria and to the protein biogenesis and homeostasis machineries in *E. coli*. Subsequently, I will describe strategies to improve recombinant production of membrane and secretory proteins in this host. I will also give one example of recombinant production of membrane proteins for vaccine development, in the closely to *E. coli* related bacterium *Salmonella typhimurium*.



**Fig. 1.** **A.** Schematic representation of a eukaryotic cell (diameter typically 10-100  $\mu\text{m}$ ). **B.** Schematic representation of a Gram-negative bacterial (prokaryotic) cell (typically 2  $\mu\text{m}$  long). **C.** Electron micrograph of *S. typhimurium* (appr. 2  $\mu\text{m}$  long). The two arrows indicate the cell envelope.

# Gram-negative bacteria

There are many different strategies to produce recombinant proteins. In my PhD studies, I have used the Gram-negative bacterium *E. coli* and, for one particular application, *S. typhimurium* as protein production hosts. In the following sections I will give an overview of the four compartments that constitute Gram-negative bacteria: the cytoplasm, the cytoplasmic membrane, the periplasm and the outer membrane (1) (Fig. 1B).

## 1 The cytoplasm

The cytoplasm, the aqueous inner compartment of Gram-negative bacteria, contains soluble proteins covering a great diversity of functions. Importantly, the cytoplasm contains the chromosome, which carries the genetic information. Most Gram-negative bacteria have only one chromosome, which, in contrast to the chromosomes of eukaryotic cells, is not enclosed by a membrane barrier. The genetic information encoded on the chromosome, made up of a deoxyribonucleic acid (DNA) molecule, is used as a recipe for protein synthesis, as briefly described in the section ‘Proteins and their biogenesis in *E. coli*’. The lower complexity of the organization of bacteria as compared to eukaryotic cells is not restricted to the absence of a nucleus: All organelles within the eukaryotic cytoplasm are absent in the bacterial cytoplasm (Fig. 1).

## 2 The cytoplasmic membrane

The cytoplasmic membrane, a.k.a. the inner membrane (IM), is the innermost membrane of the Gram-negative bacterial cell envelope and has thus direct contact with the cytoplasm. Biological membranes consist of lipids and proteins.

Membrane lipids consist of a hydrophilic head-group and a hydrophobic tail (Fig. 2A). When lipids are exposed to an aqueous environment, the hydro-

phobic tails from different lipid molecules gather together such that they are shielded from the water molecules by the hydrophilic head-groups that will face outwards towards the environment. Due to this ‘hydrophobic effect’ membranes are structured as a lipid bilayer, with the hydrophobic tails facing inwards and the hydrophilic head-groups facing outwards (2) (Fig. 2A). The lipid composition of different kinds of membranes varies and can be affected by many different factors. The *E. coli* IM consists mainly of glycerophospholipids (phosphatidylethanolamine, phosphatidylglycerol and cardiolipin). Modifying the degree of saturation of the fatty acids as well as their length will lead to altered fluidity of the membrane (3–5). Note that the IM is symmetric, meaning that the inner and outer leaflets contain the same type of lipids. Ions and polar molecules cannot cross the IM by diffusion but have to be transported by proteins residing in the membrane. This controlled transport is important for *e.g.* establishing and maintaining ion gradients.

The proteins of the IM (IMPs) are either integral or peripheral membrane proteins (6) (Fig. 2A). The integral membrane proteins are either embedded in the membrane or covalently attached *via* a lipid anchor. The lipid-anchored membrane proteins are referred to as ‘lipoproteins’. All embedded integral IMPs are so-called “ $\alpha$ -helical bundle proteins” consisting of one or multiple transmembrane domains (TMs). The TMs adopt an  $\alpha$ -helical secondary fold and are, in multiple TM proteins, connected to each other by hydrophilic loop-regions protruding out from the membrane either on the periplasmic or the cytoplasmic side (7) (Fig 2B, right side). The peripheral membrane proteins adhere to one of the leaflets of the membrane *via* non-covalent bonds to the lipids and/or to integral membrane proteins (Fig. 2A).

The view of a biological membrane has been changing a lot over the years. The fluid-mosaic model was for a long time the most accepted model (8). This model suggests that lipids make up the main part of the membrane and that the lipids simply serve as a solvent for the proteins to float in. Nowadays, our view of a membrane is more complex and it is generally assumed that the lipids and proteins affect each other’s function and that the membrane structure is highly organized (9–11).

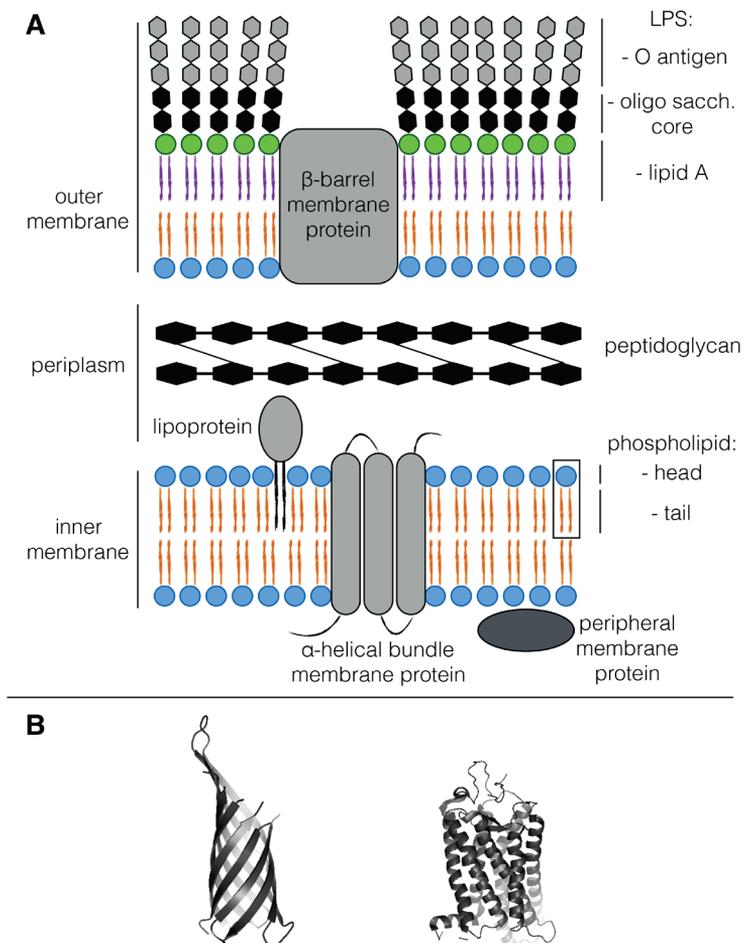
### 3 The periplasm

In between the inner and the outer membrane is the periplasm, which contains a plethora of proteins. The periplasm has been thought of as a gel-like compartment, but more recent studies suggest that the viscosity of the periplasm is very similar to that of the cytoplasm (12,13). The proteins in the periplasm are transported there from the cytoplasm across the inner membrane *via* protein-conducting channels embedded in the membrane (Fig. 2A). The periplasm contains a peptidoglycan layer, a mesh-like polymer consisting of amino acids and sugars, which gives the bacterial cell its shape and rigidity (14) (Fig. 2A). Without the peptidoglycan layer the bacterial cell would, under most conditions, lyse due to osmotic pressure. In contrast to the cytoplasm, which is reducing, the periplasm is oxidizing. This is a very important property for the folding of proteins that contain disulfide bonds. Disulfide bonds cannot be formed in the reducing environment of the cytoplasm. Notably, the periplasm does not contain the energy transfer molecule adenosine triphosphate (ATP).

### 4 The outer membrane

The outer membrane (OM) is, just like the IM, composed of a lipid bilayer containing integral and peripheral proteins. However, the two membrane systems differ significantly in some aspects (15) (Fig. 2A). The absolute majority of the proteins that reside in the OM (OMPs) consists of  $\beta$ -strands that fold into a cylinder, a so-called  $\beta$ -barrel (7) (Fig 2B, left side). The OM contains  $\beta$ -barrel proteins that function as porins, through which ions and polar compounds can diffuse, making the OM more permeable than the inner membrane (16). Another important difference is that the OM is asymmetric. The inner leaflet consists, just like the IM, mainly of glycerophospholipids, whereas the outer leaflet consists of glycolipids (lipopolysaccharides (LPS)) (1,3) (Fig. 2A). The main function of LPS is to serve as a permeability barrier towards external threats such as antibiotics. Importantly, LPS is highly immunogenic and involved in pathogenicity of Gram-negative bacteria. LPS consists of three domains: lipid A, the core oligosaccharide and the so-called O-antigen (1) (Fig. 2A). Lipid A is the hydrophobic domain anchoring the LPS molecule to the outer leaflet of the OM and is the primary component behind the endotoxic shock associated with septicemia caused by Gram-negative bacteria (17). The number and length of the

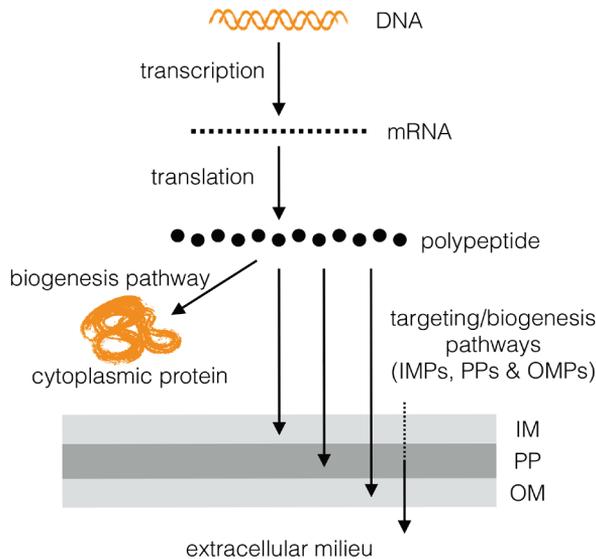
fatty acids determine the toxicity of lipid A. The core oligosaccharide is attached to lipid A, projecting outwards from the bacterial surface, and consists of a short chain of sugar residues. A repetitive glycan polymer, referred to as the ‘O antigen’, is subsequently bound to the core oligosaccharide. The O antigen is highly diverse between and within bacterial species and is required for virulence due to its role in colonization and protection against host-mediated bacterial lysis mechanisms (17). Most laboratory Gram-negative bacterial strains lack the O antigen, which renders them less virulent (17). LPS without O antigen is referred to as “rough LPS” in contrast to “smooth LPS” which indicates a complete LPS molecule.



**Fig. 2. A.** Schematic representation of the cell envelope of a Gram-negative bacterium. **B.** Examples of the two major types of integral membrane proteins. Left: The  $\beta$ -barrel outer membrane protein OmpA (PDB: 1QJP). Right: The  $\alpha$ -helical membrane protein Rhodopsin (PDB: 2I35).

# Proteins and their biogenesis in *E. coli*

In bacteria, all proteins are synthesized in the cytoplasm. According to the ‘central dogma’, first described by Francis Crick in 1970, a gene is ‘transcribed’ into a ‘messenger RNA’ (mRNA), during a process known as ‘transcription’, and the mRNA is subsequently ‘translated’ into a protein, during the ‘translation process’ (18) (Fig. 3). Proteins are made up of so-called polypeptides, which consist of amino acids. There are 22 naturally occurring amino acids that all have different structures and characteristics. Therefore, the amino-acid sequence of a polypeptide determines the characteristics of a protein. The surface of water-soluble proteins are enriched in polar and charged amino acids and hydrophobic regions are buried inside the protein, protected from the aqueous environment (19). In contrast, membrane-spanning regions of IMPs and OMPs are enriched in hydrophobic amino acids, which are exposed on the surface of the protein to



**Fig. 3.** The Central Dogma (in the context of a Gram-negative bacterium): from gene to protein. Abbreviations: IM(Ps); inner membrane (proteins), PP(s); periplasm (periplasmic proteins), OM(Ps); outer membrane (proteins).

interact with the lipid environment (20,21). Once the polypeptide chain of a protein has been synthesized, the biogenesis pathway towards a fully folded and functional protein depends on which compartment the protein is destined for (Fig. 3). At the N-terminus, proteins destined for any of the three compartments of the cell envelope are equipped with specific ‘signal/targeting sequences’ directing them to the correct location. These proteins are all confronted with the challenge of traversing membrane barriers (Fig. 3).

Transcription is carried out by the RNA polymerase (RNAP). RNAP binds at the so-called ‘promoter-region’, a sequence element located upstream of the gene, and initiates there the transcription process during which an mRNA transcript, complementary to the DNA sequence of the gene, is generated. Importantly, the recruitment of RNAP to the promoter-region is dependent on the binding to a so-called ‘ $\sigma$  factor’ (22). There are different  $\sigma$  factors and they determine to which promoters RNAPs are recruited and are therefore important regulators of gene expression (23). The most relevant *E. coli*  $\sigma$  factors for my studies will be described in section 10. More than one transcribing RNAP molecule can be present at different positions on the same gene simultaneously, generating more than one mRNA transcript at a time (24,25). The “strength” of the promoter and the activity of the RNAP will affect the number of mRNA transcripts that are produced, which subsequently will affect the kinetics of protein synthesis. The role of promoter strengths and RNAP activity for recombinant protein production will be discussed in further detail in section 13.

During translation, the nucleotide sequence of the mRNA transcript, generated during transcription, is translated into an amino-acid sequence by the ribosome, which is a highly conserved macromolecular machinery consisting of both protein and RNA subunits. The bacterial ribosome (70S) consists of the large (50S) and the small ribosomal (30S) subunits (26). The 50S subunit is a complex made up of two rRNA molecules and more than 30 proteins, and the 30S subunit consists of one rRNA molecule and approximately 20 proteins (26). The ribosome contains the ‘decoding center’, in which the mRNA is decoded, and the ‘peptidyl transfer center’ (PTC), in which the bonds between amino acids are formed, generating the growing polypeptide chain (26). The insertion of the correct amino acid at the correct position in the polypeptide chain is dictated by so-called transfer RNAs (tRNAs). tRNAs contain both an anti-codon sequence specific for a certain codon on the mRNA and carries the specific amino acid for that codon. The synthesized polypeptide chain leaves the ribosome

through the ‘exit-tunnel’ on the large ribosomal subunit (26). The efficiency of ribosome binding to the mRNA transcript largely depends on the sequence at, and around, the ‘ribosomal binding site’ (RBS) of the mRNA (27,28). The role of the RBS sequence will be discussed in more detail in sections 13.1.2 and 13.3.1. Just like multiple RNAPs can generate multiple transcripts simultaneously, multiple ribosomes can attach to the same mRNA for the synthesis of multiple polypeptide chains simultaneously, creating co-called polysomes (29). Not surprisingly, the ribosome density, which is affected by the sequence at and around the RBS, on an mRNA transcript affects protein accumulation kinetics (26).

## 5 Protein folding in the cytoplasm

The folding of a protein is a complex process in which a protein can be ‘trapped’ in non-functional/non-native conformations (30). For soluble proteins, these non-productive conformations often cause hydrophobic amino acids of the protein to be exposed on the surface of the protein instead of being buried inside the protein (31). Exposure of hydrophobic amino acids to the aqueous environment can lead to aggregate formation (32). Importantly, in the case of membrane proteins, the hydrophobic residues are (in their native conformation) exposed on the surface of the protein, which is compatible with the lipid environment of the membrane (21). Therefore, membrane proteins are even more sensitive to aggregation in the cytoplasm than soluble proteins. To prevent protein misfolding, all kingdoms of life possess chaperones and targeting factors that interact with the polypeptide as it emerges from the ribosome and assist its folding and targeting (33,34). Some chaperones accelerate folding (‘foldases’) whereas others rather prevent premature folding (‘holdases’). Chaperones can be ribosome-bound and interact with the polypeptide chain upon its emergence from the exit-tunnel, *i.e.* in a co-translational fashion. Other chaperones are not bound to the ribosome and can act on the polypeptide chain either co- or post-translationally. The lack of chaperones leads to extensive aggregate formation and misfolding of essential proteins (34), which is the reason why some chaperones are essential to cell survival. The majority of chaperones known today are proteins, but it has been shown that lipids and RNAs also can affect the folding of some proteins (35–37). Apart from being involved in *de novo* protein folding, chaperones can also assist in the maintenance of native proteins and in the refolding or degrada-

tion of misfolded proteins, as discussed in the section ‘Protein homeostasis in *E. coli*’. Below will follow a brief description of the most prominent chaperone systems of the *E. coli* cytoplasm: Trigger Factor, DnaKJ/GrpE and GroEL/ES. Notably, these three chaperone systems collaborate with each other, forming a cytoplasmic ‘folding network’ (Fig. 4, pathway I) (38). This chaperone network appears to also play a role in the biogenesis pathways of proteins residing in the cell envelope, as recently reviewed in (39). Recombinant production of membrane and secretory proteins commonly triggers cellular stress responses in which these chaperones are central players, as discussed in section 10. For the sake of clarity, in the following I will use the term ‘secretory proteins’ for all proteins destined for the periplasm and the OM. Proteins destined for the extracellular milieu will be referred to as ‘extracellular proteins’.

## 5.1 Trigger Factor

Trigger Factor (TF) is an ATP-independent chaperone that is highly abundant in the *E. coli* cytoplasm (40). It has gained most attention as a ribosome-bound chaperone that prevents misfolding of polypeptides as they emerge from the ribosome exit tunnel (41,42) (Fig. 4). Ribosome-bound TF likely interacts with most polypeptides but it has a preference for hydrophobic polypeptide stretches flanked by positively charged amino acids (43). Although most TF substrates are cytoplasmic proteins, also secretory proteins can be co-translationally bound by TF (34,44). TF plays an important role in directing secretory proteins through the post-translational translocation pathway (45,46), as described in section 6 (Fig. 4, pathway III & IV). TF bound to the ribosome nascent chain complex (RNC) may dissociate from the ribosome as the polypeptide grows longer, which allows for another TF molecule to bind (47,48). Although it is not completely clear as to how long TF stays attached to the polypeptide, it has been suggested that the release of TF is initiated as the hydrophobic TF binding motif has been successfully buried into the interior of the protein (34). In many cases, the release of TF is followed by further co- and post-translational folding assistance by the non-ribosome-associated chaperones DnaKJ/GrpE and GroEL/ES.

Interestingly, also non-ribosome bound TF appears to have a function in protein folding. The function of ‘free’ TF is not yet completely elucidated, but it has been suggested that it can assist posttranslational folding of multisubunit complexes (48).

## 5.2 DnaKJ/GrpE

DnaK has a central role in the cytoplasmic chaperone network (38). It receives *de novo* substrates from TF and passes many of them on to the downstream chaperone GroEL/ES (Fig. 4, pathway I). It should be noted that although the majority of the *de novo* DnaK substrates are cytoplasmic proteins, DnaK appears to also play a role in the targeting and biogenesis of some IM and secretory proteins (39) (Fig. 4). DnaK can interact with polypeptides both in a co- and post-translational manner. The function of DnaK is dependent on ATP and on its two co-chaperones DnaJ and the GrpE, as reviewed in *e.g.* (34). DnaJ and GrpE regulate the substrate binding cycle and ensure proper ATP hydrolysis, which drives the opening and closing of the substrate binding domain of DnaK (34,49,50). Upon release from DnaK, the cytoplasmic substrate protein folds into its native conformation or is delivered to GroEL/ES for further folding assistance (31,34). DnaK is only essential under stress-induced conditions when protein aggregation is elevated (51,52). Notably, DnaK can, under stress-induced situations, rescue mutants lacking TF, but mutants lacking both DnaK and TF cannot survive under such conditions (51,52). Importantly, DnaK also plays a key role in quality control and protein homeostasis, as described in the section ‘Protein homeostasis in *E. coli*’.

## 5.3 GroEL/ES

GroEL receives the majority of its *de novo* substrates from DnaK (Fig. 4, pathway I), as reviewed in *e.g.* (34). Obligate substrates of GroEL often have a complex native structure, which is usually not accomplished by DnaK alone (31,53). GroEL belongs to a type of chaperone systems referred to as ‘chaperonins’, which have a typical “double doughnut” shape forming two oligomeric barrel cavities in which proteins can fold in a protected environment (54,55). GroEL function is dependent on ATP hydrolysis and on its co-chaperone GroES, which functions as a lid closing the barrel after the substrate has been accommodated inside (56). The interior of the GroEL barrel is initially hydrophobic, allowing the substrate to bind. Subsequently, ATP binding causes a large conformational change of the barrel, making the interior hydrophilic (57,58). If the substrate has still not attained its native structure, it will re-bind for a second attempt of GroEL-mediated folding (31,57,59). Recently, it has been shown that GroEL does not seem to only provide a protected cavity for folding but also to actively cata-

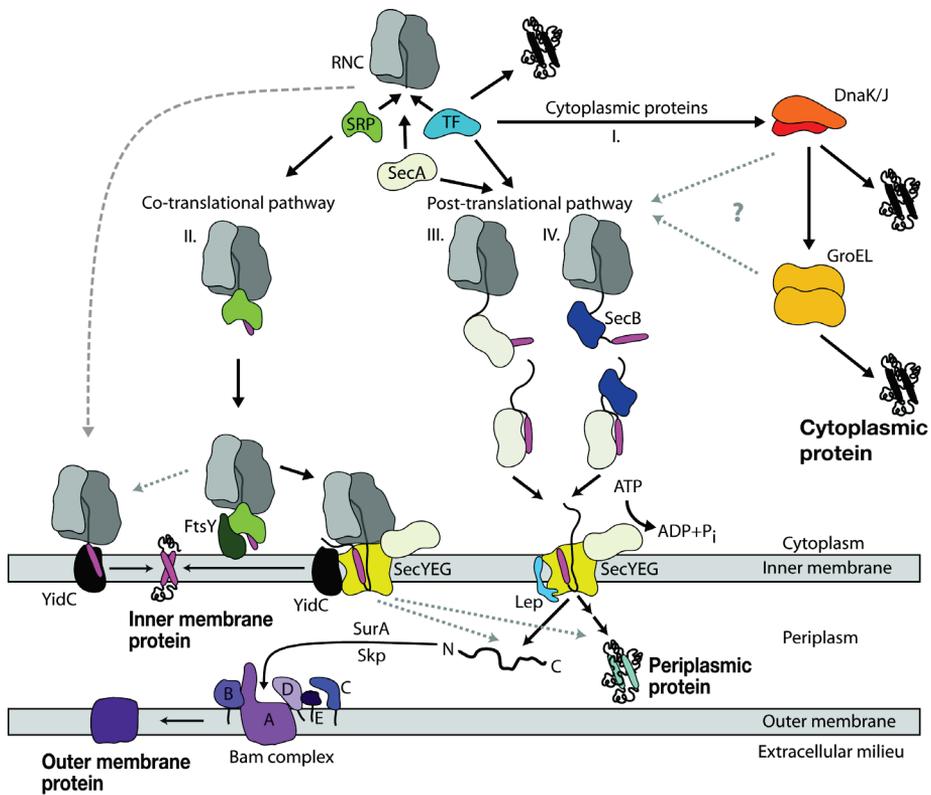
lyze the folding of the substrate (60). GroEL is an essential chaperone (61), even under non-stressed conditions, and plays just like DnaK an important role in quality control and protein homeostasis.

## 6 Protein targeting to the inner membrane

Newly synthesized polypeptides destined for any of the compartments of the cell envelope have first to be delivered to the membrane-bound machineries that execute their integration into/translocation across the IM (Fig. 4). There are two main insertion/translocation machineries in the *E. coli* IM: the Sec-translocon and the TAT-translocon, see *e.g.*, (62). During my PhD studies, I have only focused on insertion/translocation *via* the Sec-translocon. Therefore, I will here only describe the Sec-translocon. A protein can be inserted into/translocated across the IM *via* the Sec-translocon either co- or post-translationally, *i.e.* while translation is still ongoing or after the complete polypeptide has been synthesized and released from the ribosome. These two modes of insertion into/translocation across the IM are mediated by different targeting pathways: the ‘co-translational targeting pathway’ (section 6.1) and the ‘post-translational targeting pathway’ (section 6.2), respectively (Fig. 4, pathways II *versus* III & IV).

The targeting pathway of a protein to the Sec-translocon mainly depends on the hydrophobicity of the cleavable N-terminal ‘signal sequence’ of secretory proteins or of the non-cleavable so-called ‘signal anchor sequence’ of IMPs, which subsequently will constitute the first TM (62–67). Additionally, the targeting pathway can be affected by the translation rate: A fast translation rate increases the risk of protein misfolding and thus also the need of early chaperone-assistance (39,63,68,69). The recruitment of early chaperones and targeting factors, in turn, plays a role in the determination of targeting pathway.

Membrane translocation/insertion by the Sec-translocon requires, independent on targeting pathway, that the protein is kept in a ‘translocation competent’ state, *i.e.* the protein must be kept unfolded prior to interactions with the Sec-translocon. Evidently, in the co-translational pathway the polypeptide is per definition protected from pre-mature folding since it is inserted into the IM *via* the Sec-translocon as it is synthesized. In the post-translational targeting pathway, however, the fully synthesized polypeptide chain must be protected from premature folding by chaperones.



**Fig. 4.** Folding/Biogenesis pathways of cytoplasmic, inner membrane and secretory proteins. **I.** TF interacts strongly with ribosomes synthesizing cytoplasmic proteins. TF, DnaK/J and GroEL/ES chaperone systems constitute a cytoplasmic chaperone network, assisting the folding of cytoplasmic proteins. The cytoplasmic chaperone network may also play a role in the biogenesis of some membrane and secretory proteins. **II.** Highly hydrophobic signal sequences of IMPs and secretory proteins favour SRP binding and thus co-translational protein insertion into/translocation across the IM. **III & IV.** RNC binding to SecA and/or TF is stabilized by secretory signal sequences of proteins that are post-translationally translocated across the IM. Soluble SecA targets the polypeptide chain to the Sec-translocon, either in a Sec-independent (**III**) or SecB dependent (**IV**) manner. Upon translocation, Lep cleaves off the signal sequence from secretory proteins. When the polypeptide has been translocated across the IM, it is either folded into its native conformation (folding steps not represented in this figure, see section 7.4), or it can be targeted for further transport to the OM *via* SurA/Skp where the BAM-complex assists insertion of OPMs into the OM, or secreted to the extracellular milieu (not represented in this figure, see section 8). IM-associated SecA fuels the translocation of secretory proteins and large periplasmic loops of inner membrane proteins across the IM. The SecDFYajC complex and other auxiliary components play a role in Sec-mediated IM insertion and translocation (not included in this figure, see section 7.2). YidC can insert IMPs independently on the Sec-translocon, but the targeting pathway has not been characterized in detail yet. Additionally, mRNAs encoding membrane proteins may be targeted to the IM independently on ribosome binding (not represented in this figure). Figure modified after (64).

## 6.1 The co-translational insertion/translocation pathway

Most IMPs are inserted into the membrane co-translationally and are thereby targeted to the IM *via* the ‘co-translational targeting pathway’ (Fig. 4, pathway II). Secretory proteins can also be targeted for co-translational translocation across the IM, although less frequently (62,63,68). Co-translationally inserted/translocated proteins are, according to the commonly accepted view, delivered to the IM by the targeting factor ‘signal recognition particle’ (SRP). SRP-mediated targeting is triggered by the emergence of a highly hydrophobic ‘signal anchor sequence’, which often, already in the ribosome exit tunnel, has folded into an  $\alpha$ -helix, which will constitute the first TM of the IMP upon IM insertion (62–64,67). Once SRP has bound to the ‘signal anchor sequence’, it targets the RNC to the IM with the help of the SRP receptor FtsY (70–72). Both SRP and FtsY bind GTP, and upon SRP-FtsY interaction, GTP hydrolysis induces conformational changes resulting in the transfer of the RNC to the Sec-translocon while SRP and FtsY are released (62,73). The ribosome remains attached to the translocon until the complete polypeptide has been inserted into, or translocated across, the IM.

Notably, it has been observed that some mRNAs encoding membrane proteins can localize to the IM before association to the ribosome (71,74).

## 6.2 The post-translational translocation pathway(s)

The post-translational translocation targeting pathway is the main pathway for secretory proteins (Fig. 4, pathways III & IV) (62,64). Importantly however, there are examples of both periplasmic proteins and OMPs that are targeted to the IM *via* the SRP pathway. In fact, the two model proteins used for periplasmic protein production and OMP production (DsbA-BL1 and HbpD-ESAT6) in papers III, IV and V are delivered to the IM in an SRP-mediated fashion (68,75).

The branching of the ‘co-/post-translational’ pathways occurs already early on during translation. As described above, SRP binding to the RNC is favored when the signal sequence is highly hydrophobic. When other, less hydrophobic, signal sequences emerge from the ribosome, the binding to other targeting factors is favored. TF may be the first factor in the ‘post-translational’ targeting pathway and competes with SRP for binding to the RNC (45,46). It has also been observed that the ATPase SecA can bind to the RNC and, alone or in concert with other chaperones, target the nascent chain to the Sec-translocon for

post-translational translocation (76–79) (Fig. 4, pathway III). However, most attention has been focused on another post-translational translocation targeting pathway (Fig. 4, pathway IV). Many post-translationally translocated secretory proteins are dependent on the cytoplasmic chaperone SecB, which prevents premature folding and mediates targeting to the IM (79–81). The prevailing view is that SecB binds (both co- and post-translationally) to various hydrophobic stretches along the polypeptide subsequent to that TF-RNC association has occurred. Then, after release from the ribosome, SecA is recruited for targeting the nascent polypeptide chain to the Sec-translocon, see *e.g.*, (78,79,82). SecA exists both as a soluble protein in the cytoplasm and as a peripherally associated IMP. The IM-associated SecA has an essential role in Sec-translocon mediated translocation across the IM, as described in the following section.

## 7 Destination: inner membrane & periplasm

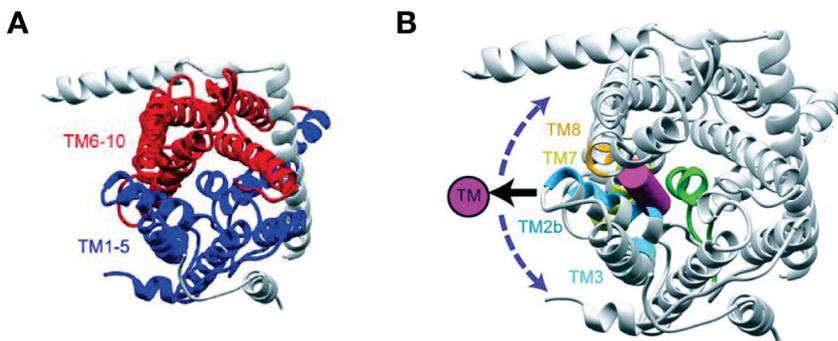
As mentioned before, insertion of membrane proteins into the IM is, in most of the cases, achieved in a co-translational manner. Translocation of proteins across the IM into the periplasm occurs mostly in a post-translational manner, but can also occur co-translationally. In any case, both insertion and translocation is mediated by the SecYEG-translocon, with or without assistance of SecA, YidC, the SecDFYajC complex or other auxiliary components, as described in the two following sections.

### 7.1 The Sec-translocon

The Sec-translocon is a hetero-oligomeric complex forming an aqueous pore that spans the IM. It translocates proteins across and inserts proteins into the IM in an unfolded state. The core unit of the Sec-translocon is composed of three integral membrane proteins (Fig. 5): SecY (composed of two halves of 5 TMs each, forming the hourglass shaped pore of the translocon), SecE (acts like a clamp to hold the SecY pore together) and SecG (function not known), in a 1:1:1 ratio (83). Interestingly, both SecY and SecE are essential and conserved throughout all kingdoms of life (62). Auxiliary components (*e.g.* SecA, YidC and SecDFYajC) and their different organizations with the SecYEG-complex are described in the next section. The crystal structure of the Sec-translocon, from the thermophilic methanogenic archaea *Methanococcus jannaschii*, was

solved in 2004 by Bert van den Berg *et al.* (83). The elucidation of the structure generated models of the mechanisms of translocation and insertion (Fig. 5). These models have by subsequent studies (using other pro- and eukaryotic organisms) shown to be, in its essence, mostly correct (84–88). Also, the similarity of the structures of the translocon generated from evolutionary distant organisms shows that it is highly conserved (89). Notably, however, SecE has 3 TMs in *E. coli* but only 1 in *M. jannaschii*. According to the *M. jannaschii* structure and later studies, there is a hydrophobic ring at the constriction of the ‘SecY-hourglass’, and at the interface between the two halves of SecY the translocon can be opened laterally towards the lipid bilayer (83,86,87,90) (Fig. 5). Furthermore, the translocon is partially filled by an  $\alpha$ -helical “plug”, which was suggested to be displaced upon signal sequence recognition (83). Recently, however, it was reported that the movement of the plug during translocation is only modest (87,88). Rather, the function of the plug, and the hydrophobic ring at the constriction of the pore, may be to keep the translocon impermeable to ions and other small molecules (91). Co- and post-translational protein transport is initiated by interactions between the Sec-translocon and the ribosome and/or SecA, respectively (87,88,90,92).

During translocation, the hydrophobic signal sequence binds to the Sec-translocon at the lateral gate, and the polypeptide immediately downstream of the signal sequence is inserted into the pore as a hairpin (88,93). The complete polypeptide is then threaded through the translocon and the signal sequence is,



**Fig. 5.** Structure of the *Methanococcus jannaschii* Sec-translocon. **A.** The two halves of SecY (blue and red) are ‘clamped’ together by SecE (white). **B.** The opening of the lateral gate (made up of TM 2, 3 and 8) is indicated by the two purple arrows. The movement of an inserting TM (pink) through the lateral gate is illustrated by the black arrow. The plug is represented in green. Reprinted with permission from (83).

in *E. coli*, cleaved by the signal peptidase (Lep), releasing the protein into the periplasm (94). Importantly, translocation of proteins through the pore is dependent on the previously mentioned membrane-bound *E. coli* ATPase SecA. Through multiple rounds of ATP-hydrolysis, SecA generates the energy needed for the translocation process (79). Translocation is also dependent on the proton motif force (PMF) (81,95) and presumably also on additional “pulling forces” from the periplasmic side of the IM (88). Recently, it has been suggested that interactions between charged residues of the translocating polypeptide and the periplasmic surface of the IM affect the dynamics of the translocation process (96). For a comprehensive review on the energetics during protein translocation, see (97). A completely translocated polypeptide can either obtain its native fold in the periplasm (including lipoproteins), as described in section 7.4, or it can be targeted to the OM, as described in section 8.

In the case of protein insertion into the IM, interactions between the highly hydrophobic ‘signal anchor sequence’ and the Sec-translocon promote the opening of the lateral gate (98). The pushing force exerted by the (ribosome-mediated) elongation of the polypeptide chain (at the expense of GTP hydrolysis) may partially drive the co-translational partitioning of the  $\alpha$ -helices into the lipid bilayer (81). Interactions between different TMs and between a TM and the lipids in the membrane also promote the insertion (99). Furthermore, Cymer and von Heijne recently showed that hydrophilic interactions between polar residues of the inserting and preceding  $\alpha$ -helices contribute to the partitioning (100). Importantly, however, efficient translocation of large periplasmic domains/loops requires SecA (68,101,102). The PMF also aids in the insertion of membrane proteins into the IM, see *e.g.*, (81). An active role of the Sec-translocon in regulating the insertion kinetics has also been proposed (98). Further insights regarding membrane protein insertion, folding and topology in the inner membrane will be discussed in section 7.3.

## 7.2 Auxiliary Sec-translocon subunits

The SecYEG complex is assisted by a variety of auxiliary subunits making the translocon composition highly modular (103,104). Here, I will present those subunits that have been most thoroughly characterized so far. Note that SecA has already been described in the previous section.

YidC is an essential and highly conserved integral IMP consisting of 6 TMs (105). The prevailing view is that YidC receives TMs as they emerge from the lateral gate of the SecYEG translocon and assists their insertion into the IM (106–108). YidC also appears to play a role in the folding and assembly of IMPs (109–112), as will be discussed in the following section. Importantly, YidC does not only assist the SecYEG complex, but can also by itself promote membrane protein insertion, thus constituting a second type of “insertase machinery”, see *e.g.*, (113–115). The targeting pathways to the YidC insertase are still elusive. Some YidC-dependent proteins appear to follow a co-translational route (supported by that *E. coli* YidC is able to bind to the ribosome, *e.g.*, (116)), but other studies have shown that YidC is also able to mediate post-translational membrane protein insertion (114,117–119). Presumably, there are specific ‘YidC pathways’ not yet discovered (116) (Fig. 4). Recently, the crystal structures of the *E. coli* YidC and of another YidC homologue were solved, both showing that YidC forms a hydrophilic groove opened towards both the cytoplasm and the IM (120–122). Proteins inserted in a SecYEG-independent manner by YidC are limited in size, which is in agreement with the geometry of the hydrophilic groove. Their insertion is independent of SecA, *i.e.* they do not contain any large periplasmic loops (118). Interestingly, YidC also seems to play a role in the biogenesis of the autotransporter OMP Hemoglobine protease (Hbp), the protein used in paper V, expanding the current idea of YidC function (123).

In papers I, III and IV, YidC is used as a model protein for the optimization of recombinant membrane protein production. Due to the function of YidC, one can question the use of YidC in these studies since elevated levels of YidC may generate a situation in which improved model protein production yields are a result of the intrinsic function of YidC itself. However, it has been shown that the recombinant production of a dysfunctional YidC gives the same production yields and cellular stress responses as for production of functional YidC (124).

Another complex involved in SecYEG-mediated protein insertion and translocation is the SecDFYajC complex. This complex associates with the SecYEG core in different structural arrangements: either with a SecA-bound SecYEG, for post-translational translocation, or with a ribosome-bound SecYEG for co-translational transport. The presence of SecDFYajC affects the efficiency of integration/translocation and makes the protein transport more dependent on the PMF (104), in correlation with the previously assumed function of SecDF, *i.e.* transducing energy from the PMF. Thus, SecDFYajC may regulate the inser-

tion/translocation process according to the properties of the substrate protein and also in order to modulate protein composition of the membrane during different growth phases/stress situations (104). Additionally, SecDFYajC has been suggested to constitute a link between SecYEG and YidC, although YidC is capable of contacting SecYEG also in the absence of SecDFYajC (108,125).

The membrane-associated protein Syd has been shown to interact with SecYEG and to have a destabilizing affect on the trimeric core unit, especially when improperly assembled (126). This observation, made *in vitro*, suggests that Syd has a ‘SecYEG proofreading’ function (126).

The IM-localized protein YidD has been shown to be involved in the insertion and processing of some YidC-dependent IMPs (127). PpiD and YfgM are two other auxiliary Sec-translocon factors presumably involved in the link between Sec-translocon mediated protein transport and the periplasmic chaperoning network, further mentioned in section 7.4 and 8.

### 7.3 Insertion and topology of inner membrane proteins

$\alpha$ -helical bundle IMPs can adopt different topologies. The N-terminus can either be positioned on the cytoplasmic side or on the periplasmic side of the IM, and the same holds for the C-terminal end of the protein. Furthermore, a TM can be tilted or can even constitute a so-called ‘re-entrant’ loop, with its N- and C-termini on the same side of the IM, see *e.g.*, (128). The topology of IMPs can be set during insertion, mainly dictated by the topology of the first TM, or it can be “flipped” after insertion into the IM, see *e.g.*, (129,130). It is known that the hydrophobicity and length of the TM, the distribution of positively charged residues of the cytoplasmic loops connecting the TMs, known as the ‘positive inside rule’ (131), and the lipid composition of the membrane all can affect the topology of IMPs (132–136). There are however many open questions in this field. Whether a cytoplasmic localization of the N-terminus of the first TM is acquired by a ‘flip’ post-membrane-insertion, or whether it is acquired pre-membrane-insertion by being inserted into the Sec-translocon as a ‘hairpin’ is still a matter of debate (132,137–139). It is also not clear as to how subsequent TMs are inserted, *i.e.* whether they can be accommodated in the Sec-translocon pore and inserted into the IM simultaneously or not. Highly hydrophobic TMs may never enter the Sec-translocon pore but rather ‘slide’ directly *via* the lateral gate into the IM (140). It has also been shown that the translation kinetics can

affect the insertion propensity of marginally hydrophobic TMs (141). Some poorly hydrophobic TMs can initially be positioned at either one of the membrane leaflets and only post-translationally be inserted into the membrane, driven by interactions with adjacent intramolecular TMs (142–144). This late insertion of a TM may then reorient the overall topology of the entire protein. Poorly hydrophobic TMs can also be retained near to the Sec-translocon pore *via* polar protein-protein interactions, which allows for proper folding before membrane partitioning (145). In fact, pre-mature membrane partitioning of marginally hydrophobic TMs can cause membrane protein aggregation in the IM (146).

According to Cymer and von Heijne (100), tertiary folding of membrane proteins can be initiated already at the immediate proximity of the Sec-translocon, in a co-translational fashion. As previously mentioned, YidC is believed to receive TMs as they emerge from the lateral gate of the translocon and assist their insertion and folding. It has been shown that the *E. coli* IMPs LacY and MalF require YidC for proper folding (109–111). It has also been seen that depletion of YidC leads to cell envelope stress responses due to membrane protein misfolding, further pointing towards a significant role of YidC as a membrane protein foldase (109,147,148). Folding of IMP loop regions is presumably assisted by cytoplasmic chaperones such as DnaK.

## 7.4 Folding and processing of periplasmic proteins

When a polypeptide emerges into the periplasm, it can be targeted further to the OM insertase machinery (BAM-complex), as described in the following section, or it can with the help of chaperones and folding catalysts attain its native structure in the periplasm. In addition, lipoproteins are attached to the periplasmic leaflet of either the IM or the OM, as briefly described at the end of this section.

Many periplasmic proteins, and also OMPs, contain one or more intramolecular disulfide bond(s). The periplasmic model protein BL1, a single-chain variable antibody fragment (scFv), used as a model protein in papers III and IV is a disulfide bond-containing protein. Disulfide bonds are usually crucial for protein structure and function and are only formed in the oxidizing environment of the periplasm. Disulfide bond formation has recently been reviewed, see (149). The generation of a disulfide bond requires oxidation of the two substrate cysteines. This is carried out by the soluble, periplasmic oxidoreductase DsbA, which contains one disulfide bond itself that is reduced during the disulfide bond

formation in the substrate protein. The electrons transferred from the substrate to DsbA are then further passed on to the IMP DsbB, which subsequently transports them to the Q-pool in the IM. Thus, both DsbA and DsbB are re-activated and ready for another round of substrate disulfide bond formation. A third component of the disulfide-bond formation machinery is the periplasmic protein DsbC. This is an isomerase that ‘senses’ erroneous disulfide bonds and subsequently reduces them. The substrate can then undergo another attempt of disulfide bond formation by DsbA. DsbC is re-activated by the IMP DsbD in a process that requires cytoplasmic thioredoxin.

In addition to disulfide bond formation, periplasmic proteins usually require assistance from other chaperones for attaining their native structure. The majority of the most prominent and/or best-characterized periplasmic chaperones have however been shown to mainly be involved in chaperoning/targeting of OMPs. SurA is one of the best-studied periplasmic chaperones and has, so far, only been shown to be essential for OMP biogenesis, see *e.g.*, (150–152), as described in the following section. Anyhow, there is also evidence that many of the periplasmic chaperones can assist the folding of both periplasmic proteins and OMPs. One rather recently discovered periplasmic chaperone, Spy, was first found to act mainly on periplasmic substrates, but recent observations point towards a role in OMP biogenesis as well (153,154). Also, the IM-anchored chaperone PpiD appears to support the folding of a general scope of newly translocated polypeptides, including those destined for the periplasm (103,155,156). The recently discovered Sec-translocon auxiliary component YfgM is a single spanning IMP believed to play a role in the interplay between the Sec-translocon and the periplasmic chaperone machinery and stress responses (157–159). Furthermore, Skp, one of the main OMP chaperoning factors, has shown to also have periplasmic substrates (154,160,161). FkpA is yet another periplasmic chaperone that has been shown to mediate the biogenesis/folding of both soluble proteins and OMPs (162,163). Lastly, DegP was previously thought to have a chaperoning role in the periplasm, but recent discoveries suggest that DegP mainly exhibits a protease function (164), see section 9. The role of periplasmic chaperones in OMP targeting will be discussed in section 8.

As mentioned before, lipoproteins are also present in the periplasm. Their biogenesis has been reviewed in (165). Most lipoproteins are anchored to the OM, rather than to the IM. Lipoproteins are equipped with a cleavable N-terminal signal sequence, and they contain at the cleavage site a so-called ‘lipo-

box'. At the very N-terminus of the processed (cleaved) protein there is a cysteine residue, which is essential for the enzymatic acylation of the lipoprotein, enabling lipid mediated OM/IM insertion. The so-called LolABCDE-system recognizes the lipoproteins destined for the OM and transports them across the periplasm upon which, they are inserted into the OM.

## 8 Destination: outer membrane & extracellular milieu

Proteins targeted to the OM are, in the majority of the cases, post-translationally translocated across the IM. When a polypeptide destined for the OM emerges into the periplasm, chaperones are recruited in order to prevent misfolding and to keep the polypeptide chain in a partially unfolded, OM-insertion-competent state. Some chaperones also function as targeting factors, directing the polypeptide to the  $\beta$ -barrel assembly machinery (BAM) complex in the OM (Fig. 4).

As mentioned before, many of the periplasmic chaperones exhibit a dual function. They are involved both in the folding of periplasmic proteins and in the biogenesis of OMPs. Also, most of them appear to have overlapping functions since many of them can rescue the lack of another. Anyhow, the currently most accepted model is that the chaperone SurA plays a major role in OMP biogenesis and that the chaperone Skp acts in a secondary 'rescue' pathway (166,167). Simultaneous deletion of these apparently parallel pathways is lethal to *E. coli* (168).

OMPs are equipped with a C-terminal 'signature sequence' enriched in aromatic residues (169). SurA has shown to recognize unfolded OMPs by these conserved, OMP-specific, C-terminal signature sequences (170). Importantly, SurA is the only periplasmic chaperone that, so far, has been shown to interact directly with the BAM-complex (171). Together, these observations have led to the idea that SurA may bind polypeptides during IM-translocation and stay attached until the substrate is delivered to the BAM-complex for OM insertion (167). Skp appears to bind to conserved hydrophobic patterns (that subsequently fold into a  $\beta$ -barrel) in yet unfolded polypeptides (172–174). It is not known whether Skp can contact the BAM-complex.

The BAM-complex assists OMP insertion into the OM and consists of the OMP BamA and four lipoprotein subunits: BamB, BamC, BamD (presumably involved in substrate recognition and regulation of BamA activity) and BamE, most likely in a 1:1:1:1 ratio (Fig. 4) (151,175). BamA and BamD constitute

the central core of the BAM-complex and are both essential for cell survival. Their deletion leads to accumulation of aggregated OMPs in the periplasm (176,177). However, also the BamB,C,E subunits are required for efficient substrate recognition and insertion, see *e.g.*, (167). The BamA protein consists of a  $\beta$ -barrel integrated into the OM with soluble loops protruding towards the periplasm, known as ‘polypeptide transport associated’ (POTRA) domains. The *E. coli* BamA has five POTRA domains, which are involved in association with the lipoprotein subunits and in substrate recognition (169) and insertion into the OM (178). Recently, two crystal structures of the *E. coli* BamA  $\beta$ -barrel, and of two additional homologs, were reported (179–181). This, together with previous findings, has shed light on how BamA assists in the folding and insertion of OMPs into the OM in *E. coli*. Two models have been suggested: the ‘BAM-assisted model’ and the ‘BAM-budding model’ (167,182). Roughly, the BAM-assisted model suggests that the main function of the BAM-complex is restricted to destabilizing, or ‘priming’, the OM for subsequent spontaneous OMP folding into the OM, thereby being responsible for localizing the OMP substrate to these primed regions. The BAM-budding model implies a more active role of BamA during folding and insertion. According to this model, the opening of a ‘lateral gate’ of BamA is triggered upon recognition of the C-terminal signature sequence, which is presumably inserted into the cavity of the  $\beta$ -barrel as a hairpin (181,183). Subsequently, POTRA domains, together with the lipoprotein subunits, thread the polypeptide into the BamA  $\beta$ -barrel. The lateral gate acts as a nucleation point for the  $\beta$ -barrel folding of the substrate. In a last step, the substrate OMP ‘buds’ off from the BamA  $\beta$ -barrel and is released into the lipid bilayer. The priming function of BamA is thought to be required also in the BAM-budding model, due to energy requirements of the insertion process.

Most proteins destined for the extracellular milieu traverse the cell envelope by other mechanisms, commonly divided into six principal groups: type I - VI. The groups differ in *e.g.* their targeting mode to the inner membrane and whether they span both the IM and the OM at once or in a two-step process. Their fates after translocation are also diverse. They can remain in the OM, be released into the extracellular milieu or be inserted into an adjacent target cell. For a comprehensive and recent review on Gram-negative secretion pathways, see *e.g.*, (184). The secretion mechanism of type Va autotransporters will be described in section 15.

# Protein homeostasis in *E. coli*

To maintain protein homeostasis and to enable regulation of cellular processes, all proteins must, at some point, be degraded. In addition, proteins can be incorrectly synthesized or the chaperone system can fail in its attempt to fold a substrate. The dynamic properties of proteins and their surroundings may cause unfavorable changes, which negatively affect protein structure and organization. To maintain a healthy physiology, *E. coli* produces proteases and peptidases that are involved in the degradation of all kinds of proteins and peptides, respectively (section 9). Importantly, the chaperone-network described in section 5 also contributes in the cellular combat against aggregation by functioning as “re-foldases”. However, under stressed conditions the aggregation prone misfolded proteins tend to accumulate to levels not tolerable for cell survival. Therefore, *E. coli* has developed stress response systems, which enhances the expression of certain genes encoding *e.g.* proteases and chaperones (section 10). These stress-responses are commonly observed during recombinant protein production.

## 9 Protein degradation

There are five major proteolytic systems in the *E. coli* cytoplasm: ClpXP, ClpAP, HslUV, Lon and FtsH. They are all ATP-dependent ATPases, of the AAA+ type, containing one peptidase domain and one ATPase domain (185,186). The peptidase domain is located in the inner cavity of a  $\beta$ -barrel structure, which can be accessed through a narrow pore (187–192). The ATPase domain is located at the top of the  $\beta$ -barrel and unfolds protein substrates prior to their entry into the barrel. Once the protein substrate is inside the barrel, it is exposed to the catalytic site and fragmented into smaller peptides. These peptide fragments are then released for further degradation by peptidases.

FtsH is the only one of these proteases that has shown to be essential for *E. coli* viability (due to its role in LPS synthesis) (193). This protease is further

distinguished from the others by the fact that it is a membrane protein. It has a soluble catalytically active domain in the cytoplasm, which is anchored to the IM *via* two N-terminal TMs (190,194–197). FtsH is involved both in degradation of cytoplasmic proteins and IMPs, and appears to play a key role in quality control of IMPs. FtsH has the ability to extract IMPs from the IM and subsequently degrade them (198,199). Interestingly, FtsH has been shown to degrade SecY upon clogging of the Sec-translocon (194,200). It has been suggested that FtsH works in close collaboration with YidC and the two ‘FtsH inhibitory proteins’ HflK and HflC (201–203), forming a quality control network. Two other proteases that play a role in IMP homeostasis are HtpX and GlpG. The production of the membrane anchored HtpX is induced during the heat shock response (described in the next section) and has been shown to have overlapping functions with FtsH (204). The IMP GlpG is a rhomboid protease cleaving misfolded IMPs inside the membrane (205–207).

DegP, DegQ and DegS are the best-studied periplasmic proteases shown to play key roles in quality control in the periplasm. These are all serine proteases and belong to the ‘High temperature requirement A (HtrA)’ protease family. HtrA proteases are important for the *E. coli* ‘heat shock’ stress response, described in the following section. As mentioned before, DegP has both a chaperone and protease activity, but seems to mainly function as a protease (164,208–210). Binding to specific degradation sequences induces oligomerization of DegP, generating a cavity in which the substrate is accommodated (211,212). Seemingly, proteins that are able to fold in this cavity are released and only those that fail to fold will be degraded (211). DegQ appears to be highly similar to DegP regarding mechanism and architecture (213–216). However, expression of the gene encoding DegQ is not up-regulated upon heat shock response (213). Also, it is believed that the ‘chaperone activity’ dominates over the ‘protease activity’ for DegQ, as opposed by the situation for DegP (214). The third member of the HtrA family is the essential IMP DegS, of which the proteolytic activity is also induced only upon substrate recognition (94,213,217). This protease does, in contrast to the other two, not form higher oligomeric states. Importantly, DegS is part of the cell envelope stress response, as described in the following section.

One important aspect in protein degradation is substrate recognition. The selection of proteins, native or denatured, has to be highly accurate not to harm important cellular functions. Some proteases such as *e.g.* FtsH, which has a

weak unfolding capacity (198,218), mainly recognize denatured proteins based on hidden ‘degradation sequences’ (degrons) that only become accessible upon misfolding of substrate proteins. Other proteases act specifically on certain (native) proteins that *e.g.* upon proteolysis activate cellular pathways (159,199,219–222). This implies that some proteases exhibit more of a “quality control” function whereas others primarily function in protein homeostasis maintenance and signaling/regulation. Notably, proteins can also be tagged for destruction. The best known example of this is the SsrA-tagging of truncated peptides arisen due to stalled translation (199). Further regulation of proteolysis is achieved by so-called ‘adaptors’, which bind to and modulate the activity of proteases (199).

## 10 Stress responses

*E. coli* has developed various types of stress-responses in order to be able to react efficiently to different kinds of adverse conditions. As mentioned previously, recruitment of RNAP to promoters is regulated by ‘ $\sigma$  factors’.  $\sigma$  factors direct RNAP to promoters controlling the expression of certain sets of genes (223). These genes encode proteins and regulatory RNAs that act to combat the effects of the stress. For the sake of clarity: Many proteins that are part of a stress response are also produced under non-stressed conditions, *via*  $\sigma^{70}$  dependent RNAP promoter recruitment, but to lower levels (224). Here, I will describe the *E. coli* stress responses that are most relevant for my PhD studies, *i.e.* those involved in protein misfolding and aggregation.

The ‘heat shock response’ is orchestrated by  $\sigma^{32}$  and is triggered by *e.g.* elevated temperatures and/or misfolded (overproduced) proteins (225,226). Some of the proteins that are upregulated during the heat shock response have previously been described: DnaKJ, GroEL, FtsH, Lon and HtpX. Additional important heat shock response proteins are the AAA+ protease ClpB and the inclusion body binding proteins IbpA and IbpB. Both ClpB and IbpA/B are believed to assist DnaK in attempting to resolve protein aggregates and assisting re-folding of misfolded proteins (227). The levels of  $\sigma^{32}$  are regulated mainly on a translational (upon elevated temperatures) and post-translational level (upon protein misfolding and aggregation). The  $\sigma^{32}$  mRNA has been shown to form secondary structures, which inhibit association with the ribosome. Upon elevated temperatures, these secondary structures are resolved, which allows the synthesis of  $\sigma^{32}$  (228). However, the presence of misfolded proteins/aggregates can,

independently of temperature shifts, induce the heat shock response. In this case it is believed that the levels of  $\sigma^{32}$  are upregulated post-translationally: DnaKJ and GroEL not occupied with a substrate (*i.e.* under non-stressed conditions) sequester  $\sigma^{32}$ . Upon protein misfolding and aggregation these chaperones bind unfolded substrates, which causes the release of  $\sigma^{32}$ , and thus induction of the heat-shock response (225,229,230). Proteases may regulate  $\sigma^{32}$  levels in a similar manner. FtsH and HslUV have been shown to degrade  $\sigma^{32}$ , a process that declines upon protein misfolding as the proteases become occupied with more important matters (231).

There are also stress responses directly linked to cell envelope stress (148,232). One of these is the ' $\sigma^{24}$  induced stress response'.  $\sigma^{24}$  directs RNAP to genes encoding various periplasmic chaperones and proteases such as the previously mentioned FkpA, Skp, SurA and DegP (233–235). RNAP bound to  $\sigma^{24}$  is also recruited to the genes encoding both  $\sigma^{32}$  and  $\sigma^{24}$  (236–239). The levels of  $\sigma^{24}$  are mainly regulated post-translationally. Under non-stressed conditions  $\sigma^{24}$  is associated to the IM and upon protein misfolding/aggregation in the periplasm a cascade reaction is triggered leading to the release of  $\sigma^{24}$ , which then can associate with RNAP (240). The periplasmic protease DegS plays an important role in the triggering of the cascade reaction (241).

Another cell envelope stress response is based on the CpxAR 'two component system', which consists of the IM-bound sensor histidine kinase CpxA, the periplasmic protein CpxP and the cytoplasmic regulator protein CpxR (236,242). Under non-stressed conditions, CpxP inhibits CpxA, and upon protein misfolding in the periplasm, CpxP becomes occupied by binding to the misfolded proteins, which subsequently leads to the activation of CpxA. Through an autophosphorylation event, CpxA transfers a phosphate to CpxR, which then acts as a transcriptional activator of genes encoding *e.g.* DegP, DsbA and  $\sigma^{32}$  (243–246).

# Recombinant protein production

Recombinant protein production has revolutionized life science research and industrial biotechnology. The possibility to produce proteins in suitable host organisms has made that cumbersome protein isolation, of *e.g.* medically important proteins such as insulin, from natural sources is not longer needed. Also, it has made it possible to engineer proteins with ‘improved’ characteristics and to equip them with *e.g.* tags that improve/facilitate their production, purification and detection.

## 11 Recombinant protein production bottlenecks

Despite the huge potential of recombinant protein production, obtaining sufficient amounts of functional protein is not always trivial. For example, the proteins produced may be inactive or degraded, or yields may not be sufficient, often due to poor growth of the host organism. Some proteins are particularly difficult to produce. Production of membrane proteins is limited by their complex biogenesis (see section 6-8) and their hydrophobic nature, which makes them prone to aggregation upon mis-targeting. The difficulty of producing membrane proteins is best illustrated by the low number of membrane proteins of which the structure has been solved compared to the number of structures solved for soluble proteins (<http://www.rcsb.org/pdb/home/home.do>). This despite the importance of membrane proteins in health and disease: Approximately 70% of all existing drugs act on membrane proteins (247). Some soluble proteins are, however, also difficult to produce, *e.g.* those containing disulfide bonds, such as *e.g.* antibody fragments and most hormones, and therefore cannot be folded properly in reducing environments such as the bacterial cytoplasm. In bacteria, these proteins have to be directed into the periplasm. Many industrially and medically valuable proteins are produced in the bacterial periplasm for yet additional reasons, such as lower proteolytic activity and facilitated protein puri-

fication due to a reduced number of contaminants in this compartment as compared to the cytoplasm. However, just like membrane proteins, the biogenesis of secretory proteins is complex (see section 6-8), which makes yields usually low. It is known that one of the main bottlenecks when producing IMPs and secretory proteins is the insufficient capacity of the Sec-translocon (124). ‘Clogging’ the Sec-translocon leads to *e.g.* accumulation of misfolded proteins (including the recombinant protein) in the cytoplasm, which leads to low cell densities due to the toxic burden, further lowering protein production yields (124). For a thorough study of the consequences of membrane protein production in *E. coli*, see (124). It should be noted that there are reports that rather favor the idea that there is no general obstacle to the production of membrane and secretory proteins (248,249).

The main aim of my PhD studies has been to improve recombinant production of membrane and secretory proteins in the Gram-negative bacterium *E. coli*. For one particular application I have also used *Salmonella*, see section 14-16. Here, it is important to state that, throughout my PhD studies, optimization strategies are aimed at producing membrane proteins inserted into the membrane, *i.e.* not in inclusion bodies (protein aggregates in the cytoplasm). Membrane-integrated membrane proteins can be extracted from the membrane in a properly folded state relatively easy with the help of detergents, whereas it is usually impossible to isolate any properly folded material from inclusion bodies. In the following sections I will describe different strategies for optimizing the production of recombinant proteins in the *E. coli* bacterial cell envelope. First however, I will briefly discuss alternative protein production hosts.

## 12 Protein production hosts

When designing a protein production strategy, the first decision to make is whether to use a prokaryotic or a eukaryotic host. Sometimes it can also be better to produce a protein in a ‘cell-free’ system. Due to space limitations I will however not describe cell-free protein production, see instead *e.g.*, (250–252). Some elements of the protein biogenesis machineries significantly differ between eukaryotic and prokaryotic cells. Therefore, not all proteins with a eukaryotic origin can be successfully produced in bacteria. The prime example of this is the lack of a glycosylation machinery in *E. coli*. Also, the differences in transcription and translation rates between bacteria and eukaryotic organisms can

negatively affect eukaryotic protein production in bacteria. In such cases, yeast, insect, fungal, mammalian, or plant cells may be a better choice. When possible, however, if the aim is to produce high quantities of protein, the use of bacteria is usually more efficient. Bacteria grow faster, they are cheaper and easier to maintain/cultivate and they are physiologically and genetically less complex. Notably, bacteria can also be manipulated such that they become more suitable for the production of eukaryotic proteins, see *e.g.*, (253).

*E. coli* is by far the most commonly used bacterial protein production host, but other bacteria have also proven to be suitable: *e.g. Lactococcus lactis*, *Bacillus subtilis*, *Pseudomonas* ssp. *Streptomyces* and *Mycobacterium*, see *e.g.*, (254–257). Usually, the best-suited host is the one that is most closely related to the organism from which the protein originates (258). It is likely that further developments will lead to an increase in the use of bacteria other than *E. coli* for the production of recombinant proteins. Thus far though, the historical dominance of the use of *E. coli* both as a bacterial model organism and as a protein production host has made that the available tool-box for this organism is still superior to that of other bacteria and to that our knowledge of this organism is unmatched.

## 12.1 *E. coli*

*E. coli*, a rod-shaped, 1-2  $\mu\text{m}$  long Gram-negative bacterium, is part of the microbial flora of the intestine of warm-blooded organisms and was first isolated about one hundred years ago by the German microbiologist and pediatrician Theodor Escherich (259,260). Existing *E. coli* strains range from harmless ‘laboratory strains’ to strongly pathogenic subspecies such as the enterohemorrhagic *E. coli* strain EHEC. Early *E. coli* ancestors gave rise to the so-called B-lineage, including BL21 and derivatives thereof, *e.g.* BL21(DE3), and to the K-lineage, which includes for example MC4100 and various cloning strains (259). In my work I have used strains belonging both to the B- and the K-lineage. It is important to note that even strains from the same lineage can differ significantly from each other due to the extensive manipulations that *E. coli* has been subjected to throughout the years. Thus, one must be cautious when comparing results obtained using different strain backgrounds.

## 13 Optimization strategies

Recombinant protein production requires, apart from a suitable protein production host, also a host-compatible expression vector. The expression vector harbors the gene encoding the recombinant protein and other essential elements. Carefully designing expression vectors often leads to improved protein production (section 13.1). Another strategy to achieve improved protein production yields is to optimize the host organism (section 13.2). Furthermore, the gene encoding the recombinant protein can be manipulated/designed in order to enhance production yields or to facilitate its detection and/or purification (section 13.3). One can also screen for optimized culturing conditions (section 13.4). Main focus will here be on those strategies that have been used for optimization of membrane and secretory protein production.

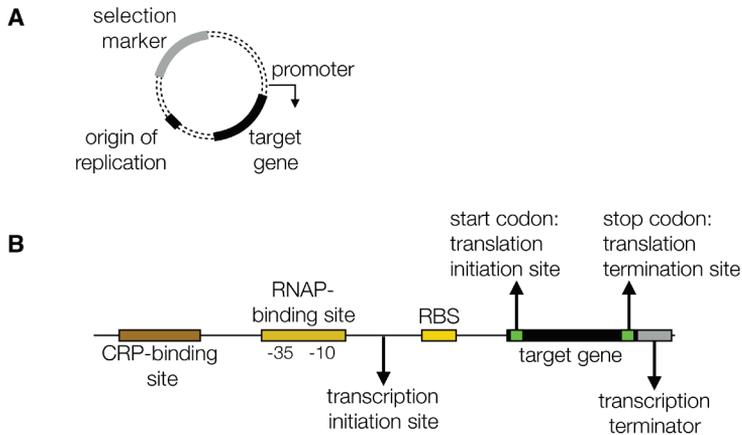
### 13.1 Expression vector design

Bacteria naturally contain small circular double stranded DNA molecules, known as ‘plasmids’, which are ‘separated’ from the chromosomal DNA. Naturally occurring plasmids give the cell an advantage under certain circumstances, *e.g.* the ability to produce proteins that make the cell resistant to antibiotics. Importantly, plasmids have been exploited as expression vectors when producing recombinant proteins (Fig. 6A). The vast repertoire of genetic manipulation and cloning techniques makes it possible to insert any desired gene/nucleotide sequence into a plasmid.

#### 13.1.1 Choice of promoter system

As mentioned previously, the transcription of a gene is dependent on the recruitment of RNAP to its ‘promoter’ sequence, located upstream of the gene (Fig. 6). In *E. coli*, promoters contain two defined regions of six nucleotides each, located 10 and 35 nucleotides upstream of the transcription initiation site, to which the RNAP binds (261) (Fig. 6B). Different genes have different promoters, which are regulated in different ways and that have different characteristics. Promoters can vary *e.g.* in their strength (*i.e.* transcription initiation frequency) and in their basal (*i.e.* non-induced) expression. The different characteristics of promoters affect both endogenous and recombinant protein production levels. The binding of RNAP to a promoter requires a certain factor or compound, which is specific for each type of promoter. These factors can *e.g.* be the

presence/absence of a small molecule (*e.g.* a metabolite) or increased/decreased temperatures. Many of the promoters used for recombinant protein production are taken from operons involved in sugar utilization. In the following sections I will focus on the promoters most relevant for the papers included in this thesis.



**Fig. 6. A.** Schematic representation of an expression vector (plasmid). Essential elements are indicated. **B.** Schematic representation of a promoter region of a sugar utilizing operon, including the binding sites for RNAP and cAMP-CRP. The ribosome-binding site (RBS), the target gene and the translation initiation and termination sites are also represented.

### 13.1.1.1 The *lac* promoter

The *E. coli* lactose-induced promoter (the *lac* promoter) (262) has been studied in great detail and is also frequently used for recombinant protein production (Fig 7A). In the absence of lactose, the *lac* promoter is repressed by the ‘*lac* inhibitor protein’ (LacI), which binds at certain ‘operator sites’ within the promoter region, as reviewed in *e.g.*, (263). When lactose enters the cell it binds to LacI, which induces a conformational change leading to the dissociation of LacI from the operator sites and, thereby, also to the initiation of transcription of the genes governed by this promoter (263). The genes that are naturally transcribed from the *lac* promoter encode proteins that enable lactose import (LacY) and utilization of lactose as an energy and carbon source (LacZ and LacA). However, when the *lac* promoter is used for recombinant protein production from a plasmid, these genes are replaced by any gene(s) encoding recombinant proteins.

Induction of expression of genes governed by the *lac* promoter additionally requires binding of the ‘cAMP receptor protein’ (CRP) to the so-called CRP

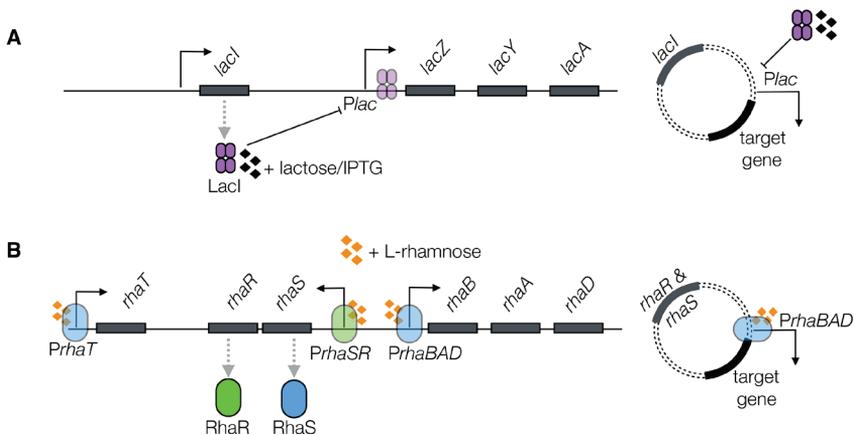
binding site upstream of the two previously mentioned RNAP binding sites (Fig. 6B) (264). The ability of CRP to bind the promoter is dependent on the presence of cAMP, which in turn depends on the glucose levels in the cell. When glucose is absent the levels of cAMP are high, whereas upon addition of glucose the levels of cAMP decrease, leading to that CRP is unable to activate transcription. Thus, glucose can be used to repress expression of genes controlled by the *lac* promoter. This mechanism is common for sugar utilization promoter systems and is referred to as ‘catabolite repression’. The purpose of catabolite repression is to enable bacteria to utilize sugars in a sequential manner (265). Catabolite repression is used in recombinant protein production to repress leaky promoters prior to the desired time point for induction of expression. However, it may sometimes be sufficient to insert a copy of the gene encoding LacI into the plasmid, which is a commonly used strategy to reduce background expression (Fig. 7A). Catabolite repression is further exploited in so-called autoinduction media, as described in section 13.4.

Importantly, also lactose-derivatives can be used for the induction of *lac*-based promoters. The non-metabolized lactose analogue ‘Isopropyl  $\beta$ -D-1-thiogalactopyranoside’ (IPTG) is the most commonly used inducer of *lac*-based promoters for recombinant protein production (Fig. 7A). In standard *E. coli* strains, transcription rates from *lac*-derived promoters are hardly titratable but rather seem to operate in an on/off mechanism. However, in mutants lacking the lactose transporter LacY target gene expression rates are more titratable (266). Papers I, II and III of this thesis are based on a frequently used *lac* promoter-derived expression system. Variations of this system are described in sections 13.2.1, 13.2.2.1 and 13.2.3.1.

#### 13.1.1.2 The *rhaBAD* promoter

Another promoter that is used for recombinant protein production is the so-called *rhaBAD* promoter, which is derived from the operon responsible for utilization of the sugar L-rhamnose (Fig. 7B) (267). The *E. coli* rhamnose operon contain three promoters that are activated upon addition of rhamnose (268,269): (i) the *rhaT* promoter governing the expression of the gene encoding the protein RhaT, which transports rhamnose into the cell, (ii) the *rhaBAD* promoter governing the expression of the genes encoding the proteins RhaB, RhaA and RhaD, which all three are involved in the utilization of rhamnose as a carbon and energy source and (iii) the *rhaSR* promoter governing the expression of the two

regulatory proteins RhaS and RhaR. Upon addition of rhamnose, background levels of RhaR activate the transcription from the *rhaSR* promoter. The subsequent production of RhaS leads to the activation of transcription from both the *rhaT* and the *rhaBAD* promoters. In addition, excess amounts of RhaS appear to downregulate transcription from its own promoter (269). Just like the *lac* promoter, the rhamnose induced promoters are susceptible to glucose-mediated catabolite repression (268). However, even in the absence of glucose these promoters are still rather silent (267). This ‘tight’ feature is generally advantageous for the production of toxic proteins such as membrane and secretory proteins. Of the three rhamnose inducible promoters in the rhamnose operon, it is usually the *rhaBAD* promoter that is exploited for recombinant protein production (267) (Fig. 7B). The protein levels produced from genes transcribed from this promoter correspond to the concentration of rhamnose added to the culture medium (267). This titratable feature has been further employed for the development of the Lemo21(DE3) protein production platform (270), as described in section 13.2.3.1. In paper IV, we made a rather unexpected observation. The correlation between the concentration of rhamnose and protein production levels in *E. coli* wild-type appear to be caused by consumption of the inducer rather than by a gradual increase in protein production rate. This has implications for the potential of the promoter in its use for production of membrane and secretory proteins.



**Fig. 7. A.** Schematic representation of the *E. coli lac* operon. **B.** Schematic representation of the *E. coli* rhamnose operon. The use of the *lac* promoter and *rhaBAD* promoter for recombinant protein production using plasmids is also shown. Note that binding sites for RNAP and cAMP-CRP are not shown.

### 13.1.1.3 Other promoters used for recombinant protein production

In this section I will briefly describe a few other promoter systems that have successfully been used for production of recombinant proteins in *E. coli* and/or *Salmonella*.

The arabinose inducible promoter, *araBAD*, is a relatively weak promoter that is susceptible to catabolite repression (271). The *araBAD* promoter-based gene expression depends on the regulator protein AraC (272). This regulator acts both as an activator and as an inhibitor. AraC represses transcription in the absence of arabinose and activates transcription in its presence. Arabinose can be transported into the cell *via* two transporters: AraE (low affinity) and AraFGH (high affinity) (272). In wild-type *E. coli* strains the expression intensity from this promoter is not adjustable, but rather exhibit an all-or-nothing expression phenotype (273,274). By producing the low affinity arabinose transporter from an arabinose independent constitutive promoter (275,276) or by substituting both arabinose transporters with a mutant LacY (277,278), cultures are homogeneous, *i.e.* the production yield is the same in all cells in a culture.

In paper V, we used the tetracycline (*tetA*) promoter for the production of the phage ‘lysis protein E’ in *Salmonella* (see section 12). In *E. coli*, the antiporter IMP TetA confers resistance to tetracycline (279,280). The *tet* operon also encode the *tetA* repressor protein TetR, which regulates the expression of *tetA*. The *tetA* promoter is a strong and tight (*i.e.* non-leaky) promoter, which, in contrast to the previously described promoters, is not susceptible to catabolite repression (279,280). Upon addition of the non-toxic tetracycline derivative anhydrotetracycline, TetR rapidly dissociates from the promoter, which results in a fast responsiveness of this promoter system.

Finally, I want to draw attention to heat-inducible expression systems in *E. coli*. These systems are usually based on the use of phage  $\lambda$  promoters and are activated by elevated temperatures, at which a heat sensitive variant of the  $\lambda$  repressor dissociates from the promoter, see *e.g.*, (281). In cases where the addition of a chemical compound is undesired, heat-inducible systems may constitute a valuable alternative.

### 13.1.2 Additional elements of an expression vector

For a plasmid to be able to replicate, and thereby be passed on to the daughter cell during bacterial cell division, it needs the so-called ‘origin of replication’ (282) (Fig. 6A). The sequence of the origin of replication determines the copy

number of the plasmid. The copy number of plasmids used for protein production can vary a lot and can affect the kinetics of gene expression. Therefore, one may contemplate to test plasmids with different ‘origins of replication’ to optimize protein production. Furthermore, not all plasmids can co-exist in the same cell and whether two different plasmids are ‘compatible’ or not depends on the ‘origins of replication’ of the two plasmids.

Additionally, a plasmid must contain a ‘selection marker’ in order to be maintained in the cell (Fig. 6A). The selection marker is usually a gene that encodes a protein that gives the cell resistance to a certain antibiotic. Thus, by adding the corresponding antibiotic to the culture medium, only the bacteria that contain the plasmid will survive. Only in the presence of the antibiotic the cell is forced to keep the plasmid in order to survive. Regarding the choice of an antibiotic selection marker, it is advisable to avoid the use of the ampicillin resistance cassette when producing membrane and secretory proteins. Ampicillin acts in the bacterial periplasm, and thus, the protein that confers the resistance to ampicillin ( $\beta$ -lactamase) has to traverse the IM, increasing the risk of saturating the Sec-translocon. Furthermore,  $\beta$ -lactamase tends to leak out from the periplasm into the surrounding medium, which causes the inactivation of ampicillin and thus makes this system rather unstable.

In line with attempts to reduce the use of antibiotics world-wide, alternative selection strategies are being developed. In paper V, I made use of mutant strains lacking the gene *asd*, which encodes a protein that is essential for cell survival due to its crucial role in peptidoglycan synthesis (283). Unpublished observations showed that these strains could be used for protein production in an antibiotic-free manner, by replacing the antibiotic resistance marker with the *asd*-gene. Antibiotics can also be avoided if the gene encoding the recombinant protein is located on the chromosome. Also, toxin-antitoxin systems can be used as alternatives to antibiotic-based systems (284).

Another important factor to consider when designing an expression vector is the position and sequence of the ‘ribosome binding site’ (RBS) (Fig. 6B), referred to as the ‘Shine-Dalgarno’ (SD) sequence. The SD-sequence is located upstream of the start codon of the gene and downstream of the transcription initiation site. Upon transcription, the SD-sequence in the mRNA will base pair with the anti-SD sequence on the 16S rRNA of the small ribosome subunit (285). If these sequences are highly similar, translation initiation is efficient whereas the use of an SD-sequence that differ from the *E. coli* anti-SD-sequence

will result in less efficient translation initiation and thus lower protein production rates. The distance between the RBS and the start codon also affects the efficiency of translation initiation and it has been shown that a distance of approximately eight nucleotides leads to the highest efficiency of translation initiation (285). Any strong secondary mRNA structures (*i.e.* hairpins) should be avoided since these disrupt accessibility of the mRNA to the ribosome, see *e.g.*, (286).

## 13.2 Optimization of the protein production host

As discussed in section 11, protein production in *E. coli* can be cumbersome. Therefore, attempts have been made to improve *E. coli* as a protein production host. Here, I will describe some of the most successful developments and important findings regarding the optimization of *E. coli* for the production of membrane and secretory proteins.

### 13.2.1 The BL21(DE3)/pT7 protein production platform

In paper III of this thesis, protein production is driven by the frequently used *lac*-promoter-derived BL21(DE3)/pT7 protein production platform, developed by Studier and colleagues in the early 80's of the last century (287) (Fig. 8). Notably, this platform was in first instance developed for production of soluble proteins rather than for production of membrane and secretory proteins. The *E. coli* strain BL21 is devoid of two major proteases, OmpT and Lon, which can interfere with protein production (288,289). With the help of the  $\lambda$ -derived phage 'DE3', the gene encoding the phage's RNAP, 'T7 RNAP', was integrated into the genome of BL21, resulting in BL21(DE3). T7 RNAP only recognizes T7 bacteriophage specific promoters, which are not naturally occurring in the *E. coli* genome and not recognized by the endogenous *E. coli* RNAP. T7 RNAP transcribes genes about five to eight times faster than the *E. coli* RNAP (290). The T7 RNAP itself is produced from a gene transcribed by the *E. coli* RNAP, under the control of the strong *lac*-derived *lacUV5* promoter (291,292). To exploit the T7 RNAP for recombinant protein production, the gene encoding the target protein is placed on an expression vector (pT7) under the control of a T7*lac* hybrid promoter (Fig. 8). This hybrid promoter is composed of the T7 promoter and a LacI-operator site, to keep the promoter 'silent' prior to induction (293). Thus, upon addition of *e.g.* IPTG, T7 RNAP is produced to high levels and subsequently used to drive the expression of the target gene from the

now also available T7lac hybrid promoter (Fig. 8). The BL21(DE3)/pT7 system generates high levels of mRNA of the target gene upon the addition of IPTG. This is often favorable for soluble protein production. However, when producing membrane or secretory proteins, this high load of mRNA often leads to protein accumulation rates that are higher than what the capacity of the biogenesis machinery of IMPs and secretory proteins can manage. This results in, amongst other things, target protein aggregation and poor growth of the host (124,270). To reduce the mRNA levels, a common approach is using the so-called pLysS and pLysE plasmids to co-express the gene encoding T7 lysozyme, which is a natural inhibitor of the T7 RNAP (294).

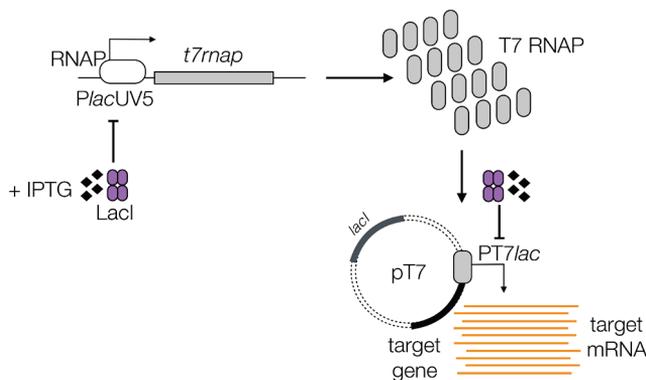


Fig. 8. Schematic representation of the BL21(DE3)/pT7 protein production system.

### 13.2.2 Isolation of improved *E. coli* protein production hosts

*E. coli* can accumulate mutations that can lead to variants with improved membrane and secretory protein production characteristics. The prime example is the isolation of the so-called Walker strains, as described in the following section. There are also other examples of the isolation of *E. coli* strains with improved membrane and/or secretory protein production characteristics, which also will be described but more briefly, in the subsequent section.

#### 13.2.2.1 Isolation of the Walker-strains

The limitations of the BL21(DE3)/pT7 system for the production of toxic proteins such as membrane proteins were recognized early on by Miroux and Walker (295). In an attempt to isolate mutants able to efficiently produce toxic pro-

teins, they exposed BL21(DE3) to the production of the bovine oxoglutarate malate carrier protein (OGCP), which naturally resides in the mitochondrial IM. Upon addition of IPTG to BL21(DE3) cells that harbored the gene encoding OGCP on a T7 promoter-based vector most cells died as a consequence of the toxicity caused by the production of OGCP. Screening for surviving (*i.e.* IPTG resistant) cells that still produced OGCP upon the addition of IPTG, led to the isolation of the first Walker-strain, C41(DE3). Subsequently, in a similar manner, C41(DE3) was screened for its ability to produce the *E. coli* IMP F<sub>0</sub>F<sub>1</sub>ATPase subunit b. The production of this protein is highly toxic even to C41(DE3). This led to the isolation of the second Walker-strain, C43(DE3).

The use of the Walker-strains is not limited to the production of the target proteins used during their isolation, but they have been proven useful for a large set of difficult-to-produce proteins (295,296). It is now known that the main mutations responsible for the improved performance of C41(DE3) and C43(DE3) are located in the sequence of the *lacUV5* promoter, which controls the expression of the gene encoding T7 RNAP in BL21(DE3) (270). Importantly, these mutations convert the *lacUV5* promoter into a much weaker promoter. Consequently, the levels of T7 RNAP, upon addition of IPTG, are much lower in the Walker strains than in BL21(DE3), which results in a lower expression rate of the gene encoding the target protein. Therefore, the capacity of the protein biogenesis machinery is not saturated in the Walker strains. This finding led to the development of the protein production strain Lemo21(DE3) (270), as described in section 13.2.3.1. Furthermore, Schlegel and colleagues could recently show that the mutations in the *lacUV5* promoter occurred by *recA* dependent recombination between the wild-type *lac* promoter and the *lacUV5* promoter (297). Importantly, C43(DE3) has additional mutations in the *lacI* copy of the DE3 region, resulting in a LacI variant that binds stronger to its operator site. This lowers basal expression levels of the target gene and results in a slower induction of target gene expression upon the addition of IPTG (298).

#### 13.2.2.2 Isolation of other *E. coli* protein production hosts

Not surprisingly, also other mutants better suited for the production of membrane and secretory proteins than wild-type *E. coli* have been isolated. In some cases, mutagenesis has been facilitated by the use of *e.g.* transposons or chemicals (mutagenic agents), whereas others rely on the spontaneous occurrence of mutations, like in the case of the isolation of the Walker-strains.

In a recent study carried out by Gul and colleagues (299), four different *E. coli* strains (NG2, NG3, NG5 and NG6) with enhanced membrane protein production characteristics were generated upon overproduction of different IMPs. These four strains were isolated using a selection strategy in which green fluorescent protein (GFP) was used as a folding reporter (see section 9.3.2.1) and the erythromycin resistance protein (ErmC) was used to select for increased IMP production by culturing cells in the presence of increasing amounts of antibiotic. Common to all four strains is that they all carry different mutations in the gene encoding the protein H-NS (299). A potential explanation behind the improved performance of these strains may be that the transcription of genes involved in the membrane protein biogenesis machinery is altered due to the role in transcriptional silencing of H-NS.

Another example is the isolation of the so-called EXP strains (300). Here, mutagenesis was forced both by chemical means and by using a strain in which the DNA polymerase proofreading function was deficient. The gene encoding the target membrane protein was simultaneously expressed from two separate plasmids. On one plasmid the target gene was fused to a gene encoding the protein that renders bacteria resistant to kanamycin and on the other plasmid to a gene that renders bacteria resistant to trimethoprim. Thus, clones with improved membrane protein characteristics could easily be selected by the ability to grow on increasing amounts of these two antibiotics. The cause behind the improved performance of these strains is not understood, but for one of the strains it is known that the copy number of the plasmid used for the recombinant protein production was lowered. Thus, it appears likely that the explanation, in that case, is linked to a reduced load on the biogenesis machinery (300).

Other means to induce genetic modifications with the aim of isolating superior protein production host strains is the use of transposon-mediated mutagenesis. This is exemplified by a study where the aim was to improve the production of a certain G-protein coupled receptor (GPCR) (301). The GPCR was fused to GFP (see section 13.3.2.1) enabling the isolation of clones with improved GPCR production characteristics by fluorescence activated cell sorting (FACS). A strain in which a transposon insertion had interrupted the gene encoding the DnaK co-chaperone DnaJ was found to exhibit an improved ability to produce the GPCR. Potential explanations behind the improved production characteristics of this strain may be that DnaJ otherwise competes with the tar-

getting factor SRP for binding to the signal anchor sequence of the GPCR or that DnaJ is involved in the degradation of this particular target protein (301).

In a subsequent study published by the same authors it was found that co-expressing the *ybaB* gene, which encodes a putative DNA-binding protein (YbaB), led to a prominent increase in the production levels of several IMPs (302). This result was found after screening the co-expression of a library containing all known *E. coli* open reading frames. Since transcription of *ybaB* is mediated *via*  $\sigma^{24}$ , it is likely that its co-expression positively affects the quality control of overproduced IMPs (302). In this study the target proteins were fused to the gene encoding  $\beta$ -lactamase, rendering bacteria resistant to increasing amounts of ampicillin.

Yet more recently, the same research group published a study in which they screened for improved GPCR production by co-expressing a library of approximately  $10^7$  different *E. coli* chromosomal fragments (249). Here, selection of clones was based on two parallel systems, one monitoring production levels by fusing the GPCR to GFP and the other monitoring activity of the GPCR *via* binding to a fluorescent ligand. Co-expression of three different gene fragments resulted in improved GPCR production: the *nagD*, *nlpD* genes and the *ptsN-yhbJ-npr* gene cluster. Since it was observed that some of the stress-response-related genes were upregulated upon the co-expression of *nagD* and *nlpD*, one possible mechanism by which the co-expression of these gene fragments improves yields of functional GPCRs could be that it ‘primes’ the cell for the production of IMPs (249).

A screening method for the generation of strains with improved performance of producing secretory proteins (scFv antibody fragments) was developed by Chen and colleagues (303). In this method (periplasmic expression with cytometric screening (PECS)), *E. coli* cells producing scFvs in the periplasm are incubated with a fluorescent conjugate of the target ligand, which can traverse the OM and bind to the scFv of concern (which cannot leak out from the periplasm). Only ligands that are bound to its corresponding scFv are retained in the periplasm. Subsequently, the cells exposing the highest fluorescence are selected using FACS. This method was later used in combination with exposing cells to a mutagenic agent to isolate strains with improved full-length antibody (IgG) production characteristics (304).

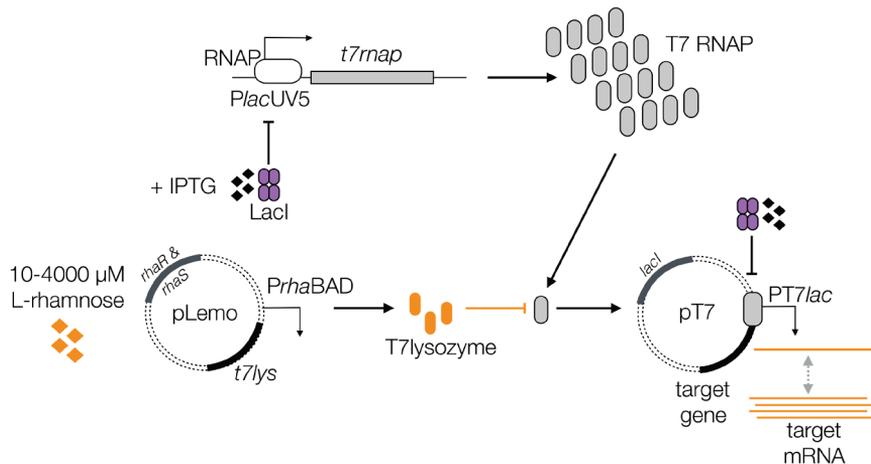
### 13.2.3 Developing *E. coli* hosts via targeted engineering strategies

If the factors that limit the production of a certain type of protein are known, one may use targeted engineering approaches to develop *E. coli* strains with improved protein production characteristics. Such strategies can be e.g. co-production of certain chaperones (section 13.2.3.3) or deletion of genes encoding proteases known to negatively affect protein production yields (section 13.2.3.2). Other approaches can involve modifications of the recombinant production regime, which can lead to altered protein production kinetics that results in improved protein production yields (following section).

#### 13.2.3.1 Tuning membrane protein production rates in Lemo21(DE3)

It has become clear that fine-tuning protein production rates provides an efficient strategy to overcome bottlenecks in the production of both membrane and secretory proteins (305–311). This is further supported by the development of the Lemo21(DE3) strain, as described below, and by papers I-IV of this thesis. The sequence of the gene encoding the target protein can also be modified to decrease transcription and/or translation rates, as discussed in section 13.3.

The development of the Lemo21(DE3) strain (Fig. 9), used in papers I and II, was influenced by the characteristics of the Walker-strains (270). Lemo21(DE3) is a BL21(DE3)-derivative and thus, the gene encoding the recombinant protein is placed under a *T7lac* promoter on a pT7 vector and is transcribed by the chromosomally encoded T7 RNAP, which itself is under the control of the IPTG inducible *lacUV5* promoter (see section 13.2.1). The advantageous feature of Lemo21(DE3) is that protein production rates can be tuned precisely, which makes that this strain is suitable for the production of a wide range of increasingly ‘toxic’ targets (270,312). The key behind the tunable characteristics of Lemo21(DE3) is the *rhaBAD* promoter-based co-expression of the gene encoding T7 lysozyme, a natural inhibitor of T7 RNAP, from a plasmid named pLemo (Fig. 9). As previously mentioned, levels of proteins expressed from the *rhaBAD* promoter depend on the concentration of the inducer rhamnose (see section 13.1.1.2). Thus, the levels of T7 lysozyme, and thereby also the activity of T7 RNAP and the protein production rate, depend on how much rhamnose is added to the culture medium (Fig. 9). By varying the concentration of rhamnose, an optimal protein production rate can be found for each protein target. At this “sweet-spot” the Sec-translocon is not saturated, which leads to optimal protein production yields.



**Fig. 9.** Schematic representation of the Lemo21(DE3) strain. For comparison of the Lemo21(DE3) strain with the BL21(DE3) strain, see also Fig. 8.

### 13.2.3.2 Generation of knockout strains

If a certain protein is likely to negatively affect protein production yields, one may delete the gene encoding the protein of concern. In paper IV of this thesis, genes of the rhamnose utilizing operon have been deleted/inactivated in favor of a more controlled protein production regime of the *rhaBAD* promoter.

In a study by Nannenga and Baneyx it was shown that deleting the gene encoding TF from the *E. coli* genome had beneficial effects on the recombinant production of some IMPs (313). In line with the reasoning regarding the transposon insertion in the gene encoding DnaJ, as mentioned in section 13.2.2.2, the authors suggested that the improved performance of the TF deficient strain is due to a reduced competition with SRP. However, co-production of SRP did not improve production yields. Also for secretory protein production it has been shown that the deletion of the gene encoding TF leads to higher protein production yields (314).

Membrane protein production yields could be improved in *E. coli* strains in which genes of the phosphoenolpyruvate:phosphotransferase system (PTS) were deleted (315). These mutants had a slower growth rate and accumulated lower levels of acetate. The inhibitory role of acetate accumulation in production of IMPs was corroborated by the observation that the ‘low-acetate-producing strain’ BL21(DE3) (316) could also produce the IMPs used in the study.

When producing proteins in the periplasm, or on the surface of the OM, production yields have, in some cases, been shown to be limited by proteolytic activity, see *e.g.*, (317–319). In these cases, the use of mutants lacking one or more proteases can be beneficial (320). It has also been shown that inactivation of the heat shock response by deleting the gene encoding  $\sigma^{32}$  can lead to improved protein production yields in the periplasm (320,321). It should however be kept in mind that the deletion of proteases (and chaperones) can result in lower quality of the protein produced. This is exemplified by a study in which it was shown that DegP is necessary for the secretory production of functional alkaline phosphatase (PhoA) (322). Furthermore, the use of deletion mutants can increase the leakiness of secretory proteins into the surrounding media (323). Secretory proteins can also be produced in OM vesicles by the use of ‘hyper-vesiculating’ deletion mutants, such as the ones used in (324).

### 13.2.3.3 *Co-production of protein biogenesis factors*

The opposite of deleting a gene encoding a protein that negatively affects protein production is to co-express a gene encoding a protein that is likely to improve production yields. It can be a component of the protein biogenesis machinery or another factor specifically assisting the production of a certain protein. Again, note that I only will focus on examples dealing with the production of membrane and secretory proteins.

In an attempt to increase the production levels of five different human GPCRs, different components of the Sec-translocon and of the chaperone machinery assisting IMP biogenesis were co-produced (325). This exercise resulted in the finding that the IMP protease FtsH can improve production yields of this kind of membrane proteins significantly, whereas co-production of FtsY, SecYE, DnaKJ, SecB, GroEL/ES and TF had only modest positive effects. The increased levels of FtsH did however not result in more active recombinant proteins. Corroborated by a subsequent study, the authors attribute the beneficial effect of overproduction of FtsH to an additive induction of heat-shock related genes, which “primes” the cell for the production of the GPCR of interest (325,326). This is in line with the reasoning regarding the co-production of NagD and NlpD, as mentioned in section 13.2.2.2.

Chen and colleagues could show that the co-production of DnaK and DnaJ resulted in higher levels of membrane-integrated CorA, an *E. coli* magnesium transporter (327). The increase in levels of properly folded CorA was mainly

explained by a reduction of inclusion body formation of this target upon its overproduction. Co-production of the targeting factors SRP, SecB and SecA negatively affected CorA-levels. The results of this study are - to some extent - contradicting the study in which it was found that a transposon mediated disruption in the gene encoding DnaJ was beneficial for GPCR production (301), as discussed in section 13.2.2.2. This discrepancy is probably due to the strongly protein target dependent effect of co-expressing or deleting genes on production yields.

In the study in which it was shown that deleting the gene encoding TF could improve IMP production yields, as mentioned in section 13.2.3.2, it was also observed that the co-production of YidC in some cases led to a higher amount of membrane-integrated IMPs (313). Although the co-production of SRP has not shown to positively affect IMP production (313,327), it appears to increase yields of co-translationally translocated produced proteins in the periplasm (314,328). These differences may be explained by that the main bottleneck of the targets produced in these different studies is not the same.

Co-production of folding catalysts and chaperones, both cytoplasmic and periplasmic, has been extensively studied for production of periplasmic proteins (304,329–331). The co-production of DsbC when producing proteins containing disulfide bonds, such as antibodies and antibody fragments, *i.e.* scFvs and Fabs, has proven especially successful (332–335). For additional strategies on improving disulfide bond formation in the periplasm, see *e.g.*, (149). It has also been shown that co-production of FkpA, Skp and SurA can improve secretory protein production yields of *e.g.* antibody fragments in *E. coli* (335–338).

#### 13.2.3.4 Disulfide bond formation in the *E. coli* cytoplasm

An alternative approach to produce disulfide bond-containing proteins in *E. coli* is to engineer strains that allow for disulfide bond formation in the cytoplasm. There are different ways of how disulfide bonds can be obtained in the cytoplasm (149): (i) disruption of the main reducing pathways of *E. coli* by deleting the gene encoding glutathione reductase (339) and/or thioredoxin reductase (340), together with suppressor mutations in the gene encoding alkyl hydroperoxidase (341,342), (ii) co-production of the mitochondrial intermembrane space sulfhydryl oxidase (Erv1) (343,344) or (iii) redirecting the native periplasmic Dsb-system such that DsbA is produced in the cytoplasm and the topology of the IMP DsbB is inverted (345). Additionally, producing DsbC in

the cytoplasm facilitates the biogenesis of proteins containing multiple consecutive disulfide bonds (341,344,346). Some of these approaches have generated commercially available strains, *e.g.* Origami (342) and SHuffle (346).

### 13.3 Modifying the target protein and the gene encoding it

Rather than manipulating the protein production host, the target protein can itself be modified by altering the sequence of the gene encoding it (section 13.3.1) or by the use of N- or C-terminal tags (section 13.3.2). Whatever the modification, one should always keep in mind that it may affect the folding/function of the protein in one way or another.

#### 13.3.1 Protein engineering and codon optimization

Modifying the sequence of the gene encoding the target protein is a common approach to improve protein production and/or purification. This can be done such that the amino-acid sequence is altered or by substituting rare codons with synonymous codons, more frequently used in the current host organism, leaving the amino-acid sequence unaltered. Also, the sequence upstream of the start codon of the gene can be manipulated to affect transcription- and/or translation initiation. I will first briefly describe some of the numerous reports of successful amino-acid substitutions resulting in improved production yields followed by a short paragraph on how synonymous codon usage, or modification of the sequence upstream of the start codon of the gene, have been used for optimizing membrane and secretory protein production.

It is difficult to predict which amino-acid substitutions will be beneficial for production of the target protein. Therefore, proteins with optimized amino-acid sequences are often generated *via* directed evolution approaches or random mutagenesis. Large libraries of sequences can be generated by *e.g.* error-prone PCR, recombination techniques, chemical mutagenesis or by using proofreading deficient DNA polymerases. To select the desired mutants from these large libraries efficient selection methods are required. One example of a recently developed selection method is the ‘cellular high-throughput encapsulation, solubilization, and screening method’ (CHESS) (347). This method was specifically developed for the generation of functional GPCRs with improved stability in a given detergent and the simultaneous isolation of the DNA carrying the gene encoding the ‘superior’ GPCR. The CHESS method has successfully been used for isolating CPCR suitable for crystallization, see *e.g.*, (348). Increased deter-

gent stability of IMPs is crucial for purification and structural analysis and is a field of extensive activity, pioneered by the laboratory of James Bowie, see *e.g.*, (349). Two recent reviews of the field of increasing stability of IMPs (GPCRs in particular) nicely covers the evolution of this field, see (350,351). The increase in IMP stability is often accompanied with higher production levels of the target protein, which have been attributed to enhanced IM insertion efficiency and/or reduced protein degradation, see *e.g.*, (352,353).

Random mutagenesis or directed evolution approaches are useful when pre-existing knowledge of how to improve certain protein characteristics is lacking. More rational designs have been achieved by computational strategies as exemplified by a study where advantage was taken of a pre-existing GPCR crystal structure for predicting stabilizing mutations (354). Another more controlled manner of changing the sequence of the gene encoding the target protein is the use of alanine scanning. Here, all residues in the protein are exchanged one by one to alanine, and the different mutants are then probed for its improved performance. This approach has successfully been used for the generation of more thermostable GPCRs, see *e.g.*, (355).

The coding sequence of a protein can also be modified without changing its amino-acid sequence: there is more than one codon for each amino acid and the use of a certain codon can be abundant in one organism whilst it is rare in another. A gene with many abundantly used codons has a high 'codon adaptation index' (CAI), whereas a gene containing many rare codons has a low CAI. The occurrence of a certain codon in an organism is correlated with the concentration of the respective tRNA in the cell (356). The CAI, together with occurrence of mRNA secondary structures, thus affects protein elongation rate (356–358). Some codon optimization strategies therefore aim at increasing the CAI as much as possible. Synonymous codon substitutions can, however, negatively affect mRNA stability, mRNA structure, translation initiation and/or elongation and protein folding (359). Thus, codon optimization must be done with care. Rather than optimizing codon usage in order to increase the elongation rate one must consider other important features. For example, it was recently observed that translation elongation of membrane proteins is paused at certain sites to facilitate membrane insertion (360). Also, various studies have shown that clusters of rare codons are crucial for co-translational folding, see *e.g.*, (356). Codon optimization strategies that mimic the original elongation rate profile of the gene seem to be the most promising approach to use.

It has been shown that the rate-limiting step in protein synthesis is translation initiation rather than elongation. The effect of CAI seems to rather play a role in global fitness of the cell: Genes with low CAI give rise to ribosome stalling, which leads to that endogenous essential proteins cannot be synthesized (361). Thus, changing the nucleotide sequence at the region of translation initiation may be more efficient to modify the protein accumulation rate rather than optimizing the codon usage. Recent studies point towards that the degree of mRNA folding around the translation initiation site is the main determinant of the rate of translation initiation, as assessed *via* GFP fluorescence accumulation (361,362). Thus, the idea is that a lower degree of mRNA folding at the translation initiation site leads to higher levels of recombinant protein production. However, these results may be a consequence of unusually strong mRNA secondary structures in the reporter genes used (*gfp*). Simmons and Yansura tested the effect of translation initiation strength on production levels of secretory proteins in the periplasm by using a library of vectors with modified translation initiation regions. It was observed that production levels of heterologous secretory proteins were improved by optimizing the translation initiation frequency for each target protein (309). Similar approaches can also improve the production of membrane proteins (286,363). Furthermore, by the use of transcriptional fusions upstream of the gene encoding the target protein membrane protein production levels could be dramatically improved (364). How this relates to protein synthesis rate is however not yet clear. Interestingly, in a study where random synonymous codons at the 5' coding region were screened for improving the activity of a cytoplasmic exocellulase protein, rather than for production levels of GFP (compare to (361,362)), no trend towards relaxed mRNA structure and thus faster translation rates, could be found (365).

### **13.3.2 Fusion partners and tags**

A fusion partner is usually attached either to the N-terminus or to the C-terminus of the target protein. The position of the fusion partner is important since the N-terminus of membrane and secretory proteins contains the information required for proper targeting and insertion/translocation across the IM. Thus, care must be taken when designing N-terminal fusions in order not to compromise the biogenesis of the target protein. Design of fusion strategies in the case of membrane proteins is also affected by the topology of the target IMP, *i.e.* in which compartment the fusion partner will end up.

Some of the most frequently used N-terminal fusions when producing membrane proteins are *e.g.* ‘maltose binding protein’ (MBP) and ‘mistic’. MBP has been shown useful to equip heterologous membrane proteins with a targeting signal recognized by the *E. coli* targeting machinery, see *e.g.*, (366,367), but has also been shown to increase yields of IM-integrated IMPs in general and to facilitate their purification, see *e.g.*, (368). Notably, MBP can be targeted to the Sec-translocon both co- and post-translationally. Mistic is a small hydrophobic protein from *Bacillus subtilis* that has been used successfully for enhancing the efficiency of membrane integration of heterologous membrane proteins (369,370). The mistic-mediated integration of IMPs into the IM appears to be independent on the Sec-translocon. Thus, the use of this tag may lead to that some of the main bottlenecks in membrane protein production are avoided (370–372). However, it appears as if the activity of IMPs can be negatively affected by mistic (373). The so-called N-terminal P8CBD-tag was used for the production of a functional YidC homologue and is targeted in a SRP-dependent manner to the Sec-translocon (306). This tag was constructed by combining the M13 phage major coat protein and one TM of *E. coli* Lep. Recently discovered proteins shown to be useful as N-terminal fusion partners for membrane protein production are bacteriorhodopsin and  $\beta$ -lactamase (249,302,374). However, the use of  $\beta$ -lactamase is thus far not known to increase IMP production levels but is rather used as a tool to monitor membrane protein topology. Furthermore, the use of different N-terminal ‘leader sequences’ has been shown to improve membrane protein production yields by reducing translation rates (307). Thioredoxin and GFP (see the following section) are frequently used C-terminal fusions and are thought to improve IMP production levels by increasing their stability.  $\beta$ -lactamase has been used as a C-terminal fusion partner of secretory proteins to probe proper folding in the periplasm (375). Furthermore, N-terminally fusing MBP to secretory proteins can improve yields in the periplasm, as exemplified by (317). However, N-terminal fusions to secretory proteins are mainly represented by the use of different signal sequences. This will be described in section 13.3.3.

#### 13.3.2.1 *The GFP pipeline*

The use of the jellyfish protein ‘green fluorescent protein’ (GFP) as a C-terminal fusion partner for monitoring production levels and topology of membrane proteins was first described by Drew and colleagues (376). It was based on the ob-

servations that the GFP moiety of the fusion only folds correctly, and thus fluoresces, on the cytoplasmic side of the membrane and not on the periplasmic side (377). Importantly, when IMP-GFP fusions are not correctly inserted into the membrane but instead form inclusion bodies in the cytoplasm, the GFP moiety does not fold properly, and does thus not fluoresce (378). GFP is a very stable protein and can easily be monitored in whole cells and can be detected in standard SDS-PAGE allowing for determination of the integrity of the IMP-GFP fusion, see *e.g.* (379). Also, the use of GFP as a C-terminal tag accelerates optimization of membrane solubilisation and purification strategies (380) and enables the use of fluorescence-detection size-exclusion chromatography (FSEC) for evaluation of the quality of IMP-GFP fusions (381). GFP can also be utilized to monitor protein production homogeneity of cell cultures and to determine protein production per cell using flow cytometry. Geertsma and colleagues took advantage of the folding properties of GFP and developed a method that allows to quantify not only the fluorescent IM-integrated IMP-GFP fusions, but also of non-fluorescent aggregated material (382). Thereby, optimization of IMP production not only relies on increasing properly folded IMPs, but also on decreasing the pool of IMPs not inserted into the IM. The distinction is based on differential migration rate of folded *versus* aggregated IMP-GFP fusions upon analysis by SDS-PAGE. The GFP-pipeline is extensively used in papers I-IV of this thesis.

One limitation of the GFP-pipeline is that only those IMPs with a topology such that the C-terminus is located on the cytoplasmic side of the membrane can be detected. However, it has been shown that the majority of *E. coli* IMPs have such a topology (383). Furthermore, IMPs with the opposite topology can be converted by simply adding one more TM to the C-terminus (384). Recently, the development of the bimolecular split-GFP complementation system has made it more feasible to also monitor the topology of the N-terminus. Here, only a 16 residue long fusion is required, which is thought to disturb targeting/translocation less than a fusion to the complete GFP protein (385). It should also be noted that other variants of fluorescent proteins have been developed with altered folding characteristics enabling also proper folding in the periplasm, *e.g.* superfolder GFP (386). Superfolder GFP is used in papers III and IV of this thesis as a model protein for secretory protein production.

### 13.3.3 The use of different signal sequences for secretory proteins

As described in section 6, the choice of targeting pathway to the Sec-translocon depends mainly on the hydrophobicity of the N-terminal signal sequence of the polypeptide emerging from the ribosome exit tunnel. The production of periplasmic proteins can sometimes be improved by increasing the hydrophobicity of their signal sequences and thus re-routing their IM targeting from the SecA/SecB-dependent pathway to the SRP-dependent pathway (63,68,69). This is often due to that the risk of misfolding/aggregation in the cytoplasm is lower during co-translational translocation, see *e.g.*, (329,387). However, different signal sequences destined for co-translational translocation can differ in their ability to promote high-level production of proteins in the periplasm. Thus, the selection of signal sequence for secretory protein production is important. The use of also heterologous signal sequences has in some cases shown to improve protein production in the periplasm (388).

## 13.4 Culture conditions

Another important factor when optimizing protein production protocols is the way the production host is cultured. One may modify *e.g.* temperature, composition of culture medium, gene expression induction strategy, feeding method (batch *versus* fed-batch), culturing volume or aeration. Here, I will mainly address the use of different culture media.

The use of different kinds of media will affect the physiological state of the host, which can lead to altered cell densities and production yields. Some media are better suited for incorporation of labeled amino acids whilst others have been optimized for *e.g.* straightforward upscaling or 'autoinduction'. *E. coli* cells, producing recombinant proteins, are routinely cultivated in the enzymatic casein digest- and yeast extract-based culture medium Lysogeny Broth (a.k.a. LB medium), which is the culture medium most frequently used for the *E. coli*-based production of membrane proteins for structural studies (<http://blanco.biomol.uci.edu/mpstruc/>). Other commonly used complex media are *e.g.* Terrific broth and 2xTY, both also based on enzymatic casein digest and yeast extract. There are also media where the components are well-defined. The main example of this is the minimal medium M9, consisting of essential salts to which desired carbon/energy sources, metals, amino acids or vitamins can be added. In paper II, a set of complex and defined media was used to screen for

improved membrane protein production in C43(DE3) and Lemo21(DE3) cells. Here, it was shown that the use of the minimal 'PASM-5052' autoinduction medium was best suited for membrane protein production. Also other studies support the finding that the use of autoinduction media can improve membrane protein production, see *e.g.*, (389,390). The key principle of autoinduction media is based on glucose mediated catabolite repression, here described in the context of *lac*-promoter based expression. Induction of expression of the gene encoding the recombinant protein, and of the chromosomally located *lacY*, *lacZ* and *lacA* genes, is first repressed by the presence of glucose (see section 13.1.1.1). This enables cells to grow in the absence of the toxic burden imposed by recombinant protein production. When the cells have consumed all the glucose, repression of the *lac*-based promoters is released, which leads to expression of the genes of the *lac*-operon. Production of LacY enables the cells to take up lactose, which subsequently leads to full induction of the gene encoding the recombinant protein. In this way, induction of the recombinant protein is thought to be 'smoother' than upon IPTG-mediated induction. William Studier was the first to develop autoinduction media by carefully examining the effect on protein production by the use of different components (391). In a subsequent study using autoinduction media developed by Studier for production of membrane proteins it was shown that it is beneficial to decrease the concentration of lactose in the media (389), which is in line with the idea that reducing protein production rates leads to improved membrane protein production yields. Notably, most autoinduction media are glycerol-based. It has been noticed that, while supporting high cell densities, the induction in glycerol-based autoinduction media is negatively affected by high aeration levels. Ukkonen and colleagues therefore developed a more robust autoinduction formulation based on a fed-batch like slow glucose feed to the culture media (392). Another strategy to circumvent the aeration sensitivity of glycerol-based autoinduction media is the somewhat contradictory addition of IPTG during exponential growth, as presented in paper II of this thesis.

It has been observed that M9-minimal medium can lead to improved protein secretion into the periplasm and to the OM surface as compared to LB-medium (310,319,388). This effect was, in some cases, attributed to a lower growth rate in the M9-medium allowing to balance target gene expression levels with the protein secretion capacity of the cell.

# Autotransporter-mediated surface display

In paper V of this thesis we have simultaneously produced the lytic bacteriophage  $\phi$ X174 membrane protein E, which converts Gram-negative bacteria into ghosts (see section 16), and the *E. coli* autotransporter hemoglobin protease-derived surface display platform (HbpD) in *E. coli* and *Salmonella*. The co-production of these two different membrane proteins can be used for the generation of multivalent vaccine platforms that are supposedly safer than live vaccines based on attenuated pathogens. In the following sections I will introduce the Gram-negative bacterium *Salmonella typhimurium* and the two aforementioned types of membrane proteins and their role in vaccine development.

## 14 *Salmonella typhimurium*

*Salmonella* was isolated more than a hundred years ago by Theobald Smith and Daniel E. Salmon and is just like *E. coli* a Gram-negative bacterium that has been extensively studied (393,394). *Salmonella* is rod-shaped and 1-2  $\mu$ m long, and has thus a very similar morphology as compared to *E. coli*. *Salmonella typhimurium* (*S. typhimurium*), used in our study, is a serovar of the species *Salmonella enterica* (395). Most *Salmonella* strains are highly pathogenic and cause gastroenteritis and typhoid fever in humans and other mammals. However, attenuated derivatives have been isolated which are suitable for laboratory work as well as for vaccine development. The immune-stimulatory properties of this bacterium have proven to be very suitable for development of bacterial based vaccines (394,396,397). Therefore, we chose, in addition to *E. coli*, an attenuated derivative of *S. typhimurium* (398) as a host for the dual membrane protein production strategy. Antigens from *Mycobacterium tuberculosis* were used as model antigens for surface display.

## 15 Autotransporters

Gram-negative bacteria have evolved various mechanisms to secrete proteins across the cell envelope (184). One of the most widely distributed and simple is the type V secretion mechanism, which contains five similar but distinct sub-groups. They are all exploited by bacteria to secrete factors involved in *e.g.* pathogenesis (399). Here, I will briefly introduce the type Va secretion pathway, a.k.a. the classical autotransporter (AT) pathway (400). Type Va ATs are synthesized as large precursor proteins that contain three domains (Fig. 10A): (i) a cleavable signal sequence (ss) at the N-terminus for targeting to and translocation across the IM *via* the Sec-translocon, (ii) an extracellular passenger domain that carries the effector function of the AT and (iii) a  $\beta$ -domain at the C-terminus that adopts a  $\beta$ -barrel conformation in the OM (Fig. 10B) and plays a role in transfer of the passenger domain across the OM. The  $\beta$ -barrel and the passenger are connected by a short linker, which usually forms an  $\alpha$ -helix embedded inside the  $\beta$ -barrel (Fig. 10B) and is essential for translocation of the passenger domain (401–406). Following translocation across the OM, some AT passenger domains remain attached to the  $\beta$ -barrel whereas others are cleaved and secreted into the extracellular milieu. The majority of the AT passengers tend to adopt the same fold: A long repetitive  $\beta$ -helix to which different side chains are attached, see *e.g.*, (399) (Fig. 10B).

Our understanding of the translocation mechanism of ATs across the OM is still rather poor. For a long time it has been thought that the  $\beta$ -barrel alone is responsible for the translocation of the passenger. While there is no doubt that the  $\beta$ -barrel is important for targeting and translocation (407–410), growing evidence show that also the BAM-machinery is involved (408,411,412). The current, but yet somewhat speculative model, suggests that the  $\beta$ -barrel of the AT in a first step is inserted into the OM *via* assistance from the BAM-machinery. Then, the AT  $\beta$ -barrel, in a non-completely folded state, stays attached to BamA, generating a large hybrid translocation channel, through which the passenger is threaded as a hairpin in a C-to-N-terminal direction (399,410). This model is supported by the observation that the pore diameter of the AT  $\beta$ -barrel is too narrow to alone translocate the passenger domain, which has been shown to form tertiary structure prior to translocation (413–415), (402–406). It also fits with the observations made by Noinaj and colleagues that BamA seems to be able to open laterally towards the OM (181,183), as mentioned in section

8. Several periplasmic chaperones have been shown to play roles in the secretion of ATs across the OM, albeit differently for different ATs, see *e.g.* review (416).

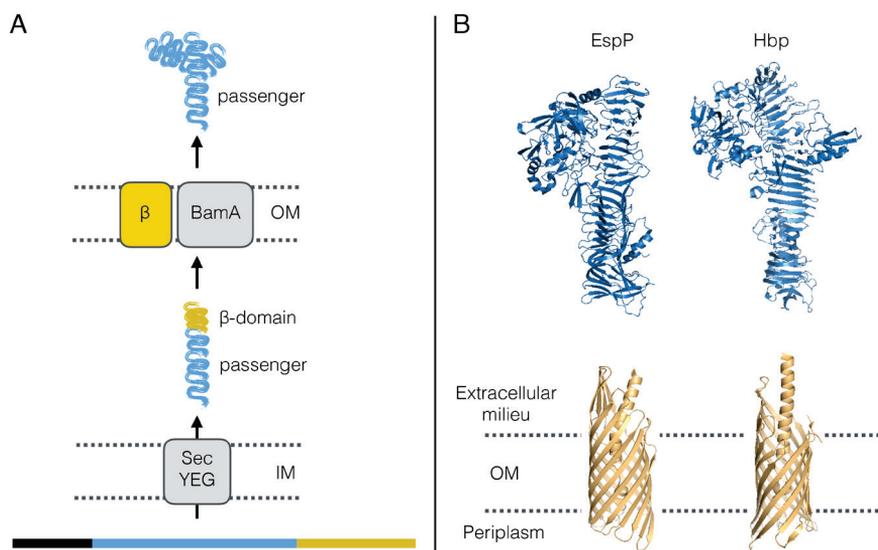
Notably, the periplasm does not contain any ATP and there is no electrochemical gradient across the OM that could energize the AT-mediated translocation process. A greatly supported model suggests that sequential folding of the passenger domain could drive the translocation through a ‘Brownian ratchet mechanism’ (417–420). This model fits well with the  $\beta$ -helical structure of most passenger domains. However, folding of the passenger domain is perhaps not the only driving force but rather supported by other factors such as charge interactions between the passenger and the outer surface of the OM (414).

These mechanistic and structural insights have paved the way for the exploitation of the AT pathway for surface display of heterologous proteins for development of screening tools *e.g.*, (421), biocatalysts *e.g.*, (422), and vaccines *e.g.*, (423). Due to space limitations, I will, despite of the numerous reports of successful use of ATs in biotechnology, in the following section focus only on the use of the *E. coli* AT Hemoglobin protease (Hbp) and its use in antigen display. The type Va AT Hbp was used in paper V of this thesis.

## 15.1 The hemoglobin protease-based display platform

The type Va AT Hbp is a key virulence factor of the human pathogen *E. coli* EB1, which causes peritonitis. Hbp binds to and degrades hemoglobin (424). The catalytic function (the serine protease site) of Hbp is located to one of the five side domains of the  $\beta$ -helical passenger domain (425). Importantly, this domain (D1) is absent in the Hbp display variant used in paper V of this thesis.

Hbp belongs to the Serine Protease Autotransporters of Enterobacteriaceae (SPATE) family. A common characteristic of ATs of this group, apart from the protease function, is the unusually long signal sequence, carrying an approximately 25 amino acid long N-terminal extension (416,426). This highly conserved SPATE specific signal sequence is dispensable for IM targeting and translocation but seems to play a role in preventing premature folding of the AT in the periplasm. Many ATs (including SPATEs) are post-translationally translocated across the IM and it has been observed that the cytoplasmic chaperone DnaK is important for the biogenesis of such ATs, see *e.g.*, (39). Hbp however, uses the SRP dependent pathway to the Sec-translocon (75). Furthermore, YidC has been shown to be involved in the translocation of Hbp across the IM (123).



**Fig. 10.** **A.** Schematic representation of the AT secretion pathway (type Va pathway). **B.** Top panel, in blue: Passenger domains of the SPATEs EspP (PDB 3SZE) and Hbp (PDB 3AK5). Bottom panel, in yellow:  $\beta$ -domains of EspP (PDB 3SLJ) and Hbp (PDB 3AEH).

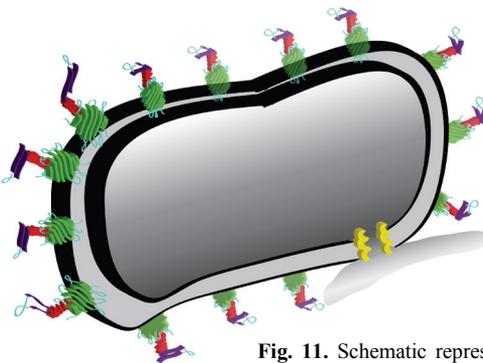
The passenger domain of Hbp (and of all SPATEs) is cleaved and released from the  $\beta$ -barrel upon translocation, but it can be kept attached to the surface by disrupting the cleavage site between the passenger domain and the  $\beta$ -barrel (427). Jong and colleagues took advantage of the available structural knowledge of Hbp and designed a platform suitable for the secretion and display of heterologous antigens of varying size (425,427,428). Key to optimal secretion/display was to maintain the  $\beta$ -helical stem of the passenger intact and rather exchange the passenger side domains with the heterologous antigens of interest. Production of the Hbp chimeras resulted in high-level display of antigens on the surface of both *E. coli* and of the aforementioned attenuated *Salmonella* strain. To highlight the potential of this platform for vaccine development, the gene encoding the chimera was integrated into the *Salmonella* chromosome under the control of the *lacUV5* promoter for stable production. The absence of a *lac*-operon, and thus also of LacI, in *Salmonella* makes production from *lac*-derived promoters constitutive in this bacterium. Although live vaccines are ‘ideal’ based on their feasibility and effectiveness, the potential risk of reversion from an attenuated into a pathogenic state causes regulatory authorities to slow down development. Therefore, other platforms such as outer membrane vesicles (324) or bacterial ghosts may be useful alternatives.

## 16 Bacterial ghosts and lysis protein E

To overcome the potential threat of reversion of the attenuated *Salmonella* vaccine strain to a pathogenic state, the basic concept of autodisplay can be extended by combining the HbpD antigen display platform with the formation of bacterial ghosts (Fig. 11).

Bacterial ghosts (BGs) are empty cell envelopes from Gram-negative bacteria, which are produced by the controlled expression of the cloned gene *E* from bacteriophage  $\phi$ X174, see reviews (429,430) and references therein. This gene product forms a lysis pore across the envelope of the living bacteria, which enables the cytoplasmic content to be expelled through the tunnel. Thus, BGs are devoid of cytoplasmic content but still possess all bacterial bio-adhesive surface properties, important for stimulation of the immune system, in their original state. Therefore, just like for live vaccines, the addition of adjuvants is not required. Furthermore, ghosts do not possess any potential for reverting back to an infectious state. Lysis protein E inhibits the function of MraY, which is a protein that catalyzes the first membrane-localized step of peptidoglycan synthesis. The mechanism of how lysis protein E generates defined pores through the cell envelope is however yet elusive (429–432).

BGs derived from different Gram-negative bacteria have shown potential as vaccine vectors and as delivery vehicles of both protein- and DNA-based drugs (429,430). For recent advancements see *e.g.*, (433–437). To increase the safety of the platform two main approaches have been employed: the co-production of nucleases (438,439) and the use of chemical inactivating agents (429). It has been shown that a lyophilization step renders BGs stable for long time storage (429). Effort has also been aimed at generating antibiotic free production of BGs (440).



**Fig. 11.** Schematic representation of an *E. coli/Salmonella*-derived ghost displaying an HbpD-antigen chimera on its surface.

# Summaries of papers I-V and Outlook

## **Paper I:** Optimizing membrane protein overexpression in the *Escherichia coli* strain Lemo21(DE3)

### **Background**

In the previously developed BL21(DE3)-derived membrane protein production strain Lemo21(DE3) expression of the gene encoding T7 lysozyme, which is a natural inhibitor of T7 RNA, is under control of the *E. coli* rhamnose promoter system. This setup makes that the expression rate of the gene encoding any given membrane protein can be precisely modulated by adding varying amounts of rhamnose to the culture medium (section 13.2.3.1 and Fig. 9). This makes Lemo21(DE3) suitable for the production of membrane proteins. Here, we have studied in detail how the production yields of membrane proteins are optimized in Lemo21(DE3). We also investigated the effects of altering cultivation strategies on production yields to explore the potential of Lemo21(DE3) as a tool for membrane protein production.

### **Results & Conclusions**

First, as expected it could be determined that T7 lysozyme accumulation levels correlated in a manner corresponding to the amount of rhamnose added to the culture medium. At the rhamnose concentration giving highest yields of the model proteins YidC-GFP and GltP-GFP, toxic effects caused by the production of these proteins were not detectable. Also, a larger fraction of total protein produced was inserted into the IM as the rhamnose concentration was increased. At the optimal rhamnose concentration almost no YidC-GFP or GltP-GFP accumulated in inclusion bodies in the cytoplasm. When these two model proteins were produced at a rate higher than optimal, the cells ‘evaded’ membrane protein production, as demonstrated by the appearance of non-producing sub-populations of cells. Notably, the cells that had evaded IMP-GFP production had

also recovered from the toxic burden conferred by the overproduction of these IMP-GFP fusions. In contrast, cells in cultures producing these IMP-GFP fusions in the presence of the optimal rhamnose concentration (or higher) were homogeneous (*i.e.* all cells produced the IMP-GFP fusion to equal levels). Furthermore, our results show that increasing the temperature from 30 °C to 37 °C drastically lowered YidC-GFP levels. When lowering the temperature to 20 °C, yields could be marginally improved if the induction time was increased from 8 to 24 hrs. As expected, the amount of rhamnose needed to reach optimal production yields increased with increased temperature. Finally, increasing production yields by optimizing production rate also led to more active IMP material as shown for the glutamate transporter GltP. The IMPs NhaA and MhpI produced in Lemo21(DE3) were suitable for crystallization upon recovery from the GFP-fusion.

## **Paper II: MemStar: A one-shot *Escherichia coli*-based approach for high-level bacterial membrane protein production**

### **Background**

The improved yields obtained in Lemo21(DE3) are mainly due to increased biomass formation. For structural studies this is posing a problem since detergent solubilization efficiencies decrease as the amount of total membranes to be solubilized increase. Therefore, we set up a screening strategy aiming to find a culturing condition in which membrane protein production is increased per cell rather than per volume of culture. This would make it feasible to work with a larger selection of protein homologues, increasing the rate at which membrane protein structures can be solved.

### **Results & Conclusions**

A set of ten *E. coli* membrane proteins was produced both in the BL21(DE3)-derived strain C43(DE3) and in Lemo21(DE3) (see sections 13.2.2.1 and 13.2.3.1), and production was monitored by the previously described GFP-pipeline (see section 13.3.2.1). The two strains were cultured in four different kinds of both rich and defined media: Lysogeny Broth (LB medium), Terrific Broth (TB medium) and the two autoinduction media ZYM-5052 and PASM-

5052. Exploring the effect of the inducer IPTG further extended the number of expression regimes tested. Surprisingly and somewhat contradictory, the addition of IPTG to Lemo21(DE3) cells cultured in PASM-5052 dramatically boosted production yields per OD<sub>600</sub> unit for all ten proteins. To avoid the necessity of performing rhamnose titration screens, we further determined the rhamnose concentration at which the probability of producing high amounts of membrane proteins was highest. This led to establishing the Membrane protein Single shot amplification recipe: MemStar. Using MemStar, production yields for another set of 24 membrane proteins were improved to reach an average of 5 mg/L per OD<sub>600</sub> unit, which is significantly higher than yields obtained with other common production strategies. MemStar enabled to obtain new structural information for several transporters, including the sodium/proton antiporter NapA.

### **Paper III: High-level production of membrane proteins in *E. coli* BL21(DE3) by omitting the inducer IPTG**

#### **Background**

The BL21(DE3)/pT7 expression system in combination with LB medium is a setup that is routinely used for the production of membrane and secretory proteins (see section 13.2.1). However, from previous studies we know that production of membrane proteins in the cytoplasmic membrane and secretory proteins in the periplasm is often hampered in this setup due to saturation of the capacity of the Sec-translocon, resulting in toxicity to the host and consequently in low production yields. The isolation of the Walker strains and the development of Lemo21(DE3) represents examples of strategies of how this toxicity can be circumvented. We know that the improved production in these strains is due to a lowered intensity of T7RNAP-mediated transcription of the target protein. The BL21(DE3)/pT7-based protein production system is known to be leaky. Leakiness is generally an unwanted characteristic, especially for the production of toxic targets such as membrane and secretory proteins. Here, however, contrary to expectation we observed that omission of IPTG from BL21(DE3)/pT7 cells cultured in LB medium led to significantly higher membrane and secretory protein production yields than when IPTG was added.

## Results & Conclusions

Using a set of cytoplasmic membrane and secretory model proteins, we studied how protein production accumulation rates and yields were affected when omitting IPTG to cultures of BL21(DE3)/pT7 cells. We also explored how these altered target gene expression intensities affected the quality of the recombinant membrane and secretory proteins. Furthermore, flow cytometry was used to study the homogeneity of cell cultures as well as protein accumulation levels per cell. In the complete absence of IPTG, cultures stably produced membrane proteins in the cytoplasmic membrane, whereas upon the addition of IPTG membrane proteins aggregated in the cytoplasm and non-producing clones were selected for. Furthermore, in the absence of IPTG, membrane proteins were produced at a lower rate than in the presence of IPTG. These observations indicate that in the absence of IPTG the Sec-translocon capacity is not saturated, leading to enhanced membrane protein production yields in the cytoplasmic membrane. Importantly, for more than half of the targets tested the yields obtained using uninduced BL21(DE3) cells were higher than the yields obtained in the widely used membrane protein production strains C41(DE3) and C43(DE3). Since most secretory proteins reach the periplasm *via* the Sec-translocon, we also monitored the production of three secretory recombinant proteins in the periplasm of BL21(DE3) cells in the presence and absence of IPTG. For all three targets tested omitting IPTG led to the highest production levels in the periplasm. Conclusively, omission of IPTG from BL21(DE3) cells cultured in LB medium provides a very cost and time effective alternative for the production of membrane- and secretory proteins. Therefore, it is recommended to incorporate this condition in protein production screens.

## **Paper IV: Engineering *E. coli* for rhamnose P<sub>BAD</sub> promoter-based production of membrane and secretory proteins**

### **Background**

Key to the successful production of membrane and secretory proteins in *E. coli* is to be able to precisely set the protein production rates of these proteins. We have previously shown, in paper I, that this can be done by using the Lemo21(DE3) strain in combination with pT7-based protein production vectors. As described in the summary of paper I and in section 13.2.3.1, key to the tunable protein production characteristics of Lemo21(DE3) is the titratable *PrhaBAD*-governed production of lysozyme. Thus, we reasoned that a simpler alternative could be to directly produce the target protein in a *PrhaBAD* governed manner. However, when producing membrane and secretory proteins from the rhamnose inducible P<sub>BAD</sub> in *E. coli*, we were not able to optimize production in a satisfactory manner. Although this promoter system has previously been used for the production of membrane and secretory proteins in *E. coli*, the *modus operandi* of this system has not yet been studied in much detail. We reasoned that studying protein production kinetics of the system could help to improve it.

### **Results & Conclusions**

To characterize *PrhaBAD*-mediated protein production kinetics, we monitored the production of the stable and fast-folding protein SfGFP in *E. coli* wild-type and rhamnose catabolism and RhaT-mediated rhamnose transport deficient single and double mutants. This revealed that in *E. coli* wild-type the tunable protein production characteristic is based on rhamnose consumption rather than on tuning protein production rates. In mutants deficient in RhaT-mediated uptake of rhamnose, protein production rates could be precisely set by varying the amounts of inducer added to the culture medium. Furthermore, in the strain deficient in both RhaT-mediated rhamnose uptake and rhamnose catabolism, protein production rates became more stable over time and only minimal amounts of rhamnose were required. These characteristics appeared to be critical for maximizing *PrhaBAD*-mediated membrane and secretory protein production yields.

## **Paper V: Autotransporter-Based Antigen Display in Bacterial Ghosts**

### **Background**

In this study we simultaneously produced the bacteriophage  $\phi$ X174 lysis protein E and the engineered OM antigen-display platform *E. coli* autotransporter hemoglobin protease (HbpD). The co-production of these two, non-related, membrane proteins can be used for the generation of a multivalent vaccine platform that is most likely safer than live vaccines. The production of lysis protein E generates defined channels in the cell envelope, which leads to the formation of so-called ghosts. In ghosts, the cytoplasmic content is released into the extracellular milieu whilst the morphology of the cell envelope remains mostly intact. The antigen-display platform HbpD has been shown to enable the efficient presentation of antigens on the surface of both *E. coli* and *Salmonella* cells. Here, we used the *Mycobacterium tuberculosis* vaccine target ESAT6 (early secreted antigenic target of 6 kDa) as a model antigen for surface display, fused to the protruding  $\beta$ -helix of HbpD generating the chimera protein HbpD-ESAT6.

### **Results & Conclusions**

The use of different promoter systems/expression regimes enabled the concerted production of HbpD-ESAT6 and lysis protein E in both *E. coli* and *Salmonella*. Ghost formation was monitored by determining lysis efficiency based on determining CFUs, the localization of a set of cellular markers, fluorescence microscopy, flow cytometry, and electron microscopy. HbpD-mediated surface display of ESAT6 was monitored using a combination of a protease accessibility assay, fluorescence microscopy, flow cytometry and (immuno-) electron microscopy. We have shown that the concerted production of HbpD-ESAT6 and lysis protein E in *E. coli* and *Salmonella* can be used to produce ghosts that efficiently display antigens on their surface. This system holds promise for the development of safe and cost-effective vaccines with optimal intrinsic adjuvant activity and effective exposure of antigens to the immune system.

## Outlook

Papers I-IV, as summarized above, all indicate that being able to precisely and stably set protein production rates is key to optimizing membrane and secretory protein production yields in *E. coli*. These results are all based on host organism optimization exercises. With this in mind, it would be interesting to look into whether also other reported strategies, such as modifying translation initiation regions or using N-terminal tags, result in a more controlled protein production regime, thereby improving yields. The results in paper IV were rather unexpected. It would be interesting to compare the expression regime of the *rhaBAD* promoter to that of other sugar-utilizing systems, such as the *araBAD* promoter, in appropriate strain backgrounds. To understand the expression regime from the *rhaBAD* promoter in more detail, experiments that can reveal whether the non-tunable characteristic is generated at a transcriptional or translational level are required. In paper III, it cannot be excluded that trace amounts of lactose are responsible for the results obtained when no IPTG is added. Therefore, a comprehensive study testing different types of media (+/- lactose) in combination with a lactose permease deficient strain would probably further our understanding behind the mechanism of “un-induced” protein production. Lemo21(DE3) is a strain that has been widely and successfully used by academic laboratories and is considered a well-performing protein production strain. However, there is still room for improvement, *e.g.* a combination of Lemo21(DE3) with the rhamnose catabolism and uptake deficient strain developed in paper IV may be explored. As shown in paper II, Lemo21(DE3) can successfully be combined with other media than LB. The MemStar condition has proven extremely robust, and presumably, since Lemo21(DE3) offers the possibility to cover a very large window of target gene expression intensities, only modest results would be expected from modifying MemStar further. As to paper V, ongoing studies aim at further improving production yields of other more challenging (*i.e.* complex) HbpD-chimeras. Initially, the effect of deleting periplasmic and outer membrane proteases on HbpD-chimera production will be studied. Also, the production of these chimeras in the ‘rhamnose optimized protein production strain (paper IV)’ may provide a better starting point for extending the number and size of antigens attached to the Hbp passenger domain. Finally, it would be exciting to study the efficacy of a ghost-based HbpD-mediated vaccine platform in animal models.

# Populärvetenskaplig sammanfattning

Alla levande organismer består av celler. Beroende på vilken organism cellen tillhör ser den lite olika ut. Bakterier är encelliga organismer och deras uppbyggnad och biologiska processer är lite enklare jämfört med celler från till exempel människa. Trots det finns det även mycket som fungerar på samma sätt i dessa avlägset besläktade organismer vilket gör att man kan använda sig av bakterier för att studera komponenter eller processer som även är relevanta för mänskliga celler.

I alla celler finns DNA-molekyler (kromosomer) som innehåller gener som kodar för hur cellen ska bygga ihop olika byggstenar (amino-syror) till proteiner. Gener är uppbyggda av fyra olika molekyler (Adenin, Tymin, Guanin, Cytosin) som bildar en lång ”kedja” av olika kombinationer av dessa fyra molekyler. Beroende på sekvensen av dessa molekyler i genen, kommer proteinet som genen kodar för få olika egenskaper och se olika ut. Denna process delas av alla organismer, oberoende av hur komplexa eller enkla de är. Proteiner är cellens maskiner och kan uträtta en otalig mängd olika aktiviteter. De proteiner som jag studerat sitter vid ytan av cellen, i det så kallade membranet som omsluter den inre delen av cellen. Cellmembraner är uppbyggda av lipidmolekyler (fetter). De proteiner som sitter i membranet måste därför vara fettlösliga, hydrofoba, till skillnad från de proteiner som finns i den inre, vattenlösliga, delen av cellen. Membranproteiner är ofta involverade i viktiga processer, som till exempel att ta upp näring, att exportera ut sådant som är giftigt för cellen eller att reagera på stimuli från sin omgivning. Ett felaktigt membranprotein kan därför ge upphov till olika sjukdomar och faktum är att de flesta mediciner är riktade mot just proteiner som sitter i cellmembraner. För att till exempel kunna utveckla mediciner måste man först ta reda på exakt hur proteinerna fungerar och hur de ser ut. Dock gör membranproteiners hydrofoba egenskap att det är svårt att studera dem. Ibland kan det hjälpa om man har tillgång till stora mängder av proteinet ifråga eller om man kan manipulera det så att de blir lättare att hantera. För att kunna producera proteiner i stora mängder och för att kunna manipulera

dem använder man sig av rekombinant protein produktion. Det innebär att man kan syntetisera en gen som kodar för vilket protein som helst och få till exempel bakterier att ta upp genen och sedan producera det önskade proteinet som genen kodar för. Denna teknik gör det till exempel möjligt att producera insulin med hjälp av bakterier istället för att, som innan, rena upp det från djur.

Min avhandling handlar till största del om hur man kan optimera rekombinant produktion av membranproteiner med hjälp av bakterien *E. coli*. Jag har även tittat på hur produktion av en annan typ av proteiner kan optimeras. Dessa proteiner sitter inte i membranet utan transporteras ut genom membranet så att de hamnar i ett vattenlösligt hölje, den så kallade periplasman, som finns mellan de två membranerna som omger *E. coli* bakterier. Det är vanligt att man producerar industriellt/medicinskt viktiga proteiner som till exempel antikroppar och hormoner i den delen av bakterien.

Den främsta slutsatsen i min avhandling är att det är viktigt att göra det möjligt för bakterien att tillverka dessa typer av proteiner under en låg hastighet. Eftersom alla proteiner är olika har det visat sig att det är allra bäst om man har en bakterie i vilken man kan optimera tillverkningshastigheten för varje enskilt protein. Om proteintillverkningshastigheten är för hög är det lätt att det maskineri som är involverat i att leverera proteinerna till sin korrekta plats, i membranet eller i periplasman, i sin korrekta form blir överlastat. Det leder till att proteinerna blir inaktiva samt att cellen får svårigheter att fungera och växa som den ska. Denna insikt har lett till att vi utvecklat dels nya protokoll för proteinproduktion i *E. coli* men också till utveckling av en ny mutant *E. coli* stam som lämpar sig bättre för produktion av dessa proteiner.

I den sista studien i min avhandling har jag producerat membranproteiner i ett annat syfte än att rena upp dem för att studera dess funktion och form. Här har jag producerat två membranproteiner samtidigt för att utveckla en ny typ av bakterie-baserat vaccin. För detta använde jag bakterien *Salmonella*, vilken är en bakterie som visat sig vara lämpad för vaccinutveckling. Det ena proteinet som jag producerade i *Salmonella* kan användas som en plattform för att ”visa upp” antigener på ytan av bakterien, och det andra produceras för att lysa bakterien genom att skapa definierade hål genom membranet. På så sätt oskadliggör man bakterien, men behåller samtidigt strukturen av membranet innehållande det antigen man valt att klä bakterieytan med. Tanken är att detta system kan ha färre bieffekter i jämförelse med de vacciner som baserar sig på levande bakterier, men ändå vara lika effektivt.

# Acknowledgements

**Jan-Willem**, thanks so much for these five and a half (!!!) years and for everything you have taught me. For all discussions, all the agreeing and disagreeing. The Sec-translocon on the cover is for you...! ☺ Thanks also for all the rounds of proofreading my thesis!! I liked the comparison of it to a kladdkaka...! ☺

**Gunnar von Heijne**, thanks for being my co-supervisor!

**Stefan Nordlund**, thanks for everything you do for DBB and for taking me through all those checkpoints!

## **All the people at DBB:**

Writing up my thesis made me truly realize how much excellent research that is coming out from this department. Thanks for all the inspiration! Special greetings to the groups of **Gunnar von Heijne, Dan Daley, Rob Daniels, Elzbieta Glazer and Ing-Marie Nilsson**.

**David Drew** and **Chiara**, thanks for having me on-board on the Memstar project!

**Nurzian** and **David** and the rest of the **Xbrane crew**, thanks for cloning expertise and inspiration! ☺

**Joel Luirink** and the **Amsterdam/Abera crew**, thanks for nice collaborations!

**Rickard**, thanks for running into our office so often and sharing “worries” and happy moments! I’m so happy we did the Krakow trip together.

**Patricia**, you are the coolest person! So much fun with you as my ‘partner’ in Okinawa! Thanks for making me feel at home here at DBB!

**Bill**, thanks for colouring the vaccine project ☺. I’m happy you are finding use for it now over there, on the far away tropical paradise island as well.... Thanks of course also for having Patricia and me over, it was an amazing week!

**Nina**, always so positive! Thanks for sharing your energy!

**Susan**, thanks for everything. For paving the way and for making a really good and inspiring thesis. ☺

**Eduarne**, I miss you so so much! Without you I would probably not have survived the first year <3

### **My closest lab-mates:**

**Thomas**, you have an impressive ability to be positive in every situation, thanks for sharing that! Thanks also for taking such good care of the lab and of the group. **Zhe**, thanks for all the work on the non-IPTG project, we did it! ☺ **Nienke**, I'm so happy you joined the group, thanks for all the warming support! **Alex**, welcome and good luck with those Hbp-chimeras! ☺

### **Family and Friends:**

**LTH-gänget (8xEmma)**, ni är bäst! Tack för all oändlig pepp. Tack för pepp också till mina fina Stockholmstjejer, **Emma, Nina, Mickan, Jennie o Elin** (tack för korrekturläsningen!). **Line**, du är en evig inspirationskälla, så coooool. **Giovanna!!!** Jag är så himla glad att vi fann varandra bland kalla joggingturer i den varmaste delen av Stockholm. Utan dig hade dessa år varit betydligt tuffare. Min lillstrumpa **Therese**, tack för att du alltid alltid finns där. Och för att du är min lillasyster på låtsas <3.

My second family: **Guido, Enza, Angelo e Betti!** Grazie mille per il caloroso appoggio!! Betti, thanks so much also for formatting assistance!! ☺

Mina storebröder **Thomas** och **Stefan**. Tack för er finfina uppfostran, bättre förebilder är omöjligt att hitta! Bästa och coolaste svägerskorna **Anna-Karin** och **Lotta**, tack för en massa saker men allra mest för att ni skapat de knasigaste och klokaste kidsen i stan. **Astrid, Thea, Alice** och **Alfred**. Så mycket kul med er!!

**Farmor, Mormor** och **Morfar**, tack för allt vett ni lärt mig. Jag saknar er så mycket.

**Mamma** och **Pappa**, ni är de bästa föräldrarna i hela världen. Tack för allt ni gör för mig, jämt.

**Alberto**, tack för ditt oändliga tålamod och för din förmåga att alltid förstå. Tack för allt du lär mig och för allt kul vi gör ihop! "Chi va piano va sano va lontano" ☺ <3 ☺ <3 ☺ <3 ☺ !!!

# References

1. Silhavy TJ, Kahne D, Walker S. The bacterial cell envelope. *Cold Spring Harb Perspect Biol.* 2010;2(5):a000414.
2. Tanford C. The hydrophobic effect and the organization of living matter. *Science.* 1978;200(4345):1012–8.
3. Raetz CRH, Dowhan W. Biosynthesis and function of phospholipids in *Escherichia coli*. *J Biol Chem.* 1990;265(3):1235–8.
4. Aibara S, Kato M, Ishiniga M, Kito M. Changes in positional distribution of fatty acids in the phospholipids of *Escherichia coli* after shift-down in temperature. *Biochim Biophys Acta.* 1972;270(3):301–6.
5. Zhang YM, Rock CO. Membrane lipid homeostasis in bacteria. *Nat Rev Microbiol.* 2008;6(3):222–33.
6. Klein P, Kanehisa M, DeLisi C. The detection and classification of membrane-spanning proteins. *Biochim Biophys Acta.* 1985;815(3):468–76.
7. Fairman JW, Noinaj N, Buchanan SK. The structural biology of  $\beta$ -barrel membrane proteins: a summary of recent reports. *Curr Opin Struct Biol.* 2011;21(4):523–31.
8. Singer SJ, Nicolson GL. The fluid mosaic model of the structure of cell membranes. *Science.* 1972;175(23):720–31.
9. Engelman DM. Membranes are more mosaic than fluid. *Nature.* 2005;438(7068):578–80.
10. Barák I, Muchová K. The role of lipid domains in bacterial cell processes. *Int J Mol Sci.* 2013;14(2):4050–65.
11. Goñi FM. The basic structure and dynamics of cell membranes: An update of the Singer–Nicolson model. *Biochim Biophys Acta.* 2014;1838(6):1467–76.
12. Hobot JA, Carlemalm E, Villiger W, Kellenberger E. Periplasmic gel: new concept resulting from the reinvestigation of bacterial cell envelope ultrastructure by new methods. *J Bacteriol.* 1984;160(1):143–52.
13. Mullineaux CW, Nenninger A, Ray N, Robinson C. Diffusion of green fluorescent protein in three cell environments in *Escherichia coli*. *J Bacteriol.* 2006;188(10):3442–8.
14. Vollmer W, Blanot D, De Pedro M. Peptidoglycan structure and architecture. *FEMS Microbiol Rev.* 2008;32(2):149–67.
15. Koebnik R, Locher KP, Van P. MicroReview Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol Microbiol.* 2000;37(2):239–53
16. Hancock RE. Role of porins in outer membrane permeability. *J Bacteriol.* 1987;169(3):929–33.
17. Raetz CRH, Whitfield C. Lipopolysaccharide endotoxins. *Annu Rev Biochem.* 2002;71(1):635–700.
18. Crick F. Central Dogma Of Molecular Biology. *Nature.* 1970;227:561–3.
19. Rose GD, Roy S. Hydrophobic basis of packing in globular proteins. *Proc Natl Acad Sci U S A.* 1980;77(8):4643–7.
20. Arboř A. Membrane proteins. *Curr Opin Struct Biol.* 1996;6(4):560–72.

21. von Heijne G. The membrane protein universe: what's out there and why bother? *J Intern Med.* 2007;261(6):543–57.
22. Travers AA, Burgess RR. Cyclic re-use of the RNA polymerase sigma factor. *Nature.* 1969;222(5193):537–40.
23. Chandransu P, Helmann JD. Sigma Factors in Gene Expression. *Encycl Life Sci.* 2014.
24. French SL, Miller OL. Transcription mapping of the *Escherichia coli* chromosome by electron microscopy. *J Bacteriol.* 1989;171(8):4207–16.
25. Epshtein V, Nudler E. Cooperation between RNA polymerase molecules in transcription elongation. *Science.* 2003;300(5620):801–5.
26. Steitz TA. A structural understanding of the dynamic ribosome machine. *Nat Rev Mol Cell Biol.* 2008;9(3):242–53.
27. Vellanoweth RL, Rabinowitz JC. The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli* in vivo. *Mol Microbiol.* 1992;6(9):1105–14.
28. de Smit MH, van Duin J. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc Natl Acad Sci U S A.* 1990;87(19):7668–72.
29. Brandt F, Etchells SA, Ortiz JO, Elcock AH, Hartl FU, Baumeister W. The native 3D organization of bacterial polysomes. *Cell.* 2009;136(2):261–71.
30. Jahn TR, Radford SE. The Yin and Yang of protein folding. *FEBS J.* 2005;272(23):5962–70.
31. Hartl FU, Hayer-Hartl M. Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol.* 2009;16(6):574–81.
32. Bartlett AI, Radford SE. An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nat Struct Mol Biol.* 2009;16(6):582–9.
33. Ellis RJ, Minton AP. Protein aggregation in crowded environments. *Biol Chem.* 2006;387(5):485–95.
34. Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Ulrich Hartl F. Molecular chaperone functions in protein folding and proteostasis. *Annu Rev Biochem.* 2013;82:323–55.
35. Dowhan W, Bogdanov M. Molecular genetic and biochemical approaches for defining lipid-dependent membrane protein folding. *Biochim Biophys Acta.* 2012;1818(4):1097–107.
36. Kim HK, Choi S Il, Seong BL. 5S rRNA-assisted DnaK refolding. *Biochem Biophys Res Commun.* 2010;391(2):1177–81.
37. Bogdanov M, Dowhan W. Lipid-assisted protein folding. *J Biol Chem.* 1999;274(52):36827–30.
38. Calloni G, Chen T, Schermann SM, Chang H, Genevoux P, Agostini F, et al. DnaK functions as a central hub in the *E. coli* chaperone network. *Cell Rep.* 2012;1(3):251–64.
39. Castanié-Cornet MP, Bruel N, Genevoux P. Chaperone networking facilitates protein targeting to the bacterial cytoplasmic membrane. *Biochim Biophys Acta.* 2014;1843(8):1442–56.
40. Merz F, Hoffmann A, Rutkowska A, Zachmann-Brand B, Bukau B, Deuerling E. The C-terminal domain of *Escherichia coli* trigger factor represents the central module of its chaperone activity. *J Biol Chem.* 2006;281(42):31963–71.

41. Hebert DN, Chandrasekhar KD, Gierasch LM. You got to know when to hold (or unfold) 'em. *Mol Cell*. 2012;48(1):3–4.
42. Hoffmann A, Becker AH, Zachmann-Brand B, Deuerling E, Bukau B, Kramer G. Concerted action of the ribosome and the associated chaperone trigger factor confines nascent polypeptide folding. *Mol Cell*. 2012;48(1):63–74.
43. Patzelt H, Rüdiger S, Brehmer D, Kramer G, Vorderwülbecke S, Schaffitzel E, et al. Binding specificity of *Escherichia coli* trigger factor. *Proc Natl Acad Sci U S A*. 2001;98(25):14244–9.
44. Oh E, Becker AH, Sandikci A, Huber D, Chaba R, Gloge F, et al. Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell*. 2011;147(6):1295–308.
45. Bornemann T, Holtkamp W, Wintermeyer W. Interplay between trigger factor and other protein biogenesis factors on the ribosome. *Nat Commun*. 2014;5(4180).
46. Lin K-F, Sun C-S, Huang Y-C, Chan SI, Koubek J, Wu T-H, et al. Cotranslational protein folding within the ribosome tunnel influences trigger-factor recruitment. *Biophys J*. 2012;102(12):2818–27.
47. Liu CP, Perrett S, Zhou JM. Dimeric trigger factor stably binds folding-competent intermediates and cooperates with the DnaK-DnaJ-GrpE chaperone system to allow refolding. *J Biol Chem*. 2005;280(14):13315–20.
48. Martinez-Hackert E, Hendrickson W. Promiscuous substrate recognition in folding and assembly activities of the trigger factor chaperone. *Cell*. 2009;138(5):923–34.
49. Kelley WL. The J-domain family and the recruitment of chaperone power. *Trends Biochem Sci*. 1998;23(6):222–7.
50. Rüdiger S, Schneider-Mergener J, Bukau B. Its substrate specificity characterizes the DnaJ co-chaperone as a scanning factor for the DnaK chaperone. *EMBO J*. 2001;20(5):1042–50.
51. Deuerling E, Schulze-Specking A, Tomoyasu T, Mogk A, Bukau B. Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature*. 1999;400(6745):693–6.
52. Teter SA, Houry WA, Ang D, Tradler T, Rockabrand D, Fischer G, et al. Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. *Cell*. 1999;97(6):755–65.
53. Azia A, Unger R, Horovitz A. What distinguishes GroEL substrates from other *Escherichia coli* proteins? *FEBS J*. 2012;279(4):543–50.
54. Hartl FU. Protein folding. Secrets of a double-doughnut. *Nature*. 1994;371(6498):557–9.
55. Braig K, Otwinowski Z, Hegde R, Boisvert DC, Joachimiak A, Horwich AL, et al. The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature*. 1994;371(6498):578–86.
56. Xu Z, Horwich AL, Sigler PB. The crystal structure of the asymmetric GroEL-GroES-(ADP)<sub>7</sub> chaperonin complex. *Nature*. 1997;388(6644):741–50.
57. Sharma S, Chakraborty K, Müller BK, Astola N, Tang Y-C, Lamb DC, et al. Monitoring protein conformation along the pathway of chaperonin-assisted folding. *Cell*. 2008;133(1):142–53.
58. Skjærven L, Cuellar J, Martinez A, Valpuesta JM. Dynamics, flexibility, and allostery in molecular chaperonins. *FEBS Lett*. 2015;589(19):2522–32.
59. Weissman JS, Hohl CM, Kovalenko O, Kashi Y, Chen S, Braig K, et al. Mechanism of GroEL action: Productive release of polypeptide from a sequestered position under groes. *Cell*. 1995;83(4):577–87.

60. Georgescauld F, Popova K, Gupta AJ, Bracher A, Engen JR, Hayer-Hartl M, et al. GroEL/ES chaperonin modulates the mechanism and accelerates the rate of TIM-barrel domain folding. *Cell*. 2014;157(4):922–34.
61. Fayet O, Ziegelhoffer T, Georgopoulos C. The groES and groEL heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. *J Bacteriol*. 1989;171(3):1379–85.
62. Kudva R, Denks K, Kuhn P, Vogt A, Müller M, Koch H-G. Protein translocation across the inner membrane of Gram-negative bacteria: the Sec and Tat dependent protein transport pathways. *Res Microbiol*. 2013;164(6):505–34.
63. Huber D, Boyd D, Xia Y, Olma MH, Gerstein M, Beckwith J. Use of thioredoxin as a reporter to identify a subset of *Escherichia coli* signal sequences that promote signal recognition particle-dependent translocation. *J Bacteriol*. 2005;187(9):2983–91.
64. Luirink J, Yu Z, Wagner S, de Gier J-W. Biogenesis of inner membrane proteins in *Escherichia coli*. *Biochim Biophys Acta*. 2012;1817(6):965–76.
65. Blobel G, Dofferstein B. Transfer of proteins across membranes. II. Reconstitution of functional rough microsomes from heterologous components. *J Cell Biol*. 1975;67(3):852–62.
66. von Heijne G. The Signal peptide. *J Membr Biol*. 1990;115(3):195–201.
67. Lee HC, Bernstein HD. The targeting pathway of *Escherichia coli* presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. *Proc Natl Acad Sci U S A*. 2001;98(6):3471–6.
68. Schierle CF, Berkmen M, Huber D, Kumamoto C, Boyd D, Beckwith J. The DsbA signal sequence directs efficient, cotranslational export of passenger proteins to the *Escherichia coli* periplasm via the signal recognition particle pathway. *J Bacteriol*. 2003;185(19):5706–13.
69. Steiner D, Forrer P, Stump MT, Plückthun A. Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. *Nat Biotechnol*. 2006;24(7):823–31.
70. Mircheva M, Boy D, Weiche B, Hucke F, Graumann P, Koch H-G. Predominant membrane localization is an essential feature of the bacterial signal recognition particle receptor. *BMC Biol*. 2009;7(76).
71. Bibi E. Early targeting events during membrane protein biogenesis in *Escherichia coli*. *Biochim Biophys Acta*. 2011;1808(3):841–50.
72. Holtkamp W, Lee S, Bornemann T, Senyushkina T, Rodnina M V, Wintermeyer W. Dynamic switch of the signal recognition particle from scanning to targeting. *Nat Struct Mol Biol*. 2012;19(12):1332–7.
73. Shan SO, Chandrasekar S, Walter P. Conformational changes in the GTPase modules of the signal reception particle and its receptor drive initiation of protein translocation. *J Cell Biol*. 2007;178(4):611–20.
74. Prilusky J, Bibi E. Studying membrane proteins through the eyes of the genetic code revealed a strong uracil bias in their coding mRNAs. *Proc Natl Acad Sci U S A*. 2009;106(16):6662–6.
75. Sijbrandi R, Urbanus ML, Ten Hagen-Jongman CM, Bernstein HD, Oudega B, Otto BR, et al. Signal recognition particle (SRP)-mediated targeting and Sec-dependent translocation of an extracellular *Escherichia coli* protein. *J Biol Chem*. 2003;278(7):4654–9.
76. Karamyshev AL, Johnson AE. Selective SecA Association with signal sequences in ribosome-bound nascent chains: a potential role for SecA in ribosome targeting to the bacterial membrane. *J Biol Chem*. 2005;280(45):37930–40.

77. Huber D, Rajagopalan N, Preissler S, Rocco MA, Merz F, Kramer G, et al. SecA interacts with ribosomes in order to facilitate posttranslational translocation in bacteria. *Mol Cell*. 2011;41(3):343–53.
78. Chatzi KE, Sardis MF, Economou A, Karamanou S. SecA-mediated targeting and translocation of secretory proteins. *Biochim Biophys Acta*. 2014;1843(8):1466–74.
79. Beckwith J. The Sec-dependent pathway. *Res Microbiol*. 2013;164(6):497–504.
80. Knoblauch NTM, Rüdiger S, Schönfeld HJ, Driessen AJM, Schneider-Mergener J, Bukau B. Substrate specificity of the SecB chaperone. *J Biol Chem*. 1999;274(48):34219–25.
81. Natale P, Brüser T, Driessen JM. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane - distinct translocases and mechanisms. *Biochim Biophys Acta*. 2008;1778(9):1735–56.
82. Driessen AJM, Nouwen N. Protein translocation across the bacterial cytoplasmic membrane. *Annu Rev Biochem*. 2008;77(1):643–67.
83. Berg B Van Den, Jr WMC, Collinson I, Modis Y, Hartmann E, Harrison SC, et al. X-ray structure of a protein-conducting channel. *Nature*. 2004;427(6969):36–44.
84. Spiess M. Protein Translocation: The Sec61/SecYEG translocon caught in the act. *Curr Biol*. 2014;24(8):317–9.
85. Tsukazaki T, Mori H, Fukai S, Ishitani R, Mori T, Dohmae N, et al. Conformational transition of Sec machinery inferred from bacterial SecYE structures. *Nature*. 2008;455(7215):988–91.
86. Egea PF, Stroud RM. Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. *Proc Natl Acad Sci U S A*. 2010;107(40):17182–7.
87. Gogala M, Becker T, Beatrix B, Armache J-P, Barrio-Garcia C, Berninghausen O, et al. Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion. *Nature*. 2014;506(7486):107–10.
88. Park E, Ménétret J-F, Gumbart JC, Ludtke SJ, Li W, Whynot A, et al. Structure of the SecY channel during initiation of protein translocation. *Nature*. 2013;506(7486):102–6.
89. Pohlschr M, Hartmann E, Hand NJ, Dilks K, Haddad A. Diversity and evolution of protein translocation. *Annu Rev Microbiol*. 2005;59:91–111.
90. Frauenfeld J, Gumbart J, Sluis EO Van Der, Funes S, Gartmann M, Beatrix B, et al. Cryo-EM structure of the ribosome-SecYE complex in the membrane environment. *Nat Struct Mol Biol*. 2011;18(5):614–21.
91. Park E, Rapoport TA. Preserving the membrane barrier for small molecules during bacterial protein translocation. *Nature*. 2011;473(7346):239–42.
92. Kuhn P, Weiche B, Sturm L, Sommer E, Drepper F, Warscheid B, et al. The bacterial SRP receptor, SecA and the ribosome use overlapping binding sites on the SecY translocon. *Traffic*. 2011;12(5):563–78.
93. Plath K, Mothes W, Wilkinson BM, Stirling CJ, Rapoport TA. Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. *Cell*. 1998;94(6):795–807.
94. Dalbey RE, Wang P, van Dijl JM. Membrane proteases in the bacterial protein secretion and quality control pathway. *Microbiol Mol Biol Rev*. 2012;76(2):311–30.
95. Shiozuka K, Tani K, Mizushima S, Tokuda H. The proton motive force lowers the level of ATP required for the in vitro translocation of a secretory protein in *Escherichia coli*. *J Biol Chem*. 1990;265(31):18843–7.

96. Ismail N, Hedman R, Lindén M, von Heijne G. Charge-driven dynamics of nascent-chain movement through the SecYEG translocon. *Nat Struct Mol Biol.* 2015;22(2):145–9.
97. Collinson I, Corey RA, Allen WJ, Collinson I. Channel crossing: how are proteins shipped across the bacterial plasma membrane? *Phil. Trans. R. Soc. B.* 2015;370(1679).
98. Zhang B, Miller TF. Hydrophobically stabilized open state for the lateral gate of the Sec translocon. *Proc Natl Acad Sci.* 2010;107(12):5399–404.
99. Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M, et al. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature.* 2007;450(7172):1026–30.
100. Cymer F, Heijne G Von. Cotranslational folding of membrane proteins probed by arrest-peptide – mediated force measurements. *Proc Natl Acad Sci U S A.* 2013;110(36):14640–5.
101. Dalbey RE, Chen M. Sec-translocase mediated membrane protein biogenesis. *Biochim Biophys Acta.* 2004;1694(1-3):37–53.
102. Deitermann S, Sprie GS, Koch H-G. A dual function for SecA in the assembly of single spanning membrane proteins in *Escherichia coli*. *J Biol Chem.* 2005;280(47):39077–85.
103. Sachelaru I, Petriman NA, Kudva R, Koch HG. Dynamic interaction of the Sec translocon with the chaperone PpiD. *J Biol Chem.* 2014;289(31):21706–15.
104. Schulze RJ, Komar J, Botte M, Allen WJ, Whitehouse S, Gold VAM, et al. Membrane protein insertion and proton-motive-force-dependent secretion through the bacterial holo-translocon SecYEG–SecDF–YajC–YidC. *Proc Natl Acad Sci.* 2014;111(13):4844–9.
105. Samuelson JC, Chen M, Jiang F, Möller I, Wiedmann M, Kuhn A, et al. YidC mediates membrane protein insertion in bacteria. *Nature.* 2000;406(6796):637–41.
106. Scotti PA, Urbanus ML, Brunner J, de Gier JW, von Heijne G, van der Does C, et al. YidC, the *Escherichia coli* homologue of mitochondrial Oxa1p, is a component of the Sec translocase. *EMBO J.* 2000;19(4):542–9.
107. Houben ENG, ten Hagen-Jongman CM, Brunner J, Oudega B, Luirink J. The two membrane segments of leader peptidase partition one by one into the lipid bilayer via a Sec/YidC interface. *EMBO Rep.* 2004;5(10):970–5.
108. Sachelaru I, Petriman NA, Kudva R, Kuhn P, Welte T, Knapp B, et al. YidC occupies the lateral gate of the SecYEG translocon and is sequentially displaced by a nascent membrane protein. *J Biol Chem.* 2013;288(23):16295–307.
109. Nagamori S, Smirnova IN, Kaback HR. Role of YidC in folding of polytopic membrane proteins. *J Cell Biol.* 2004;165(1):53–62.
110. Zhu L, Kaback HR, Dalbey RE. YidC protein, a molecular chaperone for LacY protein folding via the SecYEG protein machinery. *J Biol Chem.* 2013;288(39):28180–94.
111. Wagner S, Pop O, Haan GJ, Baars L, Koningstein G, Klepsch MM, et al. Biogenesis of MalF and the MalFGK2 maltose transport complex in *Escherichia coli* requires YidC. *J Biol Chem.* 2008;283(26):17881–90.
112. de Sousa Borges A, de Keyzer J, Driessen AJM, Scheffers D-J. The *Escherichia coli* membrane protein insertase YidC assists in the biogenesis of Penicillin Binding Proteins. *J Bacteriol.* 2015;197(8):2556–14.
113. Gallusser A, Kuhn A. Initial steps in protein membrane insertion. Bacteriophage M13 procoat protein binds to the membrane surface by electrostatic interaction. *EMBO J.* 1990;9(9):2723–9.

114. Facey SJ, Neugebauer SA, Krauss S, Kuhn A. The mechanosensitive channel protein MscL is targeted by the SRP to the novel YidC membrane insertion pathway of *Escherichia coli*. *J Mol Biol*. 2007;365(4):995–1004.
115. Dalbey RE, Kuhn A, Zhu L, Kiefer D. The membrane insertase YidC. *Biochim Biophys Acta*. 2014;1843(8):1489–96.
116. Geng Y, Kedrov A, Caumanns JJ, Crevenna AH, Lamb DC, Beckmann R, et al. Role of the cytosolic loop C2 and the C terminus of YidC in ribosome binding and insertion activity. *J Biol Chem*. 2015;290(28):17250–61.
117. van Bloois E, Jan Haan G, de Gier J-W, Oudega B, Luirink J. F(1)F(0) ATP synthase subunit c is targeted by the SRP to YidC in the *E. coli* inner membrane. *FEBS Lett*. 2004;576(1-2):97–100.
118. Welte T, Kudva R, Kuhn P, Sturm L, Braig D, Muller M, et al. Promiscuous targeting of polytopic membrane proteins to SecYEG or YidC by the *Escherichia coli* signal recognition particle. *Mol Biol Cell*. 2012;23(3):464–79.
119. Robinson PJ, Woolhead C. Post-translational membrane insertion of an endogenous YidC substrate. *Biochim Biophys Acta*. 2013;1833(12):2781–8.
120. Kumazaki K, Kishimoto T, Furukawa A, Mori H, Tanaka Y, Dohmae N, et al. Crystal structure of *Escherichia coli* YidC, a membrane protein chaperone and insertase. *Sci Rep*. 2014;4(7299).
121. Kumazaki K, Chiba S, Takemoto M, Furukawa A, Nishiyama K, Sugano Y, et al. Structural basis of Sec-independent membrane protein insertion by YidC. *Nature*. 2014;509(7501):516–20.
122. Shimokawa-Chiba N, Kumazaki K, Tsukazaki T, Nureki O, Ito K, Chiba S. Hydrophilic microenvironment required for the channel-independent insertase function of YidC protein. *Proc Natl Acad Sci*. 2015;112(16):5063–8.
123. Jong WSP, ten Hagen-Jongman CM, Ruijter E, Orru RVA, Genevaux P, Luirink J. YidC is involved in the biogenesis of the secreted autotransporter hemoglobin protease. *J Biol Chem*. 2010;285(51):39682–90.
124. Wagner S, Baars L, Ytterberg AJ, Klussmeier A, Wagner CS, Nord O, et al. Consequences of membrane protein overexpression in *Escherichia coli*. *Mol Cell Proteomics*. 2007;6(9):1527–50.
125. Nouwen N, Driessen JM. SecDFyajC forms a heterotetrameric complex with YidC. *Mol Microbiol*. 2002;44(5):1397–405.
126. Dalal K, Nguyen N, Alami M, Tan J, Moraes TF, Lee WC, et al. Structure, binding, and activity of Syd, a SecY-interacting protein. *J Biol Chem*. 2009;284(12):7897–902.
127. Yu Z, Lavèn M, Klepsch M, de Gier JW, Bitter W, van Ulsen P, et al. Role for *Escherichia coli* YidD in membrane protein insertion. *J Bacteriol*. 2011;193(19):5242–51.
128. Viklund H, Granseth E, Elofsson A. Structural classification and prediction of reentrant regions in  $\alpha$ -Helical transmembrane proteins: Application to complete genomes. *J Mol Biol*. 2006;361(3):591–603.
129. Nishiyama K, Suzuki T, Tokuda H. Inversion of the membrane topology of SecG coupled with SecA-dependent preprotein translocation. *Cell*. 1996;85(1):71–81.
130. Lehn RC Van, Zhang B, Miller TF, Lehn RC Van, Zhang B, Iii TFM. Regulation of multispanspanning membrane protein topology via post-translational annealing. *eLife*. 2015;4(e08697).

131. von Heijne G. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J.* 1986;5(11):3021–7.
132. Goder V, Spiess M. Molecular mechanism of signal sequence orientation in the endoplasmic reticulum. *EMBO J.* 2003;22(14):3645–53.
133. Wahlberg JM, Spiess M. Multiple determinants direct the orientation of signal-anchor proteins: the topogenic role of the hydrophobic signal domain. *J Cell Biol.* 1997;137(3):555–62.
134. Wallin E, von Heijne G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* 1998;7(4):1029–38.
135. Beltzer JP, Fiedler K, Fuhrer C, Geffen I, Handschin C, Wessels HP, et al. Charged residues are major determinants of the transmembrane orientation of a signal-anchor sequence. *J Biol Chem.* 1991;266(2):973–8.
136. Bogdanov M, Heacock PN, Dowhan W. A polytopic membrane protein displays a reversible topology dependent on membrane lipid composition. *EMBO J.* 2002;21(9):2107–16.
137. Goder V, Bieri C, Spiess M. Glycosylation can influence topogenesis of membrane proteins and reveals dynamic reorientation of nascent polypeptides within the translocon. *J Cell Biol.* 1999;147(2):257–65.
138. Devaraneni PK, Conti B, Matsumura Y, Yang Z, Johnson AE, Skach WR. Stepwise insertion and inversion of a type II signal anchor sequence in the ribosome-Sec61 translocon complex. *Cell.* 2011;146(1):134–47.
139. Shaw AS, Rottier PJ, Rose JK. Evidence for the loop model of signal-sequence insertion into the endoplasmic reticulum. *Proc Natl Acad Sci U S A.* 1988;85(20):7592–6.
140. Cymer F, von Heijne G, White SH. Mechanisms of integral membrane protein insertion and folding. *J Mol Biol.* 2015;427(5):999–1022.
141. Duong F, Wickner W. Sec-dependent membrane protein biogenesis: SecYEG, preprotein hydrophobicity and translocation kinetics control the stop-transfer function. *EMBO J.* 1998;17(3):696–705.
142. Pitonzo D, Skach WR. Molecular mechanisms of aquaporin biogenesis by the endoplasmic reticulum Sec61 translocon. *Biochim Biophys Acta.* 2006;1758(8):976–88.
143. Hedin LE, Öjemalm K, Bernsel A, Hennerdal A, Illergård K, Enquist K, et al. Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. *J Mol Biol.* 2010;396(1):221–9.
144. Ojemalm K, Halling KK, Nilsson I, von Heijne G, Helix HT, Heijne G Von. Orientational preferences of neighboring helices can drive ER insertion of a marginally hydrophobic transmembrane helix. *Mol Cell.* 2012;45(4).
145. Hou B, Lin P-J, Johnson AE. Membrane protein TM segments are retained at the translocon during integration until the nascent chain cues FRET-detected release into bulk lipid. *Mol Cell.* 2012;48(3):398–408.
146. Cross BCS, High S. Dissecting the physiological role of selective transmembrane-segment retention at the ER translocon. *J Cell Sci.* 2009;122(11):1768–77.
147. Ruiz N, Silhavy TJ. Sensing external stress: watchdogs of the *Escherichia coli* cell envelope. *Curr Opin Microbiol.* 2005;8(2):122–6.
148. Shimohata N, Nagamori S, Akiyama Y, Kaback HR, Ito K. SecY alterations that impair membrane protein folding and generate a membrane stress. *J Cell Biol.* 2007;176(3):307–17.

149. Hatahet F, Boyd D, Beckwith J. Disulfide bond formation in prokaryotes: History, diversity and design. *Biochim Biophys Acta*. 2014;1844(8):1402–14.
150. Denoncin K, Schwalm J, Vertommen D, Silhavy TJ, Collet JF. Dissecting the *Escherichia coli* periplasmic chaperone network using differential proteomics. *Proteomics*. 2012;12(9):1391–401.
151. McMorran LM, Brockwell DJ, Radford SE. Mechanistic studies of the biogenesis and folding of outer membrane proteins in vitro and in vivo: What have we learned to date? *Arch Biochem Biophys*. 2014;564:265–80.
152. Goemans C, Denoncin K, Collet J-F. Folding mechanisms of periplasmic proteins. *Biochim Biophys Acta*. 2014;1843(8):1517–28.
153. Quan S, Koldewey P, Tapley T, Kirsch N, Ruane KM, Pfizenmaier J, et al. Genetic selection designed to stabilize proteins uncovers a chaperone called Spy. *Nat Struct Mol Biol*. 2011;18(3):262–9.
154. Schwalm J, Mahoney TF, Soltes GR, Silhavy TJ. Role for Skp in LptD Assembly in *Escherichia coli*. *J Bacteriol*. 2013;195(16):3734–42.
155. Antonoaea R, Fürst M, Nishiyama K-I, Müller M. The periplasmic chaperone PpiD interacts with secretory proteins exiting from the SecYEG translocon. *Biochemistry*. 2008;47(20):5649–56.
156. Matern Y, Barion B, Behrens-Kneip S. PpiD is a player in the network of periplasmic chaperones in *Escherichia coli*. *BMC Microbiol*. 2010;10(251).
157. Gotzke H, Palombo I, Muheim C, Perrody E, Genevaux P, Kudva R, et al. YfgM is an ancillary subunit of the SecYEG translocon in *Escherichia coli*. *J Biol Chem*. 2014;289(27):19089–97.
158. Götzke H, Muheim C, Altelaar AFM, Heck AJR, Maddalo G, Daley DO. Identification of putative substrates for the periplasmic chaperone YfgM in *Escherichia coli* using quantitative proteomics. *Mol Cell Proteomics*. 2015;14(1):216–26.
159. Bittner L-M, Westphal K, Narberhaus F. Conditional proteolysis of the membrane protein YfgM by the FtsH protease depends on a novel N-terminal degron. *J Biol Chem*. 2015;290(31):19367–78.
160. Jarchow S, Lück C, Görg A, Skerra A. Identification of potential substrate proteins for the periplasmic *Escherichia coli* chaperone Skp. *Proteomics*. 2008;8(23-24):4987–94.
161. Entzminger KC, Chang C, Myhre RO, McCallum KC, Maynard J. The Skp chaperone helps fold soluble proteins in vitro by inhibiting aggregation. *Biochemistry*. 2012;51(24):4822–34.
162. Saul FA, Arié J-P, Vulliez-le Normand B, Kahn R, Betton J-M, Bentley GA. Structural and Functional Studies of FkpA from *Escherichia coli*, a cis/trans peptidyl-prolyl isomerase with chaperone activity. *J Mol Biol*. 2004;335(2):595–608.
163. Ruiz-Perez F, Henderson IR, Nataro JP. Interaction of FkpA, a peptidyl-prolyl cis/trans isomerase with EspP autotransporter protein. *Gut Microbes*. 2010;1(5):339–44.
164. Ge X, Wang R, Ma J, Liu Y, Ezemaduka AN, Chen PR, et al. DegP primarily functions as a protease for the biogenesis of  $\beta$ -barrel outer membrane proteins in the Gram-negative bacterium *Escherichia coli*. *FEBS J*. 2014;281(4):1226–40.
165. Nakayama H, Kurokawa K, Lee BL. Lipoproteins in bacteria: structures and biosynthetic pathways. *FEBS J*. 2012;279(23):4247–68.
166. Vertommen D, Ruiz N, Leverrier P, Silhavy TJ, Collet JF. Characterization of the role of the *Escherichia coli* periplasmic chaperone SurA using differential proteomics. *Proteomics*. 2009;9(9):2432–43.

167. Rollauer SE, Soorshjani MA, Noinaj N, Buchanan SK, Buchanan SK. Outer membrane protein biogenesis in Gram-negative bacteria. *Phil. Trans. Soc. B.* 2015;370(1679).
168. Rizzitello A, Harper JR, Silhavy TJ. Genetic evidence for parallel pathways of chaperone activity in the periplasm of *Escherichia coli*. *J Bacteriol.* 2001;183(23):6794–800.
169. Robert V, Volokhina EB, Senf F, Bos MP, Gelder P Van, Tommassen J. Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. *PLoS Biol.* 2006;4(11):e377.
170. Bitto E, McKay DB. The periplasmic molecular chaperone protein SurA binds a peptide motif that is characteristic of integral outer membrane proteins. *J Biol Chem.* 2003;278(49):49316–22.
171. Sklar JG, Wu T, Kahne D, Silhavy TJ. Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in *Escherichia coli*. *Genes Dev.* 2007;21(19):2473–84.
172. Qu J, Behrens-Kneip S, Holst O, Kleinschmidt JH. Binding regions of Outer Membrane Protein A in complexes with the periplasmic chaperone Skp. A site-directed fluorescence study. *Biochemistry.* 2009;48(22):4926–36.
173. Walton TA, Sandoval CM, Fowler CA, Pardi A, Sousa MC. The cavity-chaperone Skp protects its substrate from aggregation but allows independent folding of substrate domains. *Proc Natl Acad Sci U S A.* 2009;106(6):1772–7.
174. Burmann BM, Wang C, Hiller S. Conformation and dynamics of the periplasmic membrane-protein–chaperone complexes OmpX–Skp and tOmpA–Skp. *Nat Struct Mol Biol.* 2013;20(11):1265–72.
175. Hagan CL, Westwood DB, Kahne D. Bam lipoproteins assemble BamA in vitro. *Biochemistry.* 2013;52(35):6108–13.
176. Sandoval CM, Baker SL, Jansen K, Metzner SI, Sousa MC. Crystal structure of BamD: An essential component of the beta-barrel assembly machinery of gram-negative bacteria. *J Mol Biol.* 2011;409(3):348–57.
177. Voulhoux R, Bos MP, Geurtsen J, Mols M, Tommassen J. Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science.* 2003;299(5604):262–5.
178. Kim S, Malinverni JC, Sliz P, Silhavy TJ, Harrison SC, Kahne D. Structure and function of an essential component of the outer membrane protein assembly machine. *Science.* 2007;317(5840):961–4.
179. Albrecht R, Schütz M, Oberhettinger P, Faulstich M, Bermejo I, Rudel T, et al. Structure of BamA, an essential factor in outer membrane protein biogenesis. *Acta Crystallogr Sect D Biol Crystallogr.* 2014;70(6):1779–89.
180. Ni D, Wang Y, Yang X, Zhou H, Hou X, Cao B, et al. Structural and functional analysis of the  $\beta$ -barrel domain of BamA from *Escherichia coli*. *FASEB J.* 2014;28(6):2677–85.
181. Noinaj N, Kuszak AJ, Gumbart JC, Lukacik P, Chang H, Easley NC, et al. Structural insight into the biogenesis of  $\beta$ -barrel membrane proteins. *Nature.* 2013;501(7467):385–90.
182. Noinaj N, Rollauer SE, Buchanan SK. The  $\beta$ -barrel membrane protein insertase machinery from Gram-negative bacteria. *Curr Opin Struct Biol.* 2015;31:35–42.
183. Noinaj N, Kuszak AJ, Balusek C, Gumbart JC, Buchanan SK. Lateral opening and exit pore formation are required for BamA function. *Structure.* 2014;22(7):1055–62.
184. Costa TRD, Felisberto-Rodrigues C, Meir A, Prevost MS, Redzej A, Trokter M, et al. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat Rev Microbiol.* 2015;13(6):343–59.

185. Sauer RT, Baker TA. AAA+ Proteases: ATP-fueled machines of protein destruction. *Annu Rev Biochem.* 2011;80(1):587–612.
186. Dougan DA. *Regulated Proteolysis in Microorganisms.* Springer Netherlands, Inc., 2013;66.
187. Bochtler M, Hartmann C, Song HK, Bourenkov GP, Bartunik HD, Huber R. The structures of HslU and the ATP- dependent protease HslU – HslV. *Nature.* 2000;403(6771):800–5.
188. Sousa MC, Trame CB, Tsuruta H, Wilbanks SM, Reddy VS, McKay DB. Crystal and solution structures of an HslUV protease-chaperone complex. *Cell.* 2000;103(4):633–43.
189. Wang J, Hartling JA, Flanagan JM. The structure of ClpP at 2.3 Å resolution suggests a model for ATP-dependent proteolysis. *Cell.* 1997;91(4):447–56.
190. Suno R, Niwa H, Tsuchiya D, Zhang X, Yoshida M, Morikawa K. Structure of the whole cytosolic region of ATP-dependent protease FtsH. *Mol Cell.* 2006;22(5):575–85.
191. Park SC, Jia B, Yang J-K, Van D Le, Shao YG, Han SW, et al. Oligomeric structure of the ATP-dependent protease La (Lon) of *Escherichia coli*. *Mol Cells.* 2006;21(1):129–34.
192. Vieux EF, Wohlever ML, Chen JZ, Sauer RT, Baker T. Distinct quaternary structures of the AAA+ Lon protease control substrate degradation. *Proc Natl Acad Sci U S A.* 2013;110(22):E2002–8.
193. Katz C, Ron EZ. Dual role of FtsH in regulating lipopolysaccharide biosynthesis in *Escherichia coli*. *J Bacteriol.* 2008;190(21):7117–22.
194. Akiyama Y, Kihara A, Tokuda H, Ito K. FtsH (HflB) is an ATP-dependent protease selectively acting on SecY and some other membrane proteins. *J Biol Chem.* 1996;271(49):31196–201.
195. Akiyama Y. Quality control of cytoplasmic membrane proteins in *Escherichia coli*. *J Biochem.* 2009;146(4):449–54.
196. Narberhaus F, Obrist M, Führer F, Langklotz S. Degradation of cytoplasmic substrates by FtsH, a membrane-anchored protease with many talents. *Res Microbiol.* 2009;160(9):652–9.
197. Westphal K, Langklotz S, Thomanek N, Narberhaus F. A trapping approach reveals novel substrates and physiological functions of the essential protease FtsH in *Escherichia coli*. *J Biol Chem.* 2012;287(51):42962–71.
198. Kihara A, Akiyama Y, Koreaki I. Dislocation of membrane proteins in FtsH-mediated proteolysis. *EMBO J.* 1999;18(11):2970–81.
199. Langklotz S, Baumann U, Narberhaus F. Structure and function of the bacterial AAA protease FtsH. *Biochim Biophys Acta.* 2012;1823(1):40–8.
200. van Stelten J, Silva F, Belin D, Silhavy TJ. Effects of antibiotics and a proto-oncogene homolog on destruction of protein translocator SecY. *Science.* 2009;325(5941):753–6.
201. Kihara a, Akiyama Y, Ito K. A protease complex in the *Escherichia coli* plasma membrane: HflKC (HflA) forms a complex with FtsH (HflB), regulating its proteolytic activity against SecY. *EMBO J.* 1996;15(22):6122–31.
202. Kihara A, Akiyama Y, Ito K. Different pathways for protein degradation by the FtsH/HflKC membrane-embedded protease complex: an implication from the interference by a mutant form of a new substrate protein, YccA. *J Mol Biol.* 1998;279(1):175–88.
203. van Bloois E, Dekker HL, Fröderberg L, Houben ENG, Urbanus ML, de Koster CG, et al. Detection of cross-links between FtsH, YidC, HflK/C suggests a linked role for these proteins in quality control upon insertion of bacterial inner membrane proteins. *FEBS Lett.* 2008;582(10):1419–24.

204. Sakoh M, Ito K, Akiyama Y. Proteolytic activity of HtpX, a membrane-bound and stress-controlled protease from *Escherichia coli*. *J Biol Chem*. 2005;280(39):33305–10.
205. Maegawa S, Ito K, Akiyama Y. Proteolytic action of GlpG, a rhomboid protease in the *Escherichia coli* cytoplasmic membrane. *Biochemistry*. 2005;44(41):13543–52.
206. Erez E, Bibi E. Cleavage of a multispanning membrane protein by an intramembrane serine protease. *Biochemistry*. 2009;48(51):12314–22.
207. Baker RP, Young K, Feng L, Shi Y, Urban S. Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proc Natl Acad Sci U S A*. 2007;104(20):8257–62.
208. Strauch KL, Beckwith J. An *Escherichia coli* mutation preventing degradation of abnormal periplasmic proteins. *Proc Natl Acad Sci U S A*. 1988;85(5):1576–80.
209. Strauch KL, Johnson K, Beckwith J. Characterization of DegP, a gene required for proteolysis in the cell-envelope and essential for growth of *Escherichia coli* at high temperature. *J Bacteriol*. 1989;171(5):2689–96.
210. Spiess C, Beil A, Ehrmann M. A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell*. 1999;97(3):339–47.
211. Krojer T, Sawa J, Schäfer E, Saibil HR, Ehrmann M, Clausen T. Structural basis for the regulated protease and chaperone function of DegP. *Nature*. 2008;453(7197):885–90.
212. Kim S, Grant RA, Sauer RT. Covalent linkage of distinct substrate degrons controls assembly and disassembly of DegP proteolytic cages. *Cell*. 2011;145(1):67–78.
213. Waller PRH, Sauer RT. Characterization of degQ and degS, *Escherichia coli* genes encoding homologs of the DegP protease. *J Bacteriol*. 1996;178(4):1146–53.
214. Bai XC, Pan XJ, Wang XJ, Ye YY, Chang LF, Leng D, et al. Characterization of the structure and function of *Escherichia coli* DegQ as a representative of the DegQ-like proteases of bacterial HtrA family proteins. *Structure*. 2011;19(9):1328–37.
215. Malet H, Canellas F, Sawa J, Yan J, Thalassinou K, Ehrmann M, et al. Newly folded substrates inside the molecular cage of the HtrA chaperone DegQ. *Nat Struct Mol Biol*. 2012;19(2):152–7.
216. Sawa J, Malet H, Krojer T, Canellas F, Ehrmann M, Clausen T. Molecular adaptation of the DegQ protease to exert protein quality control in the bacterial cell envelope. *J Biol Chem*. 2011;286(35):30680–90.
217. Sohn J, Grant RA, Sauer RT. OMP peptides activate the DegS stress-sensor protease by a relief of inhibition mechanism. *Structure*. 2009;17(10):1411–21.
218. Herman C, Prakash S, Lu CZ, Matouschek A, Gross CA. Lack of a robust unfoldase activity confers a unique level of substrate specificity to the universal AAA protease FtsH. *Mol Cell*. 2003;11(3):659–69.
219. Smith CK, Baker TA, Sauer RT. Lon and Clp family proteases and chaperones share homologous substrate-recognition domains. *Proc Natl Acad Sci U S A*. 1999;96(12):6678–82.
220. Gur E, Sauer RT. Recognition of misfolded proteins by Lon, a AAA + protease. *Genes Dev*. 2008;22(16):2267–77.
221. Burton RE, Baker TA, Sauer RT. Nucleotide-dependent substrate recognition by the AAA+ HslUV protease. *Nat Struct Mol Biol*. 2005;12(3):245–51.
222. Lee Y-Y, Chang C-F, Kuo C-L, Chen M-C, Yu CH, Lin P-I, et al. Subunit oligomerization and substrate recognition of the *Escherichia coli* ClpYQ (HslUV) protease implicated by in vivo protein-protein interactions in the yeast two-hybrid system. *J Bacteriol*. 2003;185(8):2393–401.

223. Sharma UK, Chatterji D. Transcriptional switching in *Escherichia coli* during stress and starvation by modulation of  $\sigma^{70}$  activity. *FEMS Microbiol Rev.* 2010;34(5):646–57.
224. Wade JT, Roa DC, Grainger DC, Hurd D, Busby SJW, Struhl K, et al. Extensive functional overlap between  $\sigma$  factors in *Escherichia coli*. *Nat Struct Mol Biol.* 2006;13(9):806–14.
225. Zhao K, Liu M, Burgess RR. The global transcriptional response of *Escherichia coli* to induced  $\sigma^{32}$  protein involves  $\sigma^{32}$  regulon activation followed by inactivation and degradation of  $\sigma^{32}$  in vivo. *J Biol Chem.* 2005;280(18):17758–68.
226. Nonaka G, Blankschien M, Herman C, Gross CA, Rhodius VA. Regulon and promoter analysis of the *E. coli* heat-shock factor,  $\sigma^{32}$ , reveals a multifaceted cellular response to heat stress. *Genes Dev.* 2006;20(13):1776–89.
227. Mogk A, Deuring E, Vorderwülbecke S, Vierling E, Bukau B. Small heat shock proteins, ClpB and the DnaK system form a functional triade in reversing protein aggregation. *Mol Microbiol.* 2003;50(2):585–95.
228. Morita M, Kanemori M, Yanagi H. Heat-Induced Synthesis of  $\sigma^{32}$  in *Escherichia coli*: Structural and Functional Dissection of rpoH mRNA Secondary Structure. *J bacteriol.* 1999;181(2):401–10.
229. Guisbert E, Herman C, Lu CZ, Gross CA. A chaperone network controls the heat shock response in *E. coli*. *Genes Dev.* 2004;18(22):2812–21.
230. Rodriguez F, Arsène-Ploetze F, Rist W, Rüdiger S, Schneider-Mergener J, Mayer MP, et al. Molecular Basis for Regulation of the Heat Shock Transcription Factor  $\sigma^{32}$  by the DnaK and DnaJ Chaperones. *Mol Cell.* 2008;32(3):347–58.
231. Meyer AS, Baker TA. Proteolysis in the *Escherichia coli* heat shock response: a player at many levels. *Curr Opin Microbiol.* 2011;14(2):194–9.
232. Mecsas J, Rouviere PE, Erickson JW, Donohue TJ, Gross CA. The activity of  $\sigma^E$ , an *Escherichia coli* heat-inducible sigma-factor, is modulated by expression of outer membrane proteins. *Genes Dev.* 1993;7(12B):2618–28.
233. Kabir MS. Cell lysis directed by  $\sigma^E$  in early stationary phase and effect of induction of the rpoE gene on global gene expression in *Escherichia coli*. *Microbiology.* 2005;151(8):2721–35.
234. Danese PN, Silhavy TJ. The  $\sigma^E$  and the Cpx signal transduction systems control the synthesis of periplasmic protein-folding enzymes in *Escherichia coli*. *Genes Dev.* 1997;11(9):1183–93.
235. Dartigalongue C, Missiakas D, Raina S. Characterization of the *Escherichia coli*  $\sigma^E$  regulon. *J Biol Chem.* 2001;276(24):20866–75.
236. Raivio TL, Silhavy TJ. The  $\sigma^E$  and Cpx regulatory pathways: Overlapping but distinct envelope stress responses. *Curr Opin Microbiol.* 1999;2(2):159–65.
237. Wang Q, Kaguni JONM. A Novel Sigma Factor Is Involved in Expression of the rpoH Gene of *Escherichia coli*. *J. Bacteriol.* 1989;171(8):4248–53.
238. Ades SE, Grigorova IL, Gross CA. Regulation of the alternative sigma factor  $\sigma^E$  during initiation, adaptation, and shutoff of the extracytoplasmic heat shock response in *Escherichia coli*. *J Bacteriol.* 2003;185(8):2512–9.
239. Raina S, Missiakas D, Georgopoulos C. The rpoE gene encoding the  $\sigma^E$  ( $\sigma^{24}$ ) heat shock sigma factor of *Escherichia coli*. *EMBO J.* 1995;14(5):1043–55.
240. Ades SE. Regulation by destruction: design of the  $\sigma^E$  envelope stress response. *Curr Opin Microbiol.* 2008;11(6):535–40.

241. Regt AK de, Baker TA, Sauer RT. Steric clashes with bound OMP peptides activate the DegS stress-response protease. *Proc Natl Acad Sci.* 2015;112(11):3326–31.
242. Hunke S, Betton J-M. Temperature effect on inclusion body formation and stress response in the periplasm of *Escherichia coli*. *Mol Microbiol.* 2003;50(5):1579–89.
243. Raivio TL, Popkin DL, Silhavy TJ. The Cpx envelope stress response is controlled by amplification and feedback inhibition. *J Bacteriol.* 1999;181(17):5263–72.
244. Raivio TL, Silhavy TJ. Periplasmic stress and ECF sigma factors. *Annu Rev Microbiol.* 2001;55(1):591–624.
245. Price NL, Raivio TL. Characterization of the Cpx regulon in *Escherichia coli* strain MC4100. *J Bacteriol.* 2009;191(6):1798–815.
246. Danese PN, Silhavy TJ. CpxP, a stress-combative member of the Cpx regulon. *J Bacteriol.* 1998;180(4):831–9.
247. Lundstrom K. Structural genomics and drug discovery. *J Cell Mol Med.* 2007;11(2):224–38.
248. Gubellini F, Verdon G, Karpowich NK, Luff JD, Boel G, Gauthier N, et al. Physiological response to membrane protein overexpression in *E. coli*. *Mol Cell Proteomics.* 2011;10(10).
249. Skretas G, Makino T, Varadarajan N, Pogson M, Georgiou G. Multi-copy genes that enhance the yield of mammalian G protein-coupled receptors in *Escherichia coli*. *Metab Eng.* 2012;14(5):591–602.
250. Bernhard F, Tozawa Y. Cell-free expression - making a mark. *Curr Opin Struct Biol.* 2013;23(3):374–80.
251. Henrich E, Hein C, Dötsch V, Bernhard F. Membrane protein production in *Escherichia coli* cell-free lysates. *FEBS Lett.* 2015;589(15):1713–22.
252. Matsuda T, Watanabe S, Kigawa T. Cell-free synthesis system suitable for disulfide-containing proteins. *Biochem Biophys Res Commun.* 2013;431(2):296–301.
253. Hatahet F, Blazyk JL, Martineau E, Mandela E, Zhao Y, Campbell RE, et al. Altered *Escherichia coli* membrane protein assembly machinery allows proper membrane assembly of eukaryotic protein vitamin K epoxide reductase. *Proc Natl Acad Sci.* 2015; doi: 10.1073/pnas.1521260112.
254. Schlegel S, Hjelm A, Baumgarten T, Vikström D, de Gier J-W. Bacterial-based membrane protein production. *Biochim Biophys Acta.* 2014;1843(8):1739–49.
255. Fernández FJ, Vega MC. Technologies to keep an eye on: alternative hosts for protein production in structural biology. *Curr Opin Struct Biol.* 2013;23(3):365–73.
256. Chen R. Bacterial expression systems for recombinant protein production: *E. coli* and beyond. *Biotechnol Adv.* 2012;30(5):1102–7.
257. Pohl S, Bhavsar G, Hulme J, Bloor AE, Misirli G, Leckenby MW, et al. Proteomic analysis of *Bacillus subtilis* strains engineered for improved production of heterologous proteins. *Proteomics.* 2013;13(22):3298–308.
258. Grishammer R, Tateu CG. Overexpression of integral membrane proteins for structural studies. *Q Rev Biophys.* 1995;28(3):315–422.
259. Daegelen P, Studier FW, Lenski RE, Cure S, Kim JF. Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL606 and BL21(DE3). *J Mol Biol.* 2009;394(4):634–43.

260. Shulman ST, Friedmann HC, Sims RH. Theodor Escherich: the first pediatric infectious diseases physician? *Clin Infect Dis*. 2007;45(8):1025–9.
261. Hawley DK, McClure WR, Limited IRLP. Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res*. 1983;11(8):2237-55.
262. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*. 1961;3(3):318–56.
263. Ullmann A. *Escherichia coli* Lactose Operon. *Encycl Life Sci*. 2009.
264. Kolb A, Busby S, Buc H, Garges S, Adhya S. Transcriptional regulation by cAMP and its receptor protein. *Annu Rev Biochem*. 1993;62:749–95.
265. Görke B, Stülke J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol*. 2008;6(8):613–24.
266. Khlebnikov A, Keasling JD. Effect of lacY expression on homogeneity of induction from the P(tac) and P(trc) promoters by natural and synthetic inducers. *Biotechnol Prog*. 2002;18(3):672–4.
267. Giacalone MJ, Gentile AM, Lovitt BT, Berkley NL, Gunderson CW, Surber MW. Toxic protein expression in *Escherichia coli* using a rhamnose-based tightly regulated and tunable promoter system. *Biotechniques*. 2006;40(3):355–64.
268. Wickstrum JR, Santangelo TJ, Egan SM. Cyclic AMP receptor protein and RhaR synergistically activate transcription from the L-rhamnose-responsive rhaSR promoter in *Escherichia coli*. *J Bacteriol*. 2005;187(19):6708–18.
269. Wickstrum JR, Skredenske JM, Balasubramaniam V, Jones K, Egan SM. The AraC/XylS family activator RhaS negatively autoregulates rhaSR expression by preventing cyclic AMP receptor protein activation. *J Bacteriol*. 2010;192(1):225–32.
270. Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, et al. Tuning *Escherichia coli* for membrane protein overexpression. *Proc Natl Acad Sci U S A*. 2008;105(38):14371–6.
271. Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose P(BAD) promoter. *J Bacteriol*. 1995;177(14):4121–30.
272. Schleif R. AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiol Rev*. 2010;34(5):779–96.
273. Siegele DA, Hu JC. Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. *Proc Natl Acad Sci U S A*. 1997;94(15):8168–72.
274. Megerle JA, Fritz G, Gerland U, Jung K, Rädler JO. Timing and dynamics of single cell gene expression in the arabinose utilization system. *Biophys J*. 2008;95(4):2103–15.
275. Khlebnikov A, Datsenko KA, Skaug T, Wanner BL, Keasling JD. Homogeneous expression of the PBAD promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity araE transporter. *Microbiology*. 2001;147(12):3241–7.
276. Khlebnikov A, Skaug T, Keasling JD. Modulation of gene expression from the arabinose-inducible araBAD promoter. *J Ind Microbiol Biotechnol*. 2002;29(1):34–7.
277. Morgan-Kiss RM, Wadler C, Cronan JE. Long-term and homogeneous regulation of the *Escherichia coli* araBAD promoter by use of a lactose transporter of relaxed specificity. *Proc Natl Acad Sci U S A*. 2002;99(11):7373–7.

278. Bowers LM, Lapoint K, Anthony L, Pluciennik A, Filutowicz M. Bacterial expression system with tightly regulated gene expression and plasmid copy number. *Gene*. 2004;340(1):11–8.
279. Skerra A. Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in *Escherichia coli*. *Gene*. 1994;151(1-2):131–5.
280. Muthukrishnan A-B, Kandhavelu M, Lloyd-Price J, Kudasov F, Chowdhury S, Yli-Harja O, et al. Dynamics of transcription driven by the tetA promoter, one event at a time, in live *Escherichia coli* cells. *Nucleic Acids Res*. 2012;40(17):8472–83.
281. Valdez-Cruz NA, Caspeta L, Pérez NO, Ramírez OT, Trujillo-Roldán MA. Production of recombinant proteins in *E. coli* by the heat inducible expression system based on the phage lambda pL and/or pR promoters. *Microb Cell Fact*. 2010;9(18).
282. del Solar G, Giraldo R, Ruiz-Echevarría MJ, Espinosa M, Díaz-Orejás R. Replication and control of circular bacterial plasmids. *Microbiol Rev*. 1998;62(2):434–64.
283. Galán JE, Nakayama K, Curtiss R. Cloning and characterization of the *asd* gene of *Salmonella typhimurium*: use in stable maintenance of recombinant plasmids in *Salmonella* vaccine strains. *Gene*. 1990;94(1):29–35.
284. Unterholzner SJ, Poppenberger B, Rozhon W. Toxin-antitoxin systems: Biology, identification, and application. *Mob Genet Elements*. 2013;3(5):e26219.
285. Ringquist S, Shinedling S, Barrick D, Green L, Binkley J, Stormo GD, et al. Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. *Mol Microbiol*. 1992;6(9):1219–29.
286. Mirzadeh K, Martínez V, Toddo S, Guntur S, Herrgård MJ, Elofsson A, et al. Enhanced protein production in *Escherichia coli* by optimization of cloning scars at the vector–coding sequence junction. *ACS Synth Biol*. 2015;4(9):959–65.
287. Studier FW, Moffatt BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol*. 1986;189(1):113–30.
288. Baneyx F, Georgiou G. In vivo degradation of secreted fusion proteins by the *Escherichia coli* outer membrane protease OmpT. *J Bacteriol*. 1990;172(1):491–4.
289. Shineberg B, Zipser D. The *lon* gene and degradation of  $\beta$ -galactosidase nonsense fragments. *J Bacteriol*. 1973;116(3):1469–71.
290. Iost I, Guillerez J, Dreyfus M. Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo. *J Bacteriol*. 1992;174(2):619–22.
291. Arditti RR, Scaife JG, Beckwith JR. The nature of mutants in the *lac* promoter region. *J Mol Biol*. 1968;38(3):421–6.
292. Wanner BL, Kodaira R, Neidhardt FC. Physiological regulation of a decontrolled *lac* operon. *J Bacteriol*. 1977;130(1):212–22.
293. Dubendorff JW, Studier FW. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with *lac* repressor. *J Mol Biol*. 1991;219(1):45–59.
294. Studier FW. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol*. 1991;219(1):37–44.
295. Miroux B, Walker JE. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol*. 1996;260(3):289–98.

296. Dumon-Seignovert L, Cariot G, Vuillard L. The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr Purif.* 2004;37(1):203–6.
297. Schlegel S, Genevaux P, de Gier J-W. De-convoluting the genetic adaptations of *E. coli* C41(DE3) in real time reveals how alleviating protein production stress improves yields. *Cell Rep.* 2015;10(10):1758–66.
298. Kwon S-K, Kim SK, Lee D-H, Kim JF. Comparative genomics and experimental evolution of *Escherichia coli* BL21(DE3) strains reveal the landscape of toxicity escape from membrane protein overproduction. *Sci Rep.* 2015;5(16076).
299. Gul N, Linares DM, Ho FY, Poolman B. Evolved *Escherichia coli* strains for amplified, functional expression of membrane proteins. *J Mol Biol.* 2014;426(1):136–49.
300. Massey-Gendel E, Zhao A, Boulting G, Kim H-Y, Balamotis M a, Seligman LM, et al. Genetic selection system for improving recombinant membrane protein expression in *E. coli*. *Protein Sci.* 2009;18(2):372–83.
301. Skretas G, Georgiou G. Genetic analysis of G protein-coupled receptor expression in *Escherichia coli*: Inhibitory role of DnaJ on the membrane integration of the human central cannabinoid receptor. *Biotechnol Bioeng.* 2009;102(2):357–67.
302. Skretas G, Georgiou G. Simple genetic selection protocol for isolation of overexpressed genes that enhance accumulation of membrane-integrated human G protein-coupled receptors in *Escherichia coli*. *Appl Environ Microbiol.* 2010;76(17):5852–9.
303. Chen G, Hayhurst A, Thomas JG, Harvey BR, Iverson BL, Georgiou G. Isolation of high-affinity ligand-binding proteins by periplasmic expression with cytometric screening (PECS). *Nat Biotechnol.* 2001;19(6):537–42.
304. Makino T, Skretas G, Kang T-H, Georgiou G. Comprehensive engineering of *Escherichia coli* for enhanced expression of IgG antibodies. *Metab Eng.* 2011;13(2):241–51.
305. Nannenga BL, Baneyx F. Enhanced expression of membrane proteins in *E. coli* with a PBAD promoter mutant: synergies with chaperone pathway engineering strategies. *Microb Cell Fact.* 2011;10(105).
306. Luo J, Choulet J, Samuelson JC. Rational design of a fusion partner for membrane protein expression in *E. coli*. *Protein Sci.* 2009;18(8):1735–44.
307. Kim HS, Ernst JA, Brown C, Bostrom J, Fuh G, Lee C V., et al. Translation levels control multi-spanning membrane protein expression. *PLoS One.* 2012;7(4):e35844.
308. Kadokura H, Watanabe K, Tsuneizumi K, Yoda K, Yamasaki M. Physiological and biochemical analysis of the effects of alkaline phosphatase overproduction in *Escherichia coli*. *J Bacteriol.* 1995;177(12):3596–600.
309. Simmons LC, Yansura DG. Translational level is a critical factor for the secretion of heterologous proteins in *Escherichia coli*. *Nat Biotechnol.* 1996;14(5):629–34.
310. Mergulhão FJM, Monteiro GA, Larsson G, Sandén AM, Farewell A, Nyström T, et al. Medium and copy number effects on the secretion of human proinsulin in *Escherichia coli* using the universal stress promoters *uspA* and *uspB*. *Appl Microbiol Biotechnol.* 2003;61(5-6):495–501.
311. Mergulhão FJM, Monteiro GA, Larsson G, Bostrom M, Farewell A, Nyström T, et al. Evaluation of inducible promoters on the secretion of a ZZ-proinsulin fusion protein in *Escherichia coli*. *Biotechnol Appl Biochem.* 2003;38(1):87–93.

312. Schlegel S, Rujas E, Ytterberg AJ, Zubarev RA, Luirink J, de Gier J-W. Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels. *Microb Cell Fact*. 2013;12(24).
313. Nannenga BL, Baneyx F. Reprogramming chaperone pathways to improve membrane protein expression in *Escherichia coli*. *Protein Sci*. 2011;20(8):1411–20.
314. Puertas J-M, Nannenga BL, Dornfeld KT, Betton J-M, Baneyx F. Enhancing the secretory yields of leech carboxypeptidase inhibitor in *Escherichia coli*: influence of trigger factor and signal recognition particle. *Protein Expr Purif*. 2010;74(1):122–8.
315. Bäcklund E, Ignatushchenko M, Larsson G. Suppressing glucose uptake and acetic acid production increases membrane protein overexpression in *Escherichia coli*. *Microb Cell Fact*. 2011;10(35).
316. Waegeman H, De Lausnay S, Beauprez J, Maertens J, De Mey M, Soetaert W. Increasing recombinant protein production in *Escherichia coli* K12 through metabolic engineering. *N Biotechnol*. 2013;30(2):255–61.
317. Salema V, Fernández LÁ. High yield purification of nanobodies from the periplasm of *E. coli* as fusions with the maltose binding protein. *Protein Expr Purif*. 2013;91(1):42–8.
318. Chen C, Snedecor B, Nishihara JC, Joly JC, McFarland N, Andersen DC, et al. High-level accumulation of a recombinant antibody fragment in the periplasm of *Escherichia coli* requires a triple-mutant (degP pre spr) host strain. *Biotechnol Bioeng*. 2004;85(5):463–74.
319. Gustavsson M, Bäcklund E, Larsson G. Optimisation of surface expression using the AIDA autotransporter. *Microb Cell Fact*. 2011;10(72).
320. Meerman HJ, Georgiou G. Construction and characterization of a set of *E. coli* strains deficient in all known loci affecting the proteolytic stability of secreted recombinant proteins. *Biotechnology*. 1994;12(11):1107–10.
321. Obukowicz MG, Staten NR, Krivi GG. Enhanced heterologous gene expression in novel rpoH mutants of *Escherichia coli*. *Appl Environ Microbiol*. 1992;58(5):1511–23.
322. Kadokura H, Kawasaki H, Yoda K, Yamasaki M, Kitamoto K. Efficient export of alkaline phosphatase overexpressed from a multicopy plasmid requires degP, a gene encoding a periplasmic protease of *Escherichia coli*. *J Gen Appl Microbiol*. 2001;47(3):133–41.
323. Shin H-D, Chen RR. Extracellular recombinant protein production from an *Escherichia coli* lpp deletion mutant. *Biotechnol Bioeng*. 2008;101(6):1288–96.
324. Daleke-Schermerhorn MH, Felix T, Soprova Z, Ten Hagen-Jongman CM, Vikström D, Majlessi L, et al. Decoration of outer membrane vesicles with multiple antigens using an autotransporter approach. *Appl Environ Microbiol*. 2014;80(18):5854–65.
325. Link AJ, Skretas G, Strauch E-M, Chari NS, Georgiou G. Efficient production of membrane-integrated and detergent-soluble G protein-coupled receptors in *Escherichia coli*. *Protein Sci*. 2008;17(10):1857–63.
326. Xu LY, Link AJ. Stress responses to heterologous membrane protein expression in *Escherichia coli*. *Biotechnol Lett*. 2009;31(11):1775–82.
327. Chen Y, Song J, Sui S, Wang D-N. DnaK and DnaJ facilitated the folding process and reduced inclusion body formation of magnesium transporter CorA overexpressed in *Escherichia coli*. *Protein Expr Purif*. 2003;32(2):221–31.
328. Valent QA, de Gier JW, von Heijne G, Kendall DA, ten Hagen-Jongman CM, Oudega B, et al. Nascent membrane and presecretory proteins synthesized in *Escherichia coli* associate with signal recognition particle and trigger factor. *Mol Microbiol*. 1997;25(1):53–64.

329. Lee YJ, Jeong KJ. Challenges to production of antibodies in bacteria and yeast. *J Biosci Bioeng.* 2015;120(5):483–90.
330. Sonoda H, Kumada Y, Katsuda T, Yamaji H. Effects of cytoplasmic and periplasmic chaperones on secretory production of single-chain Fv antibody in *Escherichia coli*. *J Biosci Bioeng.* 2011;111(4):465–70.
331. Hu X, O'Hara L, White S, Magner E, Kane M, Gerard Wall J. Optimisation of production of a domoic acid-binding scFv antibody fragment in *Escherichia coli* using molecular chaperones and functional immobilisation on a mesoporous silicate support. *Protein Expr Purif.* 2007;52(1):194–201.
332. Kurokawa Y, Yanagi H, Yura T. Overexpression of protein disulfide isomerase DsbC stabilizes multiple-disulfide-bonded recombinant protein produced and transported to the periplasm in *Escherichia coli*. *Appl Environ Microbiol.* 2000;66(9):3960–5.
333. Maskos K, Huber-Wunderlich M, Glockshuber R. DsbA and DsbC-catalyzed oxidative folding of proteins with complex disulfide bridge patterns in vitro and in vivo. *J Mol Biol.* 2003;325(3):495–513.
334. Lee YJ, Kim HS, Ryu AJ, Jeong KJ. Enhanced production of full-length immunoglobulin G via the signal recognition particle (SRP)-dependent pathway in *Escherichia coli*. *J Biotechnol.* 2013;165(2):102–8.
335. Lee YJ, Lee DH, Jeong KJ. Enhanced production of human full-length immunoglobulin G1 in the periplasm of *Escherichia coli*. *Appl Microbiol Biotechnol.* 2013;98(3):1237–46.
336. Bothmann H, Plückthun A. The Periplasmic *Escherichia coli* Peptidylprolyl cis,trans-isomerase FkpA I. increased functional expression of antibody fragments with and without cis-prolines. *J Biol Chem.* 2000;275(22):17100–5.
337. Schlapschy M, Grimm S, Skerra A. A system for concomitant overexpression of four periplasmic folding catalysts to improve secretory protein production in *Escherichia coli*. *Protein Eng Des Sel.* 2006;19(8):385–90.
338. Wang R, Xiang S, Feng Y, Srinivas S, Zhang Y, Lin M, et al. Engineering production of functional scFv antibody in *E. coli* by co-expressing the molecule chaperone Skp. *Front Cell Infect Microbiol.* 2013;3(72).
339. Prinz WA, Aslund F, Holmgren A, Beckwith J. The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. *J Biol Chem.* 1997;272(25):15661–7.
340. Derman AI, Prinz WA, Belin D, Beckwith J. Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science.* 1993;262(5140):1744–7.
341. Bessette PH, Aslund F, Beckwith J, Georgiou G. Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proc Natl Acad Sci U S A.* 1999;96(24):13703–8.
342. Ritz D, Lim J, Reynolds CM, Poole LB, Beckwith J. Conversion of a peroxiredoxin into a disulfide reductase by a triplet repeat expansion. *Science.* 2001;294(5540):158–60.
343. Hatahet F, Nguyen VD, Salo KEH, Ruddock LW. Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of *E. coli*. *Microb Cell Fact.* 2010;9(67).
344. Nguyen V, Hatahet F, Salo KE, Enlund E, Zhang C, Ruddock LW. Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of *E. coli*. *Microb Cell Fact.* 2011;10(1).

345. Hatahet F, Ruddock LW. Topological plasticity of enzymes involved in disulfide bond formation allows catalysis in either the periplasm or the cytoplasm. *J Mol Biol.* 2013;425(18):3268–76.
346. Lobstein J, Emrich CA, Jeans C, Faulkner M, Riggs P, Berkmen M. SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb Cell Fact.* 2012;11(56).
347. Scott DJ, Plückthun A. Direct Molecular Evolution of Detergent-Stable G Protein-Coupled Receptors Using Polymer Encapsulated Cells. *J Mol Biol.* 2013;425(3):662–77.
348. Egloff P, Hillenbrand M, Klenk C, Batyuk A, Heine P, Balada S, et al. Structure of signaling-competent neurotensin receptor 1 obtained by directed evolution in *Escherichia coli*. *Proc Natl Acad Sci U S A.* 2014;111(6):E655–62.
349. Zhou Y, Bowie JU. Building a thermostable membrane protein. *J Biol Chem.* 2000;275(10):6975–9.
350. Scott DJ, Kummer L, Tremmel D, Plückthun A. Stabilizing membrane proteins through protein engineering. *Curr Opin Chem Biol.* 2013;17(3):427–35.
351. Heydenreich FM, Vuckovic Z, Matkovic M, Veprintsev DB. Stabilization of G protein-coupled receptors by point mutations. *Front Pharmacol.* 2015;6(82).
352. Sarkar CA, Dodevski I, Kenig M, Dudli S, Mohr A, Hermans E, et al. Directed evolution of a G protein-coupled receptor for expression, stability, and binding selectivity. *Proc Natl Acad Sci U S A.* 2008;105(39):14808–13.
353. Schlinkmann KM, Hillenbrand M, Rittner A, Kunz M, Strohner R, Pluckthun A. Maximizing detergent stability and functional expression of a GPCR by exhaustive recombination and evolution. *J Mol Biol.* 2012;422(3):414–28.
354. Chen K-YM, Zhou F, Fryszczyn BG, Barth P. Naturally evolved G protein-coupled receptors adopt metastable conformations. *Proc Natl Acad Sci.* 2012;109(33):13284–9.
355. Shibata Y, Gvozdenovic-Jeremic J, Love J, Kloss B, White JF, Grisshammer R, et al. Optimising the combination of thermostabilising mutations in the neurotensin receptor for structure determination. *Biochim Biophys Acta.* 2013;1828(4):1293–301.
356. Quax TEF, Claassens NJ, Söll D, van der Oost J. Codon bias as a means to fine-tune gene expression. *Mol Cell.* 2015;59(2):149–61.
357. Lithwick G, Margalit H. Hierarchy of sequence-dependent features associated with prokaryotic translation. *Genome Res.* 2003;13(12):2665–73.
358. Gorochowski TE, Ignatova Z, Bovenberg RAL, Roubos JA. Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate. *Nucleic Acids Res.* 2015;43(6):3022–32.
359. Nørholm MHH, Light S, Virkki MTI, Elofsson A, Von Heijne G, Daley DO. Manipulating the genetic code for membrane protein production: What have we learnt so far? *Biochim Biophys Acta.* 2012;1818(4):1091–6.
360. Fluman N, Navon S, Bibi E, Pilpel Y. mRNA-programmed translation pauses in the targeting of *E. coli* membrane proteins. *eLife.* 2014;3(e03440).
361. Kudla G, Murray AW, Tollervay D, Plotkin JB. Coding Sequence Determinant of Protein Expression in *E. coli*. *Science.* 2009;324(5924):255–8.
362. Goodman DB, Church GM, Kosuri S. Causes and effects of N-terminal codon bias in bacterial genes. *Science.* 2013;342(6157):475–9.

363. Nørholm MHH, Toddo S, Virkki MTI, Light S, Von Heijne G, Daley DO. Improved production of membrane proteins in *Escherichia coli* by selective codon substitutions. *FEBS Lett.* 2013;587(15):2352–8.
364. Marino J, Hohl M, Seeger MA, Zerbe O, Geertsma ER. Bicistronic mRNAs to enhance membrane protein overexpression. *J Mol Biol.* 2015;427(4):943–54.
365. Cheong D-E, Ko K-C, Han Y, Jeon H-G, Sung BH, Kim G-J, et al. Enhancing functional expression of heterologous proteins through random substitution of genetic codes in the 5' coding region. *Biotechnol Bioeng.* 2015;112(4):822–6.
366. Grisshammer R, Duckworth R, Henderson R. Expression of a rat neurotensin receptor in *Escherichia coli*. *Biochem J.* 1993;295(Pt 2):571–6.
367. Gromek KA, Meddaugh HR, Wrobel RL, Suchy FP, Bingman CA, Primm JG, et al. Improved expression and purification of sigma 1 receptor fused to maltose binding protein by alteration of linker sequence. *Protein Expr Purif.* 2013;89(2):203–9.
368. Hu J, Qin H, Philip F, Cross TA. A systematic assessment of mature MBP in membrane protein production: Overexpression, membrane targeting and purification. *Protein Expr Purif.* 2011;80(1):34–40.
369. Chowdhury A, Feng R, Tong Q, Zhang Y, Xie X-Q. Mystic and TarCF as fusion protein partners for functional expression of the cannabinoid receptor 2 in *Escherichia coli*. *Protein Expr Purif.* 2012;83(2):128–34.
370. Dvir H, Choe S. Bacterial expression of a eukaryotic membrane protein in fusion to various Mystic orthologs. *Protein Expr Purif.* 2009;68(1):28–33.
371. Kefala G, Kwiatkowski W, Esquivies L, Maslennikov I, Choe S. Application of Mystic to improving the expression and membrane integration of histidine kinase receptors from *Escherichia coli*. *J Struct Funct Genomics.* 2007;8(4):167–72.
372. Roosild TP, Greenwald J, Vega M, Castronovo S, Riek R, Choe S. NMR structure of Mystic, a membrane-integrating protein for membrane protein expression. *Science.* 2005;307(5713):1317–21.
373. Deniaud A, Bernaudat F, Frelet-Barrand A, Juillan-Binard C, Vernet T, Rolland N, et al. Expression of a chloroplast ATP/ADP transporter in *E. coli* membranes: Behind the Mystic strategy. *Biochim Biophys Acta.* 2011;1808(8):2059–66.
374. Hsu M-F, Yu T-F, Chou C-C, Fu H-Y, Yang C-S, Wang AHJ. Using Haloarcula marismortui bacteriorhodopsin as a fusion tag for enhancing and visible expression of integral membrane proteins in *Escherichia coli*. *PLoS One.* 2013;8(2):e56363.
375. Mansell TJ, Linderman SW, Fisher AC, DeLisa MP. A rapid protein folding assay for the bacterial periplasm. *Protein Sci.* 2010;19(5):1079–90.
376. Drew D, Sjöstrand D, Nilsson J, Urbig T, Chin C, de Gier J-W, et al. Rapid topology mapping of *Escherichia coli* inner-membrane proteins by prediction and PhoA/GFP fusion analysis. *Proc Natl Acad Sci U S A.* 2002;99(5):2690–5.
377. Feilmeier BJ, Iseminger G, Schroeder D, Webber H, Phillips GJ. Green fluorescent protein functions as a reporter for protein localization in *Escherichia coli*. *Bacteriology.* 2000;182(14):4068–76.
378. Drew DE, von Heijne G, Nordlund P, de Gier JW. Green fluorescent protein as an indicator to monitor membrane protein overexpression in *Escherichia coli*. *FEBS Lett.* 2001;507(2):220–4.
379. Drew D, Lerch M, Kunji E, Slotboom D-J, de Gier J-W. Optimization of membrane protein overexpression and purification using GFP fusions. *Nat Methods.* 2006;3(4):303–13.

380. Drew D, Slotboom D-J, Friso G, Reda T, Genevax P, Rapp M, et al. A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. *Protein Sci.* 2005;14(8):2011–7.
381. Kawate T, Gouaux E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure.* 2006;14(4):673–81.
382. Geertsma ER, Groeneveld M, Slotboom D-J, Poolman B. Quality control of overexpressed membrane proteins. *Proc Natl Acad Sci U S A.* 2008;105(15):5722–7.
383. Daley DO, Rapp M, Granseth E, Melén K, Drew D. Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science.* 2005;308(5726):1321–3.
384. Hsieh JM, Besserer GM, Madej MG, Bui HQ, Kwon S, Abramson J. Bridging the gap: A GFP-based strategy for overexpression and purification of membrane proteins with intra and extracellular C-termini. *Protein Sci.* 2010;19(4):868–80.
385. Toddo S, Söderström B, Palombo I, Von Heijne G, Nørholm MHH, Daley DO. Application of split-green fluorescent protein for topology mapping membrane proteins in *Escherichia coli*. *Protein Sci.* 2012;21(10):1571–6.
386. Pédelacq J-D, Cabantous S, Tran T, Terwilliger TC, Waldo GS. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol.* 2006;24(1):79–88.
387. Lee YJ, Jeong KJ. Enhanced production of antibody fragment via SRP pathway engineering in *Escherichia coli*. *Biotechnol Bioprocess Eng.* 2013;18(4):751–8.
388. Mergulhão FJ, Monteiro GA. Analysis of factors affecting the periplasmic production of recombinant proteins in *Escherichia coli*. *J Microbiol Biotechnol.* 2007;17(8):1236–41.
389. Gordon E, Horsefield R, Swarts HGP, de Pont JJHM, Neutze R, Snijder A. Effective high-throughput overproduction of membrane proteins in *Escherichia coli*. *Protein Expr Purif.* 2008;62(1).
390. Deacon SE, Roach PCJ, Postis VLG, Wright GS, Xia X, Phillips SEV, et al. Reliable scale-up of membrane protein over-expression by bacterial auto-induction: from microwell plates to pilot scale fermentations. *Mol Membr Biol.* 2008;25(8):588–98.
391. Studier FW. Stable Cultures and Auto-Induction for Inducible Protein Production in *E. coli*. *Methods Mol Biol.* 2014;1091:17–32.
392. Ukkonen K, Mayer S, Vasala A, Neubauer P. Use of slow glucose feeding as supporting carbon source in lactose autoinduction medium improves the robustness of protein expression at different aeration conditions. *Protein Expr Purif.* 2013;91(2):147–54.
393. Kass EH. A brief perspective on the early history of American infectious disease epidemiology. *Yale J Biol Med.* 1987;60(4):341–8.
394. Mcsorley SJ. Immunity to intestinal pathogens: Lessons learned from *Salmonella*. *Immunol Rev.* 2014;260(1):168–82.
395. Brenner FW, Villar RG, Angulo FJ, Tauxe R, Swaminathan B. *Salmonella* nomenclature. *J Clin Microbiol.* 2000;38(7):2465–7.
396. Janeway CA, Medzhitov R. Innate immune recognition. *Annu Rev Immunol.* 2002;20(2):197–216.
397. Roland KL, Brenneman KE. *Salmonella* as a vaccine delivery vehicle. *Expert Rev Vaccines.* 2013;12(9):1033–45.
398. Hoiseth SK, Stocker B. Aromatic-dependent *Salmonella typhimurium* are non-virulent and effective as live vaccines. *Nature.* 1981;291(5812):238–9.

399. Bernstein HD. Looks can be deceiving: recent insights into the mechanism of protein secretion by the autotransporter pathway. *Mol Microbiol.* 2015;97(2):205–15.
400. Nicolay T, Vanderleyden J, Spaepen S. Autotransporter-based cell surface display in Gram-negative bacteria. *Crit Rev Microbiol.* 2015;41(1):109–23.
401. Ieva R, Skillman KM, Bernstein HD. Incorporation of a polypeptide segment into the beta-domain pore during the assembly of a bacterial autotransporter. *Mol Microbiol.* 2008;67(1):188–201.
402. Oomen CJ, van Ulsen P, Van Gelder P, Feijen M, Tommassen J, Gros P. Structure of the translocator domain of a bacterial autotransporter. *EMBO J.* 2004;23(6):1257–66.
403. Barnard TJ, Dautin N, Lukacik P, Bernstein HD, Buchanan SK. Autotransporter structure reveals intra-barrel cleavage followed by conformational changes. *Nat Struct Mol Biol.* 2007;14(12):1214–20.
404. van den Berg B. Crystal structure of a full-length autotransporter. *J Mol Biol.* 2010;396(3):627–33.
405. Tajima N, Kawai F, Park S-Y, Tame JRH. A novel intein-like autoproteolytic mechanism in autotransporter proteins. *J Mol Biol.* 2010;402(4):645–56.
406. Zhai Y, Zhang K, Huo Y, Zhu Y, Zhou Q, Lu J, et al. Autotransporter passenger domain secretion requires a hydrophobic cavity at the extracellular entrance of the  $\beta$ -domain pore. *Biochem J.* 2011;435(3):577–87.
407. Junker M, Besingi RN, Clark PL. Vectorial transport and folding of an autotransporter virulence protein during outer membrane secretion. *Mol Microbiol.* 2009;71(5):1323–32.
408. Ieva R, Bernstein HD. Interaction of an autotransporter passenger domain with BamA during its translocation across the bacterial outer membrane. *Proc Natl Acad Sci U S A.* 2009;106(45):19120–5.
409. Sauri A, Oreshkova N, Soprova Z, Jong WSP, Sani M, Peters PJ, et al. Autotransporter  $\beta$ -domains have a specific function in protein secretion beyond outer-membrane targeting. *J Mol Biol.* 2011;412(4):553–67.
410. Pavlova O, Peterson JH, Ieva R, Bernstein HD. Mechanistic link between  $\beta$  barrel assembly and the initiation of autotransporter secretion. *Proc Natl Acad Sci U S A.* 2013;110(10):E938–47.
411. Ieva R, Tian P, Peterson JH, Bernstein HD. Sequential and spatially restricted interactions of assembly factors with an autotransporter  $\beta$  domain. *Proc Natl Acad Sci U S A.* 2011;108(31):383–91.
412. Sauri A, Soprova Z, Wickström D, de Gier J-W, Van der Schors RC, Smit AB, et al. The Bam (Omp85) complex is involved in secretion of the autotransporter haemoglobin protease. *Microbiology.* 2009;155(Pt 12):3982–91.
413. Sauri A, Hagen-jongman CM, Ulsen P Van, Luirink J. Estimating the size of the active translocation pore of an autotransporter. *J Mol Biol.* 2012;416(3):335–45.
414. Kang W, Bernstein HD. Charge-dependent secretion of an intrinsically disordered protein via the autotransporter pathway. *Proc Natl Acad Sci U S A.* 2013;110(45):E4246–55.
415. Ramesh B, Sendra VG, Cirino PC, Varadarajan N. Single-cell characterization of autotransporter-mediated *Escherichia coli* surface display of disulfide bond-containing proteins. *J Biol Chem.* 2012;287(46):38580–9.
416. Van Ulsen P, Rahman SU, Jong WSP, Daleke-Schermerhorn MH, Luirink J. Type V secretion: From biogenesis to biotechnology. *Biochim Biophys Acta.* 2014;1843(8):1592–611.

417. Soprova Z, Sauri A, van Ulsen P, Tame JRH, den Blaauwen T, Jong WSP, et al. A conserved aromatic residue in the autochaperone domain of the autotransporter Hbp is critical for initiation of outer membrane translocation. *J Biol Chem.* 2010;285(49):38224–33.
418. Peterson JH, Tian P, Ieva R, Dautin N, Bernstein HD. Secretion of a bacterial virulence factor is driven by the folding of a C-terminal segment. *Proc Natl Acad Sci.* 2010;107(41):17739–44.
419. Alsteens D, Martinez N, Jamin M, Jacob-Dubuisson F. Sequential unfolding of beta helical protein by single-molecule atomic force microscopy. *PLoS One.* 2013;8(8):e73572.
420. Drobnak I, Braselmann E, Clark PL. Multiple driving forces required for efficient secretion of autotransporter virulence proteins. *J Biol Chem.* 2015;290(16):10104-16.
421. Fleetwood F, Andersson KG, Ståhl S, Löfblom J. An engineered autotransporter-based surface expression vector enables efficient display of Affibody molecules on OmpT-negative *E. coli* as well as protease-mediated secretion in OmpT-positive strains. *Microb Cell Fact.* 2014;13(179).
422. Jose J, Maas RM, Teese MG. Autodisplay of enzymes - Molecular basis and perspectives. *J Biotechnol.* 2012;161(2):92–103.
423. Gustavsson M, Do T-H, Lüthje P, Tran NT, Brauner A, Samuelson P, et al. Improved cell surface display of *Salmonella enterica* serovar *Enteritidis* antigens in *Escherichia coli*. *Microb Cell Fact.* 2015;14(47).
424. Otto BR, Van Dooren SJM, Dozois CM, Luirink J, Oudega B. *Escherichia coli* Hemoglobin protease autotransporter contributes to synergistic abscess formation and heme-dependent growth of bacteroides fragilis. *Infect Immun.* 2002;70(1):5–10.
425. Otto BR, Sijbrandi R, Luirink J, Oudega B, Heddle JG, Mizutani K, et al. Crystal structure of hemoglobin protease, a heme binding autotransporter protein from pathogenic *Escherichia coli*. *J Biol Chem.* 2005;280(17):17339–45.
426. Szabady RL, Peterson JH, Skillman KM, Bernstein HD. An unusual signal peptide facilitates late steps in the biogenesis of a bacterial autotransporter. *Proc Natl Acad Sci U S A.* 2005;102(1):221–6.
427. Jong WS, Soprova Z, de Punder K, ten Hagen-Jongman CM, Wagner S, Wickström D, et al. A structurally informed autotransporter platform for efficient heterologous protein secretion and display. *Microb Cell Fact.* 2012;11(85).
428. Jong W, Daleke-Schermerhorn MH, Vikström D, ten Hagen-Jongman CM, de Punder K, van der Wel NN, et al. An autotransporter display platform for the development of multivalent recombinant bacterial vector vaccines. *Microb Cell Fact.* 2014;13(162).
429. Langemann T, Koller VJ, Muhammad A, Kudela P, Mayr UB, Lubitz W. The bacterial ghost platform system: Production and applications. *Bioeng Bugs.* 2010;1(5):326–36.
430. Kudela P, Koller VJ, Lubitz W. Bacterial ghosts (BGs)--advanced antigen and drug delivery system. *Vaccine.* 2010;28(36):5760–7.
431. Zheng Y, Struck DK, Young R. Purification and functional characterization of phiX174 lysis protein E. *Biochemistry.* 2009;48(22):4999–5006.
432. Tanaka S, Clemons WM. Minimal requirements for inhibition of MraY by lysis protein E from bacteriophage ΦX174. *Mol Microbiol.* 2012;85(5):975–85.
433. Barisani-Asenbauer T, Montanaro J, Inic-Kanada A, Ladurner A, Stein E, Belij S, et al. *Escherichia coli* Nissle 1917 bacterial ghosts retain crucial surface properties and express chlamydial antigen: an imaging study of a delivery system for the ocular surface. *Drug Des Devel Ther.* 2015;9:3741-3754.

434. Vinod N, Oh S, Kim S, Choi CW, Kim SC, Jung C-H. Chemically induced *Salmonella enteritidis* ghosts as a novel vaccine candidate against virulent challenge in a rat model. *Vaccine*. 2014;32(26):3249–55.
435. Wen J, Yang Y, Zhao G, Tong S, Yu H, Jin X, et al. *Salmonella typhi* Ty21a bacterial ghost vector augments HIV-1 gp140 DNA vaccine-induced peripheral and mucosal antibody responses via TLR4 pathway. *Vaccine*. 2012;30(39):5733–9.
436. Eko FO, Mania-Pramanik J, Pais R, Pan Q, Okenu DMN, Johnson A, et al. *Vibrio cholerae* ghosts (VCG) exert immunomodulatory effect on dendritic cells for enhanced antigen presentation and induction of protective immunity. *BMC Immunol*. 2014;15(584).
437. Langemann T, Mayr UB, Meitz A, Lubitz W, Herwig C. Multi-parameter flow cytometry as a process analytical technology (PAT) approach for the assessment of bacterial ghost production. *Appl Microbiol Biotechnol*. Springer Berlin heidelberg. 2015:1-10.
438. Kwon SR, Kang YJ, Lee DJ, Lee EH, Nam YK, Kim SK, et al. Generation of *Vibrio anguillarum* ghost by coexpression of PhiX174 lysis E gene and staphylococcal nuclease A gene. *Mol Biotechnol*. 2009;42(2):154–9.
439. Zhu W, Zhang Y, Liu X. Efficient production of safety-enhanced *Escherichia coli* ghosts by tandem expression of PhiX174 mutant gene E and staphylococcal nuclease A gene. *Microbiol Res*. 2015;176:7–13.
440. Jawale CV, Chaudhari AA, Lee JH. Generation of a safety enhanced *Salmonella Gallinarum* ghost using antibiotic resistance free plasmid and its potential as an effective inactivated vaccine candidate against fowl typhoid. *Vaccine*. 2014;32(9):1093–9.