Dissertation

submitted to the
Combined Faculty of Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
B.Sc. Anna Sueki
Born in: Kanagawa, Japan

Oral examination: 22nd October 2019

Systematic investigation of lipoprotein cell-surface exposure in *Escherichia coli*

Referees:

Dr. Kiran Raosaheb Patil

Prof. Dr. Bernd Bukau

Abstract

Bacteria control import and export of compounds, adhesion to surfaces, cell-to-cell interactions and interactions with host via proteins they feature on their cell surface. Recent evidence has emerged that members of a category of proteins anchored to bacterial membranes through their lipidated tails, called lipoproteins, can be exposed at the cell surface of Gram-negative bacteria. *Escherichia coli* is thought to carry around 100 lipoproteins, but what fraction of them is surface exposed and whether this is a general phenomenon remains unknown. In my PhD project, I systematically investigated whether lipoproteins are surface exposed in *E. coli*. In this work, I developed a Surface Proteome Quantification (SPQ) method, which systematically and quantitatively assesses protein cell-surface exposure at a proteome-wide level, using surface biotin labelling combined with quantitative mass spectrometry. Thereby I identified at least 23 outer membrane (OM) lipoproteins to be significantly surface exposed (out of 62 quantified OM lipoproteins) to similar levels as the bona-fide surface exposed OM β-barrel proteins. Then, 5 surface identified lipoproteins, 3 OM lipoproteins and 2 inner membrane lipoproteins, were verified for their surface exposure using independent biochemical methods. Furthermore, using the SPQ method, I identified a strong dependency of the lipoprotein cell-surface exposure on the BAM (β-barrel assembly machinery) and explored the role of the different components of the BAM complex. In parallel, I generated the first proteome-wide quantitative examination of the *E. coli* membrane proteome after physical separation of the inner- and outer-membrane, revealing an additional complexity and some mis-annotations in protein localization within the bacterial cell envelope. Finally, I investigated the physiological consequences on the Rcs stress response system through a mislocalized model surface exposed lipoprotein, RcsF. Overall, the complexity and importance of protein localization in the bacterial cell envelope, with the focus on lipoproteins, is investigated and discussed in this thesis.

Zusammenfassung

Durch Proteine auf ihrer Zelloberfäche kontrollieren Bakterien den Import und Export von Stoffen, die Anheftung an Oberflächen, Zell-Zell-Kontakte, sowie Interaktionen mit einem Wirt. Neue Erkenntnisse zeigen, dass bestimmte Proteine, die durch einen Lipidrest in bakteriellen Membranen verankert werden, sogenannte Lipoproteine, auf der Zelloberfäche von gramnegativen Bakterien exponiert werden können. Es wird angenommen, dass *Escherichia coli* in etwa 100 Lipoproteine besitzt; bisher ist aber unklar, welcher Prozentsatz davon auf der Oberfläche exponiert wird und ob deren Oberflächenexposition ein generelles Phänomen ist. In meinem PhD Projekt habe ich systematisch untersucht, ob Lipoproteine in *E. coli* auf der Oberfläche exponiert werden. Dafür habe ich in dieser Arbeit eine Methode zur Quantifizierung des Oberflächenproteoms (Surface Proteome Quantification, SPQ) entwickelt, die systematisch und quantitativ das proteomweite Oberflächenexpositionslevel mittels Markierung durch Biotin und anschließender quantitativer Massenspektrometrie abschätzt. Dies führte zur Identifizierung von mindestens 23 Lipoproteinen (von insgesamt 62 quantifizierten Lipoproteinen) der äußeren Zellmembran, die signifikant auf der Oberfäche exponiert werden, in einem ähnlichen Ausmaß wie die gut charakterisierten und oberflächenexponierten beta-barrel Proteine der äußeren Zellmembran. Unabhängige biochemische Methoden bestätigten die Exposition auf der Oberfläche für 5 hier identifizierte oberflächenexponierte Lipoproteine, 3 Lipoproteine der äußeren Zellmembran und 2 Lipoproteine der inneren Zellmembran. Außerdem konnte ich mit Hilfe der SPQ Methode feststellen, dass die Oberflächenexposition von Lipoproteinen stark von der BAM-Maschinerie (beta-barrel assembly machinery) abhängt und untersuchte dazu die Rolle der unterschiedlichen Komponenten des BAM-Komplexes. Parallel dazu entwickelte ich die erste proteomweite quantitative Untersuchung von *E. coli* Membranproteinen nach Trennung der inneren und äußeren Membran, wodurch zusätzliche Komplexitäten und einige falsche Annotierungen der Proteinlokalisation von bakteriellen Proteinen in der bakteriellen Zellhülle aufgedeckt wurden. Abschließend wurden die physiologischen Folgen einer fehlerhaften Lokalisation des Lipoproteins RcsF auf die Rcs Stressantwort untersucht. Insgesamt behandelt diese Arbeit die Komplexität und Wichtigkeit der Lokalisation von Proteinen in der bakteriellen Zellhülle, mit einem Schwerpunkt auf Lipoproteine.

Acknowledgements

First and foremost, I would like to thank my supervisor, Nassos Typas. Thank you for giving me the opportunity to do my PhD with you. As a student who came with very little scientific training, you believed in my potential more than anybody, including me. Thank you for all the mentoring and care you provided throughout these four years. It was such a precious time I had in your lab. During that time, I learned and developed a lot, both academically and personally.

I would like to thank all the current and former members in our lab. Especially, Joel Selkrig, who supervised me throughout my PhD, and taught me everything at the bench. I also thank you for your endless support and patience throughout my thesis writing. I also thank Matylda Zietek, with whom I shared the passion for the Rcs system. You were always there and helped me when I needed. I'm so lucky that I've met all these amazing people in our lab. So many unforgettable memories with all of you! Many thanks to those (i.e. the entire lab) who read and helped wrap up my thesis. Also, for all the support and encouragement!

I thank the scientific support I received from outside our lab. Firstly, to the Proteomics Core Facility at EMBL Heidelberg, especially Mandy Rettel and Frank Stein, for all their support on the mass spectrometry measurements and analysis. I also thank my TAC members, Dr. Kiran Patil and Dr. Mikhail Savitski from EMBL, Prof. Bernd Bukau from Heidelberg University, and Prof. Jean-Francois Collet from UCL, Belgium, for their support throughout my PhD time.

I thank our collaborators: Kerwyn Casey Huang and Amanda Miguel from Huang Lab (Stanford University, USA), and Jean-Francois Collet, Alexandra Gennaris and Naemi Csoma from Collet Lab (UCL, Belgium), for the fruitful collaborative works we did together, and sharing their results to be included in my thesis.

Without generous gifts of reagents, especially different antibodies, many experiments would not have been possible. Thus, I thank the following labs for sharing their precious reagents and knowledge: Collet Lab (Belgium), Lithgow Lab (Australia), Vogel Lab (Germany), Pos Lab (Germany), Tokuda Lab (Japan), and Mori Lab (Japan).

Special thanks to DFG (SPP1617) and EMBL for funding my PhD study.

I thank my friends I met before my EMBL time for all the fun times and support. You know me better than I do! I thank my friends who I met at EMBL, especially the fellow PhD students, for all the fun and hard times we went through together. I thank my family in Japan, especially my grandparents for their care, and sending my favorite snacks from there all the way to Germany.

I dedicate my PhD thesis to my parents, Mama and Papa. Thank you for all support you gave me until and throughout my PhD. Even if you are on the other side of the world, in a country I never lived in, I always knew that my home is there. I'm looking forward to giving this thesis to you, and telling what I was doing over the last four years which I think is the biggest mystery for you...

Table of contents

List of abbreviations

Abbrebiation	Full name
°C	degrees Celsius
Abs	absorbance
A.U.	arbitrary unit
BAM	β-barrel assembly machinery
ECL	enhanced chemiluminescence
ELISA	enzyme-linked immunoblot assay
FPR	false positive rate
HRP	horseradish peroxidase
IM	inner membrane
IMLP	inner membrane lipoprotein
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny broth
logFC	log fold change (= $\log_2$)
LP	lipoprotein
LPS	lipopolysaccharides
MS	mass spectrometry
OD	optical density
OM	outer membrane
OMLP	outer membrane lipoprotein
OMP	outer membrane protein
PL	phospholipids
ROC	receiver operating characteristic
rpm	revolutions per minute
SDS-PAGE	sodium dodecyl sulfate - polyacrylamide gel electrophoresis
SE value	surface exposure value
SPQ	surface proteome quantification
TMT	tandem mass tag
TPR	true positive rate
WT	wildtype
β-gal	β-galactosidase

List of figures

List of tables

Chapter 1: Introduction

1.1 Studying the phenomenon of lipoprotein surface exposure

The surface of a cell is key for many tasks essential for viability. These include the sensing of the environment and the protection from it, the import and export of cargos, cell-to-cell communication, and the adherence to other bacterial cells as well as to host or abiotic surfaces. In Gram-negative bacteria, the outer-most membrane of the cell facing the extracellular environment is the outer membrane (OM). Two types of proteins are found in the OM of *Escherichia coli*: outer membrane proteins (OMPs), which form a β-barrel structure spanning through the OM, and OM lipoproteins (OMLPs), anchored to the OM by their lipidated moieties (Konovalova et al., 2017). Recent studies have demonstrated that lipoproteins can be exposed at the cell surface in Gram-negative bacteria, yet a systematic identification of which OMLPs are surface exposed and how they are translocated to the cell-surface remains poorly understood (Konovalova and Silhavy, 2015; Szewczyk and Collet, 2016). In my PhD thesis, I will address these knowledge gaps by quantitatively and systematically mapping the surface proteome of *E. coli*, with the focus on lipoproteins and interrogate mechanisms that lead to their exposure at the bacterial cell surface.

1.2 Outer membrane biogenesis

In order to study mechanistically how lipoproteins are exposed on the bacterial cell surface, it is important to understand fundamental principles of the OM biogenesis. In Gram-negative bacteria, the cell envelope is composed of two membranes, the inner membrane (IM) and the outer membrane (OM), with the soluble periplasmic space in between, where the peptidoglycan (PG) cell-wall layer resides. The OM of Gram-negative bacteria, including the model organism *E. coli* used in this study, is composed of an asymmetric lipid bilayer of two different types of lipids: phospholipids (PL) facing the periplasmic space, and lipopolysaccharide (LPS) facing the extracellular environment, and the two types of proteins, OMPs and OMLPs, are found there (Konovalova et al., 2017).

OM biogenesis has been studied extensively over the last 2-3 decades and many protein machineries have been identified. Furthermore, the molecular mechanisms of how they carry

out their functions in the cell envelope, especially at the periplasm and the OM where the environment lacks ATP, have been studied (Konovalova et al., 2017). Precursors of all the components of the OM, both proteins and membrane lipid components, are produced in the cytoplasm, translocated across the IM and then transported through the periplasmic space. For each component, there are dedicated protein machineries for the transport across the cell envelope. In the case of lipids, the Lpt complex extracts LPS from the IM, transports it across the envelope to the OM, and locates it onto the cell surface (Figure 1.1) (Sperandeo et al., 2017). The transport of PL is not fully resolved yet, though multiple protein complexes and machineries have been shown to be involved in the maintenance of PL in the OM, such as the Mla, Pqi and Yeb pathways (Figure 1.1) (Choi and Lee, 2019; Ekiert et al., 2017).

For proteins destined for the OM, both OMPs and OMLPs are transported across the IM by protein translocon machineries, either the Sec or the Tat translocons (Konovalova et al., 2017). Once they reach the periplasm, the two types of the proteins follow different paths to the OM. OMPs are transferred across the periplasm by chaperones, such as SurA and Skp. These chaperones keep their cargo proteins from prematurely folding in the periplasm, and deliver it to the OM β-barrel assembly machinery (BAM) which folds and inserts the OMPs into the OM (Figure 1.1) (Konovalova et al., 2017). In contrast, once translocated across the IM, the N-terminal cysteine of a lipoprotein precursor is lipid-modified, i.e. lipidated, and the mature lipoprotein is anchored to the IM via its lipidated tail (Konovalova et al., 2017). Lipoproteins with an OM destination are transported by the Lol system through the periplasmic space (Okuda and Tokuda, 2011; Tokuda and Matsuyama, 2004). The recruitment by the Lol system is determined by the N-terminal sequence; if a "Lol avoidance signal" (an aspartate in +2 position) is found, the lipoprotein does not interact with the Lol system and is thus retained at the IM. Otherwise, it is transported to the OM, which is the case for most of the *E. coli* lipoproteins (Okuda and Tokuda, 2011). The IM LolCDE complex uses cytoplasmic ATP to extract lipoproteins out of the IM and passes them onto their periplasmic partner protein LolA, which transports the lipoprotein cargo across the periplasmic space (Figure 1.1) (Okuda and Tokuda, 2011). Finally, LolB, an OMLP by itself, interacts with LolA and anchors the lipoprotein to the OM facing the periplasm (Figure 1.1) (Okuda and Tokuda, 2011).

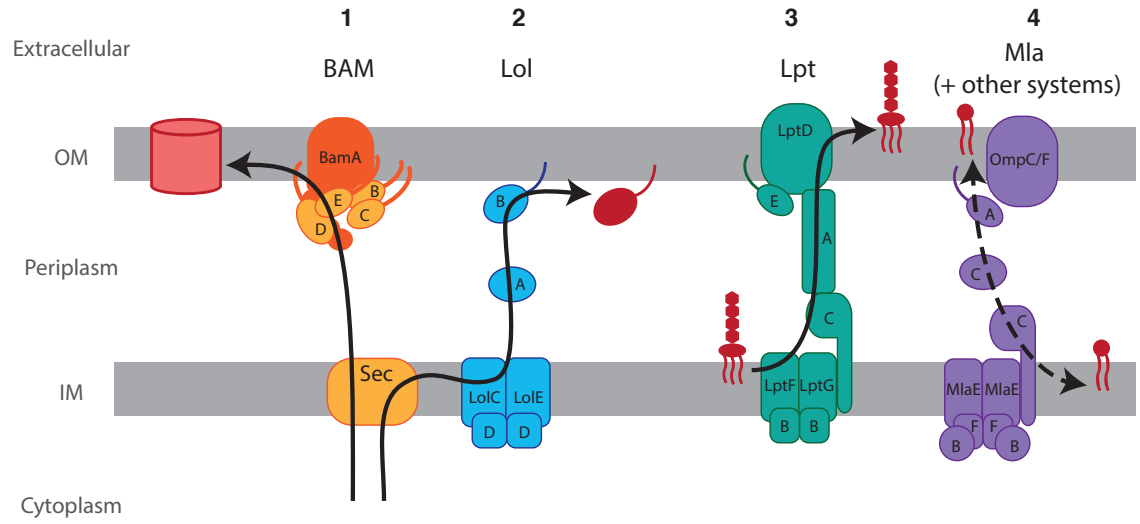


Figure 1.1: OM biogenesis machineries transporting OM components across the cell envelope. (1) The BAM complex (BamA-E) in the OM folds and inserts β-barrel proteins, which have been translocated across the IM by Sec translocons, into the OM. (2) Lol machinery. LolCDE complex in the IM takes up lipoproteins, which have been translocated across the IM by Sec translocons and lipidated. LolA transports them across the periplasm, and LolB anchors the lipoproteins to the OM, facing the periplasm. (3) The Lpt machinery is a transmembrane complex that transports LPS from the IM to the cell surface: LptBFG is in the IM, LptA spans the periplasm, and the LptDE complex inserts LPS into the outer-leaflet of the OM, facing the extracellular environment. (4) The Mla machinery and other systems transport and maintain phospholipids in the OM. The Mla complex is depicted here: The MlaBCEF complex is in the IM, MlaC in the periplasm, and MlaA is anchored to the OM where it interacts and functions together with OmpC/OmpF. Figure adapted from (Choi and Lee, 2019).

1.3 Functions of lipoproteins

Lipoproteins show diverse functions and structures in different species of Gram-negative bacteria. In *E. coli*, more than 100 proteins are predicted to be lipoproteins (numbers vary in different literature and databases), and 90 have been validated experimentally (Braun and Hantke, 2019). Approximately 95% of the lipoproteins are thought to be transported to the OM via the Lol system (Szewczyk and Collet, 2016). Some lipoproteins form parts of essential OM protein machineries, such as the BAM machinery, which consists of a core OMP called BamA and four OMLPs (BamB, C, D and E) (Figure 1.1) (Konovalova et al., 2017). Another example is the OM component of the LPS transporter LptD, an OMP that forms a stable protein complex with the OMLP LptE (Figure 1.1) (Konovalova et al., 2017). Furthermore, the most abundant protein in *E. coli,* Lpp (also known as Braun's lipoprotein), is an OMLP, and plays an important role in cell envelope integrity (Asmar and Collet, 2018; Braun and Hantke, 2019). A recent review, summarizing the current understanding of lipoproteins in *E. coli* and other organisms states that 38 lipoproteins (out of 116 predicted lipoproteins) remain unannotated for any functionality (Braun and Hantke, 2019). Thus, even in *E. coli,* which has the best studied and

understood genome among all Gram-negative bacteria, lipoproteins still hold many unexplained details.

OMLPs also play an important role in sensing the external environment. As the cell envelope is the outer-most compartment of the bacteria, many surveillance systems are found in there, to sense and adapt to changes in the extra- and intra-cellular environment. One example are two-component systems (TCSs) which transduce stimuli through a phosphorylation cascade (Hoch, 2000). Although the central components of the TCSs are histidine kinases in the IM and their partner response regulators in the cytoplasm, some TCSs have dedicated periplasmic / OM partner proteins, such as in the Rcs system (Majdalani et al., 2005; Wall et al., 2018). The Rcs is an envelope stress response TCS that is activated upon diverse perturbations in the cell envelope. RcsF, an OMLP, acts as the sensor protein in this system crucial for the phosphorelay cascade activation (Majdalani and Gottesman, 2005; Wall et al., 2018). Similarly, NlpE and QseG are OMLPs that are the sensor and activator of the respective TCSs - Cpx and QseE/F (Delhaye et al., 2019; Göpel and Görke, 2018). In all these systems, the localization of lipoproteins was shown to be crucial for their function, including the surface localization of RcsF (Cho et al., 2014; Delhaye et al., 2016; Göpel and Görke, 2018; Konovalova et al., 2014).

OMLPs also play important roles in bacterial cell biology that range from immune evasion to maintaining cell-envelope integrity. In general, many lipoproteins are known to be recognized by the host immune system through Toll-like receptors (Schenk et al., 2009). For example, the surface lipoprotein OspA in *Borrelia burgdoferi* had been used to develop a vaccine against Lyme disease, as the antibody against OspA was shown to protect animals from infection (Kenedy et al., 2012). Furthermore, essential protein machineries for the OM biogenesis, which often involve OMLP components, have been studied as potential targets of antibiotics (Choi and Lee, 2019). For example, a compound called MAC13243 was shown to have antibacterial capacity through targeting of the Lol pathway, the essential OM lipoprotein transport system (Choi and Lee, 2019; Pathania et al., 2009). Recently, BamA, the central component of the conserved OMP folding machinery in Gram-negative bacteria, has been explored as a candidate antibody target for antibiotic treatment (Storek et al., 2018). Thus, a deeper understanding of lipoproteins, especially their transport and their localization at the cell-surface, will provide information with clinical relevance, including vaccine candidates and drug targets.

1.4 Lipoprotein exposure on the bacterial cell surface

The surface of Gram-negative bacteria is heavily populated by LPS. At the same time, some proteins are known to be accessible from the extracellular environment, such as OMPs exhibiting β-barrel structures spanning the OM, whereby extracellular loops are exposed at the cell surface due to multiple anti-parallel transmembrane β-strands passing in and out of the OM (Galdiero et al., 2007; Konovalova et al., 2017). Contrarily, almost all lipoproteins have previously been thought to be anchored to the OM facing the periplasm (Okuda and Tokuda, 2011). However, multiple recent studies in diverse organisms have revealed that lipoproteins in Gram-negative bacteria can be surface exposed (Konovalova and Silhavy, 2015; Szewczyk and Collet, 2016).

In *E. coli*, to our knowledge, there are 8 lipoproteins (out of around 100 lipoproteins in total) reported to be surface exposed, also summarized in recent reviews (Konovalova and Silhavy, 2015; Szewczyk and Collet, 2016). These surface identified *E. coli* lipoproteins will be discussed in detail in the next section. In other Gram-negative or related bacteria, studies of surface lipoproteins have been carried out to an even larger extent. *Borrelia*, a spirochete bacteria lacking surface LPS, was shown to display most of its lipoproteins on the surface (Dowdell et al., 2017; Schulze and Zückert, 2006). In *Bacteroides fragilis*, a commensal gut microbe in humans, the surface proteome has been investigated by a combination of surface biotin labelling and protease accessibility, also identifying a large fraction of lipoproteins to be surface exposed (Wilson et al., 2015).

The mechanisms behind lipoprotein surface transport remain mostly unknown. Recently, in *Neisseria meningitidis*, a dedicated machinery called "Slam" has been identified, which exposes specific lipoproteins on the surface (Hooda et al., 2016). This Slam protein in *Neisseria* is specific for certain cargo, and homologs of both Slam and its cargo do not exist in many other bacteria including *E. coli* (Hooda et al., 2016). Furthermore, in *Bacteroidetes* (originally in *Capnocytophaga canimorsus*), a lipoprotein export signal (LES) was identified, which targets lipoproteins to the surface (Lauber et al., 2016). However, Slam is not conserved in *Bacteroidetes*, which suggests that a different mechanism of lipoprotein surface transport exists in these bacteria (Lauber et al., 2016). Besides the mentioned organisms, reports of lipoprotein surface localization are found and discussed in other Gram-negative bacteria, including *Brucella* (Goolab et al., 2015) and *Vibrio cholerae* (Pride et al., 2013).

Little is known regarding the function of surface lipoproteins. In *Bacteroides*, some surface lipoproteins interact with the extracellular environment for nutrient uptake, such as SusD for sugar (Glenwright et al., 2017) and BtuG for a vitamin uptake (Wexler et al., 2018). Furthermore, links between lipoprotein surface exposure and outer membrane vesicle (OMV) secretion have recently been shown also in *Bacteroides* (Valguarnera et al., 2018). There, OMVs are suggested as a possible mechanism of providing "public goods" (Valguarnera et al., 2018). These studies suggest that there are uncovered functionalities of lipoprotein surface exposure, for individual lipoproteins and also as a more general mechanism, in other bacteria as well. We are specifically interested in *E. coli,* as it is the model organism of Gram-negative bacteria, with the largest knowledge and available manipulation techniques of its genome. The elucidation of lipoprotein surface transport mechanisms in *E. coli* will open routes to the investigation of lipoprotein surface exposure in unexplored organisms as well.

1.5 Surface exposed lipoproteins in *E. coli*

The 8 lipoproteins reported to be surface exposed in *E. coli* are: TraT, Wza, CsgG, Lpp, Pal, RcsF, YaiW and BamC, discussed in details in recent reviews (Konovalova and Silhavy, 2015; Szewczyk and Collet, 2016). In 1980, TraT, a protein encoded on the F-plasmid, was shown to be surface exposed and involved in surface exclusion, however, it is not found in the lab strain *E. coli* K-12 genome (Manning et al., 1980). Then Lpp, BamC, YaiW and Pal were shown to be surface exposed each by independent studies (Arnold et al., 2014; Cowles et al., 2011; Surendran et al., 2015; Webb et al., 2012). In 2014, RcsF, the OMLP involved in the Rcs stress response system, was shown by our group and others to be surface exposed (Cho et al., 2014; Konovalova et al., 2014). In these studies, BamA was required for RcsF surface exposure, demonstrating for the first time that the BAM machinery is involved in lipoprotein surface transport in *E. coli* (Cho et al., 2014; Konovalova et al., 2014). Earlier, two other lipoproteins were identified to be surface exposed, Wza (Drummelsmith, 2000) and CsgG (Robinson et al., 2006), and they have been shown to form barrel-like structure upon oligomerization, with α-helices for Wza and β-strands for CsgG (Dunstan et al., 2015). Interestingly, surface exposure of these two lipoproteins was shown to be BAM independent, and suggested to occur via spontaneous insertion into the OM (Dunstan et al., 2015). These identified surface lipoproteins and their suggested models of surface exposure are summarized in Figure 1.2. Besides these

studies for individual lipoproteins, no systems-wide studies have been performed in *E. coli* aimed at identifying surface lipoproteins to survey the prevalence of surface exposed lipoproteins and the mechanisms behind the phenomenon.

The functional role of lipoproteins at the cell surface has been studied for some surface identified lipoproteins. For example, Wza and CsgG are assumed to be functional only on the surface as they form secretion channels across the OM (Figure 1.2) (Dunstan et al., 2015; Robinson et al., 2006). In contrast, for RcsF, the surface translocation efficiency acts as the sensor for the downstream cell envelope stress response signaling system, and its default localization, without cell envelope stress, is at the cell surface plugged inside the lumen of OMPs (Figure 1.2) (Cho et al., 2014; Konovalova et al., 2014). More recently, an additional role of RcsF has been shown: after its translocation to the surface, it monitors LPS integrity, suggesting that surface RcsF can sense envelope stress, and thus also play a functional role at the cell surface (Konovalova et al., 2016). On the other hand, Lpp, Pal and BamC have their known interacting partners on the periplasmic side of the OM (Figure 1.2): Lpp covalently binds to the peptidoglycan cell wall (Braun, 2018), Pal interacts with IM partners TolA and TolB (Bouveret et al., 1995; Cascales et al., 2000), and BamC interacts with rest of the BAM complex that is mostly located on the periplasmic side of the OM (Noinaj et al., 2017). However, the function of these lipoproteins at the cell-surface, as well as the mechanisms of their surface translocation remain unclear.

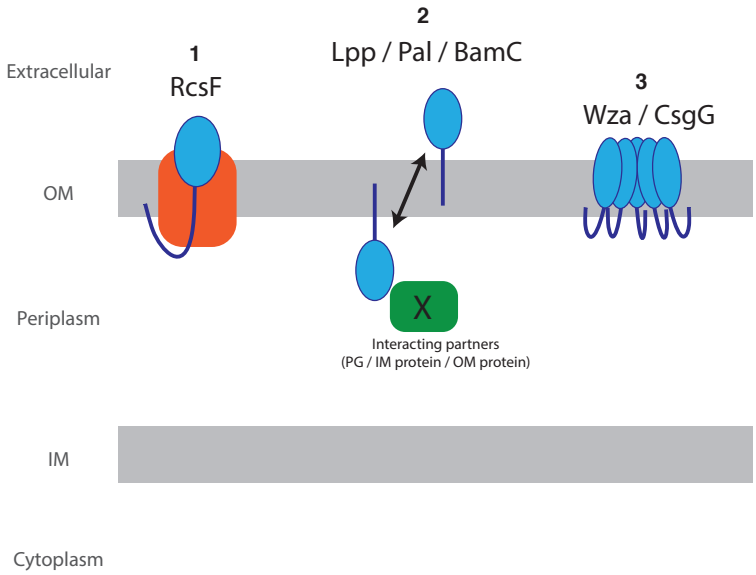


Figure 1.2: Surface exposed lipoproteins in *E. coli*. The model of their surface exposure, adapted from (Grabowicz, 2018). Lipoproteins described are shown in blue: (1) RcsF is embedded in the lumen of an OMP. (2) Lpp, Pal and BamC have dual localization facing either the periplasm or the cell surface. All of them have known interacting partners in the periplasm (Lpp with peptidoglycan (PG), Pal with TolA and TolB in the IM, and BamC with other BAM complex members in the OM). (3) Wza and CsgG forming pore-like structures upon oligomerization.

1.6 Research aims

Although several lipoproteins have been reported to be surface exposed in *E. coli*, we still lack systematic and quantitative studies of lipoprotein surface exposure. To date, out of around 100 putative lipoproteins, only 8 have been confirmed to be surface-exposed in *E. coli* (Konovalova and Silhavy, 2015; Szewczyk and Collet, 2016). Recent identification of the Slam transporter in *Neisseria* (Hooda et al., 2016) and the LES signal sequence identification for lipoprotein surface transport in *Bacteroides* (Lauber et al., 2016) further suggest a common mechanism of lipoprotein surface transport within *E. coli* as well. In addition to their localization on the surface, experimental evidence for membrane localization of lipoproteins, either in the IM or OM, is also limited. In my PhD study, I investigated lipoprotein localization in bacteria, with a focus on their surface exposure, using proteome-wide systematic and biochemical approaches. In particular, I set out to achieve the following aims (Figure 1.3):

1. Develop a quantitative proteomics-based surface labelling method to globally assess lipoprotein surface exposure in *E. coli*
2. Investigate the mechanism behind lipoprotein surface exposure by using the surface labelling method
3. Examine membrane localization of lipoproteins and other proteins quantitatively and systematically in *E. coli*
4. Explore the effect of lipoprotein mislocalization in a stress response system using a model surface exposed lipoprotein, RcsF

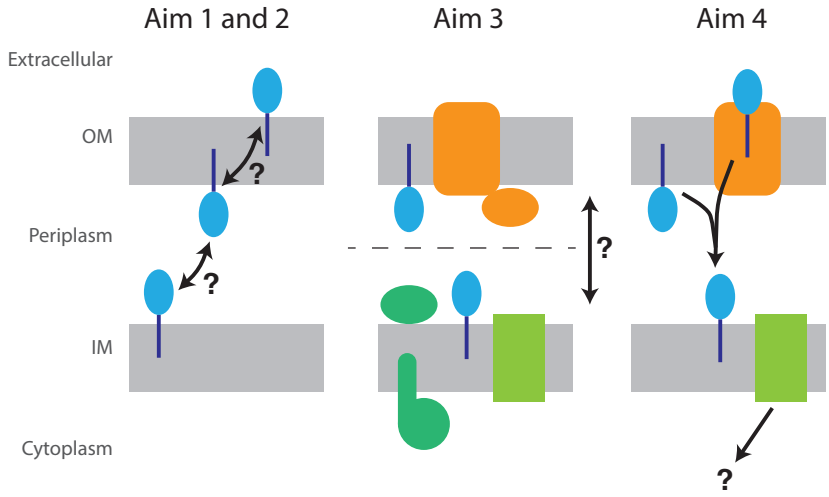


Figure 1.3: Illustration of the four research aims. Aim1 and 2 – identifying the lipoprotein localization between three possible localizations (IM, OM facing periplasm, OM facing surface), Aim 3 – Identifying protein localization between two membranes, Aim 4 – the downstream consequence of lipoprotein mislocalization from the OM to the IM.

Chapter 2: Materials and methods

2.1 Strains and cloning

2.1.1 Strains and plasmids

All bacterial strains used are listed in Table 1. The wildtype (WT) background strain used in this study is *E. coli* K-12 MG1655 carrying an *rprA::lacZ* fusion at the λ attachment site (Majdalani et al., 2002). All chromosomal gene deletions and introduction of mutations were carried out by P1 phage transduction as described in section 2.1.2.

Plasmids used are listed in Table 2. Plasmid isolation was performed using the QIAprep Spin Miniprep Kit (Qiagen, Table 4), following its product instruction. All plasmids were chemically transformed into *E. coli* using the TSS plasmid transformation method as described in section 2.1.3.

2.1.2 P1 transduction

P1 phage lysates were prepared and used to transduce mutations as follows. To prepare donor P1 phage lysates, 1 drop of 1M $CaCl_2$ and 2 drops of P1 lysate were transferred to a mid-log culture of the donor strain in Lysogeny broth (LB) medium, using a glass Pasteur pipette. Donor strains carrying the desired mutations were incubated at 37 degrees Celsius (°C) for 3-6 hours in the roller, until the cell lysis was observed, as assessed by the increased culture opacity and formation of cellular debris. 10 drops of $CHCl_3$ were added and vortexed, then was centrifuged using a Centrifuge 5810 R (Eppendorf), at 4000 revolutions per minute (rpm), with an A-4-81 rotor, for 10 minutes at room temperature to remove non-lysed cells and cell debris. The supernatant was collected and stored at 4°C. Gene deletion mutants were transduced into recipient strains by preparing P1 lysates from strains from the KEIO deletion library (Baba et al., 2006) or from the ASKA deletion library (unpublished library from H. Mori lab). *bamA101* and *bamA6* (formally referred as *yaeT6*) mutants used were previously published in (Aoki et al., 2008; Ruiz et al., 2006).

Phage transductions were carried out by harvesting 5 mL of overnight culture of recipient strains grown in LB medium, by centrifuging at 5000 *xg*, and resuspended in 2.5 mL of resuspending buffer (Table 5). 2 drops of P1 lysate carrying the desired mutation were added to 200 μL of cell suspension. After 5 min incubation at room temperature, 0.5 mL of 2x YT

media (Table 5) were added, and the mixture was incubated at 37°C for 20 minutes in the roller. 1 drop of 1M sodium citrate was added, and the mixture was plated onto a LB-agar plate (Table 5) containing 10 mM sodium citrate and an antibiotic for selection (Kanamycin – 30 μg/mL, Chloramphenicol – 10 μg/mL). Plates were then dried and incubated at 37°C overnight, and the resulting colonies were re-struck onto a fresh LB plate with the appropriate antibiotic selection in order to remove any residual phage particles.

2.1.3 TSS plasmid transformation

The protocol was adapted from a previous study as described below (Chung et al., 1989). 200 μL of a mid-log phase culture was mixed with 200 μL TSS (Table 5) and 1 μL of purified plasmid. The sample was incubated on ice for 30 minutes, and then incubated at 37°C in the roller (or 30°C for temperature sensitive plasmid pKD46) for 45 minutes. Then the mixture was spread onto a LB-agar plate with appropriate antibiotics (Ampicillin – 100 μg/mL, Spectinomycin – 100 μg/mL), dried, and incubated overnight at 37°C (or at 30°C for pDK46 transformations).

2.1.4 C-terminal Flag tagging using pKD46

The method was adapted from (Datsenko and Wanner, 2000) as described below. Flag and Kanamycin resistance cassette sequences was amplified from pJPS1 (Table 2) with forward primer (H1P1) and reverse primer (H2P2), listed in Table 6. H1P1 consists of 50 nucleotides (n.t.) before the stop-codon of the target gene (H1), two glycine residues, and an overlapping region with pJPS1, which contains the flag sequence (coding for DYKDDDK), and the following 20 n.t. upstream of the Kanamycin resistance cassette (P2). H2P2 consists of 50 n.t. after the stop codon of the target gene (H2) and 20 n.t. overlapping with pJPS1 at the end of Kanamycin cassette (P2). Using these primers, DNA fragments were amplified using Q5 Hot Start High-Fidelity DNA Polymerase (NEB, Table 4), and the fragment were separated by agarose-gel electrophoresis. The band(s) corresponding to the expected size were cut out and the DNA fragments were extracted and purified using the QIAquick Gel Extraction Kit (Qiagen, Table 4). Purified DNA was electroporated into electrocompetent cells expressing the λ Red system as described below.

2.1.5 Preparation of electrocompetent cells and DNA fragment electroporation

Overnight cultures of wildtype strain with pKD46 were diluted by 1:100 dilution in LB medium containing 100 μg/mL Ampicillin. After adding arabinose (0.02%, in order to express the λ Red

genes from pKD46), cultures were incubated for 3 hours at 30°C. Cells were cooled by incubating them on ice for 20 minutes, and all following steps were performed at 4°C or on ice. Centrifugation steps were performed using a F4-45-30-11 rotor with Centrifuge 5430 R (Eppendorf), at the indicated speed. Cells were centrifuged at 5000 rpm for 10 minutes, and resuspended to the original volume with H_2O. Cells were centrifuged again at 5000 rpm for 10 minutes, then resuspended to half of the original volume with 15% glycerol (in H_2O). Cells were again centrifuged with 5000 rpm for 10 minutes, and resuspended to 0.1% of the original volume with 15% glycerol. Finally, cells were centrifuged at 14,000 rpm for 5 minutes, and resuspended to 0.05% of the original volume with 15% glycerol. For electroporation, 50 µL of electrocompetent cells prepared were mixed with 2 µL of purified DNA (>100 ng/µL) and transferred to an electroporating cuvette (Table 4). The mixture was electroporated at 2.5 kV, using Gene Pulser Xcell Electroporation Systems (Bio-rad), and directly after electroporation, 1 mL of LB was added, and the mixture was incubated at 30°C for 1.5 hours. The culture was spread on a LB-agar plate containing appropriate antibiotics (Kanamycin – 30 µg/mL), and the plate was incubated overnight at 30°C. Colonies were re-struck onto a fresh plate with the appropriate antibiotic, and incubated overnight at 37°C in order to remove the temperature sensitive pKD46 plasmid. Insertion of the DNA fragment was confirmed by PCR, and the insert was moved to a fresh WT background using P1 transduction as previously described.

2.2 Surface and lysate biotinylation

2.2.1 Bacteria culturing

Bacterial cells were grown in LB-Lennox (referred as LB) medium at 37°C with vigorous shaking (200 rpm in incubation shaker by Infors shaker for culture volume above 5 mL in conical flasks with their size being more than 2x the culture volume, or with constant rotation in a roller within a glass vial carrying 5 mL of LB medium). Overnight cultures were prepared by inoculating 5 mL of LB medium from a single colony struck out onto an LB-agar plate from glycerol stocks (stored at -80°C). Glycerol stocks were prepared by mixing 1 mL of an overnight culture of a strain in LB with 0.5 mL 50% sterile glycerol.

2.2.2 Biotinylation with NHS-LC-LC-Biotin

The methodology for surface biotinylation of bacterial cells was adapted from a previous study (Cowles et al., 2011) and is detailed below as follows. Overnight cultures of bacterial strains tested for surface biotinylation in LB were diluted 1:100 into a fresh LB and were grown until they reached an OD_{578} of 0.6 – 0.8 (exponential phase samples). Ampicillin at 100 µg/mL was

added in the medium for strains with plasmids pJV300 and pFM1-1. For stationary phase samples, cells were grown until OD_{578} of 2.3 – 3.0. All subsequent steps were carried out on ice or at 4°C unless otherwise specified. To normalize for differences in cell density between samples, a sample of the original culture proportional to the OD_{578} was calculated according the following formula; V (in mL) = 286 / OD_{578}, which was adapted from a previous study that used a similar cell lysis method (Anwari et al., 2010). The normalized cultures were then harvested (for intact cell and lysate biotinylation) by centrifugation at 4000 rpm with a JLA8.1000 rotor, using an Avanti J-20XP centrifuge (Beckman Coulter) for 20 minutes, and washed 3 times with 40 mL PBS using a Centrifuge 5810 R (Eppendorf), at 4000 rpm with an A-4-81 rotor using 50 mL Falcon tubes.

Intact cell (surface) biotinylation: Washed cells were resuspended in 4 mL of PBS and incubated at room temperature for 5 minutes with gentle rocking. Then, 80 µL of freshly prepared 25 mg/mL stock (solubilized in DMSO) of NHS-LC-LC-Biotin (Thermo) were added to the sample, and incubated for 15 minutes with gentle rocking at room temperature. The reaction was stopped by adding 1.33 mL of 1M Tris-HCl, pH = 7.5 to a final concentration of 0.25 M. Cells were washed twice with cold PBS, and then proceeded with cell lysis (everything from here on at 4°C).

Lysate biotinylation: cellular lysates prepared (as described below in section 2.2.3, lysate volume around 60 mL) were adjusted to room temperature for 5 minutes and then incubated with 80 µL of 25 mg/mL stock of NHS-LC-LC-Biotin (Thermo) (i.e. the same amount of biotin as used for the intact cell biotinylation), and quenched with 1.33 mL of 1M Tris-HCl, pH = 7.5 (i.e. the same volume of 1M Tris-HCl as intact cell biotinylation). As the volume for the lysate sample is larger than the intact cell biotinylation sample, a conical flask was used, which was shaken at 100 rpm in an incubation shaker (Infors) instead of rocking. Samples were then processed to enrich for membrane fractions as follows, which also effectively removes unreacted NHS-LC-LC-Biotin from the sample that would otherwise interfere with downstream pull-down step.

2.2.3 Cell lysis and membrane preparation

The protocol was adapted from (Anwari et al., 2010), with the following modifications. Importantly, PBS was used as the base buffer instead of Tris, in order to avoid quenching the biotin cross-linker during the biotinylation experiments. Intact cell and lysate biotinylated

samples were treated identically, and the procedure was performed either on ice or at 4°C to prevent protein degradation, unless otherwise stated.

Cells were resuspended in 20 mL of spheroplasting buffer (Table 5) containing 1x protease inhibitor (Roche, Table 4), followed by the addition of a final concentration of 50 µg/mL lysozyme (freshly prepared, Table 4), and incubated for ~10 minutes on ice. 40 mL of lysis buffer (Table 5) was added drop-wise to the sample while mixing. Cells were lysed by passing through a high-pressure homogenizer (Emulsiflex C5 by Avestin) for a minimum of 3 times at >20K psi until the sample was no longer turbid. Cell debris were removed by centrifuging at 4000 rpm for 10 minutes with a Centrifuge 5810 R (Eppendorf, rotor A-4-81). Lysate biotinylation samples were biotinylated at this point as detailed above in section 2.2.2. Lysates were ultracentrifuged at 100,000 *xg* for 1.5 hours using a Ti45 rotor with Optima L-100 XP Ultracentrifuge (Beckman). The membrane pellet was resuspended in the same buffer (mix buffer, 1:2 mixture of spheroplasting and lysis buffer) using a dounce homogenizer, and then ultracentrifuged again at 100,000 *xg* for 1.5 hours using a Ti45 rotor with Optima L-100 XP Ultracentrifuge (Beckman), in order to remove any residual soluble proteins and biotin reagent.

The membrane pellet was resuspended in 4 mL of the mix buffer (1:2 mixture of spheroplasting and lysis buffer), and solubilized over-night by incubating with Dodecyl-β-D-maltosid (DDM, stock concentration 20 %) added to be final concentration of 1 % DDM , and incubated at 4°C rotating on a tube rotator. Non-solubilized membranes were removed by centrifuging for 1 hour at 100,000 *xg* using a TLA100.4 rotor in an Optima TLX ultracentrifuge (Beckman). Supernatants were collected as the membrane-solubilized sample, and was then used for Strep-Tactin Sepharose pull-down experiments.

2.2.4 Biotinylated protein pull-down

The batch purification method was used with the following specification. All centrifuging steps were performed using the Centrifuge 5430 R (Eppendorf) at 5000 *xg* unless otherwise specified. 100 µL of Strep-Tactin Sepharose beads (by IBA, Table 4) were equilibrated by washing once with PBS, and incubated with 1 mL of the membrane-solubilized sample for 4 hours at 4°C. Beads were washed once using PBS and then three times using PBS with 1M NaCl, followed by two times washing with PBS without NaCl (all wash buffers contained 0.01% DDM). Samples not washed with 1M NaCl (referred as “no salt” samples) were washed 5 times with PBS. Washed beads were incubated for 30 minutes at 4°C with 100 µL elution buffer (Buffer

E by IBA, Table 4, with 1x protease inhibitor and DDM to a 1% final concentration was added). The eluted biotinylated proteins were collected by centrifugation at 8200 *xg*, removing the supernatant, and this elution process was repeated for a total of two eluted samples which were then pooled into a final elution volume of 200 µL.

2.3 LC-MS/MS

2.3.1 Sample preparation and TMT labelling

Full MS sample list is included in Table 7. The mass spectroscopy (MS) measurements and data analysis were performed by EMBL Proteomics Core Facility. Specifically, sample handling was done by Mandy Rettel, as described below.

Firstly, the disulphide bonds were reduced by incubating at 56°C for 30 minutes in a buffer containing 10 mM dithiothreitol (DTT) in 50 mM HEPES buffer pH = 8.5. Reduced cysteines were alkylated by incubating 30 minutes at room temperature in dark with 20 mM 2-chloroacetamide in 50 mM HEPES buffer pH = 8.5. Samples were prepared for MS measurements using the SP3 protocol (Hughes et al., 2014, 2019), and trypsin (sequencing grade, Promega, Table 4) was added to digest proteins in an enzyme:protein ratio of 1:50 for overnight at 37 °C. Then digested peptides were recovered in HEPES buffer by collecting the supernatant on magnet, and combining the second elution wash of beads with HEPES buffer. Collected peptides were labelled with TMT10plex Isobaric Label Reagent (ThermoFisher, (Werner et al., 2014)) according to the manufacturer's instructions as described below. In brief, 0.8 mg of the TMT reagents were dissolved in 42 µL of 100 % acetonitrile and 4 µL of this stock was added to the peptide sample and incubated for 1 hour at room temperature. The reaction was quenched with 5% hydroxylamine for 15 minutes at room temperature. Then the 10 samples labelled by different TMT10plex labels were combined. The combined sample was cleaned up using OASIS® HLB µElution Plater (Waters).

The samples were separated through an offline high pH reverse phase fractionation on an Agilent 1200 Infinity high-performance liquid chromatography system which was equipped with a Germini C18 column (3 µm, 110 Å, 100 x 1.0 mm, Phenomenex). The fractionation was done as previously described (Reichel et al., 2016). For three MS runs (P0413, P0444, and P0483 in Table 7), samples were pooled in total 5 fractions after high pH fractionation. For one MS run (P0919 in Table 7), samples were pooled in total 12 fractions.

2.3.2 Mass spectrometry data acquisition

An UltiMate 3000 RSLC nano LC system (Dionex) was fitted with a trapping cartridge (µ-Precolumn C18 PepMap 100, 5µm, 300 µm i.d. x 5 mm, 100 Å) and an analytical column (nanoEase™ M/Z HSS T3 column 75 µm x 250 mm C18, 1.8 µm, 100 Å, Waters). Trapping was carried out with a constant flow of Solvent A (0.1% formic acid in H_2O, flow rate of 30 µL per minute) onto the trapping column for 6 minutes. Then peptides were eluted via the analytical column (with a constant flow of 0.3 µL per minute) with increasing percentage of Solvent B (0.1 % formic acid in acetonitrile) from 2 % to 4 % in 4 minutes, from 4 % to 8 % in 2 minutes, then 8 % to 28 % for a further 96 minutes (for P0413, P0444, and P0483 in Table 7) or 66 minutes (P0919 in Table 7), and finally from 28% to 40% for another 10 minutes. The outlet of the analytical column was coupled directly to a QExactive plus (Thermo) mass spectrometer by the Proxeon nanoflow source in positive ion mode.

The peptides were introduced into the QExactive plus via a Pico-Tip Emitter 360 µm OD x 20 µm ID; 10 µm tip (New Objective) and an applied spray voltage of 2.3 kV. The capillary temperature was set at 320°C. Then full mass scan was acquired, with mass range 375 - 1200 m/z in profile mode with resolution of 70000. The filling time was set at maximum of 10 ms with a limitation of $3x10^6$ ions. Data dependent acquisition (DDA) was performed with the resolution of the Orbitrap set to 35000, with a fill time of 120 ms and a limitation of $2x10^5$ ions. A normalized collision energy of 32 was applied. Dynamic exclusion time of 30 second was used. The peptide match algorithm was set to ‘preferred’ and charge exclusion ‘unassigned’, charge states 1, 5 - 8 were excluded. MS^2 data was acquired in profile mode.

2.3.3 MS data analysis

IsobarQuant (Franken et al., 2015) and Mascot (v2.2.07, (Perkins et al., 1999)) were used to process the acquired data, which was then searched against a Uniprot *Escherichia coli* proteome database containing common contaminants and reversed sequences (The UniProt Consortium, 2019). The following modifications were included into the search parameters: Carbamidomethyl (C) (fixed modification), Acetyl (Protein N-term), Oxidation (M), Biotin_thermo21343 (K, +452.61Da), TMT10 (K) and TMT10 (N-term) (variable modifications).

For the full scan (MS1) a mass error tolerance of 10 ppm, and for MS/MS (MS2) spectra of 0.02 Da was set. Further parameters were set: Trypsin as protease with an allowance of

maximum two missed cleavages; a minimum peptide length of seven amino acids; at least two unique peptides were required for a protein identification. The false discovery rate on peptide and protein level was set to 0.01.

2.3.4 Statistical analysis of MS data

Statistical analysis and data visualization of MS data were performed with the support from Frank Stein (EMBL Proteomics Core Facility).

The protein.txt output files of IsobarQuant (Franken et al., 2015) were processed using the R programming language (R Core Team, 2019). As a quality criterion, only proteins which were quantified with at least two unique peptides were used. Raw tmt reporter ion signals (signal_sum columns) were first batch-cleaned using the removeBatchEffect function from limma (Ritchie et al., 2015) and then normalized using the vsn package (Huber et al., 2002). Whenever value imputation was carried out, k-nearest neighbors method was used. If the experiment contained multiple replicates, limma was used again to test for differential abundance. In case of only a single replicate, ratios were calculated and transformed into a z-distribution. Probabilities were estimated by assuming a t-distribution. P-values were corrected for multiple testing using the Benjamini and Hochberg method (Benjamini and Hochberg, 1995). For clustering, k-means clustering was used for sucrose gradient fractionation samples (discussed in Chapter 5).

2.4 Protein surface localization validation

2.4.1 Dot blot

The method was adapted from a previous study (Cho et al., 2014) as described as below. Over-night cultures in LB were diluted 1:100 into a fresh LB media, and grown until mid-log phase (OD_{578} = 0.6-0.8). 1 mL of culture was then washed with PBS, and resuspended to OD = 1 in PBS for intact cell sample, or in permeabilizing buffer (Table 5) for permeabilized sample. For permeabilization, the cells were incubated in permeabilizing buffer for 30 minutes at 4°C before blotting. Samples were then spotted using a pipette onto a nitrocellulose membrane (Table 4) in triplicates of 2 µL for each sample and allowed to air dry. The dried membrane was incubated with 5% skim-milk in TBS for 30 minutes, washed three times for 5 minutes each with TBST (containing 0.05% Tween-20, and for all subsequent washing steps). Washed membranes were then incubated with the primary antibody diluted in 1% skim-milk in TBST (α-Flag: 1:1000, α-AcrA: 1:200,000) for 1 hour in TBST with 1% skim-milk. After washing three times for 5

minutes each with TBST, the membrane was incubated for 1 hour with secondary antibody conjugated to HRP in TBST with 1% skim-milk. After washing three times for 5 minutes each with TBST, the membrane was incubated with ECL substrate (Thermo, Table 4), and exposed to X-ray films (Table 4). Exposed films were developed using a film developing machine (Kodak RP X-Omat Processor, Model M6B), and were scanned by a printer to digitalize. AcrA antibody used as a negative control for intact cells.

2.4.2 Whole-cell ELISA

The method was adapted from a previous study (Webb et al., 2012) as described below. Over-night cultures were diluted 1:100 into fresh LB, and grown until mid-log phase (OD_{578} = 0.6-0.8). 1 mL of the culture was washed once in PBS, and diluted to OD = 0.2 in PBS. Then 50 µL of the cell suspension was plated onto a 96-well ELISA plate (Greiner, Table4) in two or three technical replicates, and dried for 2 hours at 65°C using a thermoblock on Thermomixer comfort (eppendorf). The dried plate was then incubated with blocking solution (100 µL 1% BSA in TBS) for 30 min at 37°C. After three washes with 200 µL TBST (with Tween-20 0.05%), the wells were incubated with 100 µL primary antibody solution (antibody in 0.1% BSA in TBST, refer to respective antibody dilution in Table 3) for 1 hour at 37°C. After washing the wells, 100 µL of secondary antibody solution (antibody in 0.1% BSA in TBST, refer to respective antibody dilution in Table 3) was added and incubated again for 1 hour at 37°C. Wells were washed with 200 µL TBST (with Tween-20 0.05%), and subsequently, 150 µL of 1-step ABTS (Thermo, Table 4) was added and absorbance kinetics at 405 nm were measured every minute using a BioTek Synergy HT Plate Reader, at room temperature, for 30 minutes. For uncoated blank wells, PBS without cells was used as a blank control, and solution without primary antibody was added to control for non-specific binding of the secondary antibody.

Data analysis – A linear regression curve was fitted to each well for the absorbance at 405 nm for the first 30 min of the incubation for every minute (measured time points are between 25 to 28). The velocity of absorbance for each well was calculated using linear model fit (lm) function in R; the velocity of wells coated with cells were blanked by subtracting the velocities of corresponding uncoated wells (i.e. with primary and secondary, or secondary only antibody incubation). Standard deviations of blanked replicates were used to calculate the error.

2.4.3 Proteinase K surface shaving

The method was adapted from (Webb et al., 2012) with the following modifications. Over-night cultures were diluted 1:100 into fresh LB, and grown until mid-log phase (OD_{578} = 0.6-0.8). 5 mL of culture was harvested, washed once with PBS and resuspended to 0.5 x OD unit in buffer (Tris 20 mM with Tween-20 0.05%). Then 80 µL of the cell suspension was aliquoted and incubated either with or without 2.5 µL of Proteinase K solution (PK, stock 20 mg/mL, Table 4)) and topped up to 100 µL final volume by the same buffer (Tris 20 mM with Tween-20 0.05%). The samples were mixed by finger-flicking and incubated at 37°C for 1 hour. After the incubation, the samples were centrifuged (5000 *xg* for 5 minutes) and washed once with 0.5 mL sterile H_2O. The cell pellet was resuspended in 10% Trichloroacetic acid (TCA), warmed-up for 5 minutes at 65°C, and then cooled-down on ice for 10 minutes. Precipitates were pelleted by centrifugation at maximum speed (12,000 rpm) at 4°C for 10 minutes using an F4-45-30-11 rotor with a Centrifuge 5430 R (Eppendorf), washed once with ice-cold acetone 200 µL, and pelleted again by centrifugation with the same condition. The pellet was then resuspended in 100 µL SDS-PAGE loading buffer (Table 5), boiled at 95°C for 5 minutes. If the pellet was not solubilized, it was incubated longer with occasional vortexing.

2.5 Sucrose gradient fractionation

The method was adapted from a previous study (Anwari et al., 2010). Twice the number of cells were used compared to the biotinylation experiment. Cells were harvested, washed, lysed, and membranes were pelleted in the same way as described for surface biotinylation in section 2.2.1 and 2.2.3, without the incubation step with biotin. After two rounds of ultracentrifugation as described in section 2.2.3, membrane pellets were resuspended in 2 mL of 25% sucrose in 5 mM EDTA. 500 µL of the membrane sample was transferred to the top of a sucrose gradient (1.7 mL of each sucrose solution, 60%, 55%, 50%, 45%, 40%, and 35% sucrose in 5 mM EDTA, pH=7.5 from the bottom of the ultracentrifuge tubes (Ultra-Clear Tube, Table 4)). The samples were ultracentrifuged at 140,000 *xg* for 16 hours at 4°C in a SW40 rotor using an Optima L-100 XP Ultracentrifuge (Beckman). Fractions of 1 mL were collected step-wise from the top of the gradient, yielding 11 fractionated samples that were analyzed by Coomassie staining and Western blotting using SDS-PAGE gels as described below in section 2.5.

Fractions 2 to 11, as well as an aliquot of the input total membrane sample diluted 10 times in H_2O, were submitted to the EMBL Proteomics Core Facility, and measured using 11-plex

TMT-labelling quantitative LC-MS/MS as described above in section 2.3, with the following modification. Before sample preparation for MS, proteins were solubilized by adding SDS to the samples (final concentration of 1 % SDS), sonicated for 5 minutes in a ultrasonic bath, heated for 10 minutes to 80°C, and sonicated again for another 5 minutes. During TMT labelling, 131C label (ThermoFisher) was added to allow 11-plex TMT measurement. During the offline high pH reverse phase fractionation described in section 2.3.1, samples were pooled in total 12 fractions.

For MS data acquisition, the same protocol as in section 2.3.2 was followed, with the following modifications. Peptides were eluted from trapping via the analytical column with a constant flow of 0.3 µL per minute with increasing percentage of Solvent B (0.1% formic acid in acetonitrile) from 2% to 4% in 6 minutes, from 4 % to 8 % in 1 minute, then 8 % to 25 % for a further 71 minutes, and finally from 25 % to 40 % in another 5 minutes. The outlet of the analytical column was coupled directly to a Fusion Lumos (Thermo) mass spectrometer using the Proxeon nanoflow source in positive ion mode.

The peptides were introduced into the Fusion Lumos via a Pico-Tip Emitter 360 µm OD x 20 µm ID; 10 µm tip (New Objective) and an applied spray voltage of 2.4 kV. The capillary temperature was set at 275 °C. Full mass scan was acquired with mass range 375-1500 m/z in profile mode in the Orbitrap with resolution of 120000. The filling time was set at maximum of 50 ms with a limitation of $4x10^5$ ions. Data dependent acquisition (DDA) was performed with the resolution of the Orbitrap set to 30000, with a fill time of 94 ms and a limitation of $1x10^5$ ions. A normalized collision energy of 38 was applied and MS^2 data was acquired in profile mode.

MS data analysis was carried out as described in section 2.3.3, with the different set of modifications included in the search parameters: Carbamidomethyl (C) and TMT10 (K) (fixed modification), Acetyl (Protein N-term), Oxidation (M) and TMT11 (N-term) (variable modifications).

2.6 Protein separation and detection

2.6.1 SDS-PAGE and Coomassie staining

SDS-PAGE gels were prepared using the following procedure with Mini-PROTEAN Tetra Handcast Systems (Bio-rad). For two 15% gels, 2.5 ml of LT buffer (Table 5), 5 mL of 30% acrylamide stock (Table 4), 2.45 mL of H_2O were mixed. 50 µL of 10% ammonium persulfate (APS) and 5 µl of Tetramethylethylenediamine (TEMED, Table 4) was added to the mixture and loaded on the casting system. For two 8% gels, 2.5 ml of LT buffer (Table 5), 2.67 mL of 30% acrylamide stock, 4.78 mL of H_2O was mixed. 50 µL of 10% APS and 5 µL of TEMED was added to the mixture and loaded onto the casting system. For both 15 % and 8 % gels, isopropanol was used to flatten the top of the separating gel. For the gel stacker, 1.25 mL of UT buffer (Table 5), 0.65 mL acrylamide stock, 2.07 mL H_2O was mixed. 25 µL of 10% APS and 5 µL of TEMED was added and the mixture was loaded to fill up the casting glass followed by comb insertion to form the wells.

Protein samples were solubilized and reduced by boiling at 95°C for 5 minutes in Laemmli loading buffer (Table 5). Solubilized samples were loaded and separated in acrylamide gels, 8% and 15% gels were prepared as described above, or gradient gels of 4-20% acrylamide (Teo-Tricine gels from Expedeon, Table 4) using the respective running buffers (Table 5). For both gels, Bio-rad systems were used, 30 mA per gel for 8% and 15% gels, 100 V per chamber for gradient gels. Gels were incubated with a Coomassie staining solution (Table 5) for 1 hour, and destained with destaining solution (Table 5) until the desirable signal was achieved. Incubations were performed at room temperature with constant moderate mixing by rocking.

2.6.2 Western blot

Proteins were separated on acrylamide gels as described above, and transferred to methanol-activated PVDF membranes (Table 4), using Western blot transfer buffer (Table 5) for 1.5 hours at 100 V. All the incubation steps from here on were performed with constant moderate agitation by using rocking platforms. Membranes were blocked for 1 hour with 5% skim milk in TBST, and then incubated with appropriately diluted primary antibodies (dilutions summarized in Table 3) in 5% skim milk in TBST overnight at 4°C or 2 hours at room temperature. After three times of 5 minutes washes with TBST, membranes were incubated for 1 hour with secondary antibodies conjugated with horseradish peroxidase (HRP) diluted in 5% skim milk in TBST (1:5,000 for α-mouse, 1:10,000 for α-rabbit). For Western blots using Streptavidin-HRP conjugate (against biotin), the membranes were blocked for 1 hour with 5%

skim-milk in TBST, and incubated for 1 hour with Streptavidin-HRP diluted in 5% BSA in TBST. After these antibody incubations, membranes were washed again three times for 5 minutes with TBST. Proteins were detected by adding ECL substrate (Table 3), then exposed, and visualized either using X-ray films (Table 3), or a digital developing machine (ChemiDoc™ Touch Imaging System).

2.6.3 Western blot quantification

Western blots were quantified using the following methods. For Streptavidin (against biotin) Western blot, the method was adapted from a previous study (Brochado et al., 2018). Each lane of the blot was quantified using the same size rectangle and the "grey mean value" was measured in Fiji software (Schindelin et al., 2012). The signal level was quantified by subtracting the measured value from 255, set as the maximum gray mean value.

For α-FtsZ Western blot, the built-in gel analysis tool in the Fiji software (Schindelin et al., 2012) was used to quantify FtsZ levels, where the bands across the entire blot were quantified from a horizontal rectangle area. After α-FtsZ Western blot data acquisition, the membrane was re-incubated and quantified for RecA (loading control protein) levels using α-RecA antibody following the procedure in section 2.6.2. RecA levels were quantified in the same way as for FtsZ, and the ratio of FtsZ / RecA signals were calculated.

2.7 RcsF-IM experiments

2.7.1 RcsF-IM expression

pNG162 plasmid is a low-copy pSC101 origin vector previously used to express wildtype RcsF (Cho et al., 2014). RcsF-IM, an *rcsF* gene carrying two amino acid substitutions relative to the wildtype (S17D/M18D), were introduced into pNG162 background (work by Matylda Zietek, unpublished). A high copy plasmid, pTrcHis2A expressing the *lac* repressor protein LacIq (Thermo, Table 2), was kept together with pNG162 during the entire procedure in order to prevent a mucoid phenotype from appearing (caused by activation of the Rcs system through leaky expression of RcsF-IM.)

Overnight cultures were grown at 37°C in LB-Lennox with Ampicillin and Spectinomycin (LB-AmpSpec, both at 100 μg/mL). 1M IPTG stock was prepared and filter sterilized prior to RcsF-IM induction. Cells were diluted to OD_{578} = 0.001 in LB-AmpSpec, and grown for ~3 hours

until around OD_{578} = 0.1. Cells were then diluted to OD_{578} = 0.025 in fresh LB with appropriate IPTG concentrations (ranging between 0 to 20 μM). In order to measure growth rate, OD_{578} was measured throughout the experiment, and once the OD_{578} = 0.3 was reached, cells were then diluted to OD_{578} = 0.025. This cycle was maintained for 6 to 9 hours.

2.7.2 β-galactosidase assay

β-galactosidase (β-gal) assay was performed as previously described (J. H. Miller, 1972). 1 mL of growing culture was sampled and harvested by centrifuging at maximum speed (14,000 rpm) at 4°C using a F4-45-30-11 rotor with a Centrifuge 5430 R (Eppendorf). The supernatant was discarded and the cell pellet was stored at -20°C. The pellet was resuspended in 1 mL of Z-buffer (Table 5). If needed, the samples were further diluted in Z-buffer to slow-down the reaction, in order for it to be within a measurable time window. 2 drops of $CHCl_3$ and 1 drop of 0.1% SDS were added, and the cells were vortexed to in order to lyse them. 200 μL of Ortho-Nitrophenyl-β-galactoside (ONPG, freshly prepared, 4 mg/mL stock in H_2O) was added and mixed. When the sample started turning yellow, the reaction was stopped by adding and mixing 500 μL of 1M Na_2CO_3. Samples were centrifuged for 5 minutes, and 300 μL of the supernatant were placed to wells on a Flat-bottom 96-well plate (Nuclon™ Delta Surface, Thermo). Absorbance at 405 nm was measured using a BioTek Synergy HT Plate Reader. Specific β-gal activity was calculated using the following equation:

Specific β-gal activity (μmol / min / mg) = $3.38*A_{405} / (\Delta t*V_{cells}*OD_{578})$, with the parameters defined as below:

A_{405} = absorbance at 405 nm on plate reader after reaction, blanked for no cell control

Δt = reaction time (from addition of ONPG to addition of Na_2CO_3)

V_{cells} = volume of cells of original OD_{578}used for the reaction

OD_{578} = OD_{578} of the culture when sampled

2.7.3 Growth-rate calculation

Growth rate was calculated from OD_{578} measurements for the cultures growing exponentially (below OD_{578} = 0.4) using three or more measurements, fitted to an exponential curve ($y=ae^{(bx)}$, b: growth rate) using a graphing and data analysis software KaleidaGraph (Kirsch and Ekerdt, 2000). For an average growth rate, the last three quantified growth rates were used.

2.7.4 FtsZ level measurement

Cells were grown and harvested in the same ways as in 2.7.1. After incubation with appropriate concentration of IPTG (between 0 and 15 μM) for minimum 1.5 hours, 1 mL of growing culture (in exponential phase) was harvested, and lysed and solubilized by boiling in Laemmli buffer for 5 minutes at 95°C. Samples were diluted and normalized using OD_{578} prior to SDS-PAGE analysis. Western blots using α-FtsZ was performed, then the same membrane was re-incubated and used for α-RecA (loading control) Western blot. The result was quantified as described in section 2.6.3.

2.7.5 Cell length measurement

This experiment and analysis were performed by Matylda Zietek (Typas Lab, EMBL Heidelberg), and manuscript is about to be submitted, thus only brief experimental procedure is described here.

After 1 hour incubation with IPTG, samples were collected, and spun down at 1000 *xg* for 5 minutes, washed 3 times with PBS, and fixed in 4% formaldehyde. Fixed cells were observed under a widefield microscope (Zeiss Cellobserver). Microscope images were quantified, for at least 500 cells, using the Morphometrics software and a custom Matlab script.

2.7.6 Single-cell microscopy

This experiment and analysis were performed by Amanda Miguel (Huang Lab, Stanford University), and manuscript is about to be submitted, thus only brief experimental procedure is described here.

A microfluidic device (CellASIC) was used to image *E. coli* cells under the microscope for at least 100 minutes. Cells were carrying a plasmid carrying an *rprA::msfGFP* to measure Rcs activity.

Agarose pads were used to measure FtsZ levels using a strain carrying a functional chromosomal FtsZ-sfGFP. For cephalexin treatment, 10 μg/mL cephalexin was used. Data analysis was carried out using the Morphometrics software and a custom Matlab script.

2.8 Data analysis and visualization

Data was analyzed and plotted using R (R Core Team, 2019), unless otherwise stated. Figures were prepared using Adobe Illustrator (Adobe).

2.9 Databases

Protein sequences were obtained from the UniProt database (The UniProt Consortium, 2019), unless otherwise stated.

Protein localization information was obtained from STEPdb database (stepdb.eu) (Orfanoudaki and Economou, 2014), unless otherwise stated. If the protein names were not found in the STEPdb 2.0, the STEPdb 1.0 was used. From the database annotation, proteins were grouped into the followings:

Localization name (symbol) in STEPdb	**Name used in thesis**
Nucleoid (N), Cytoplasmic (A), ribosomal (r)	Cytoplasm
Peripheral inner membrane protein facing cytoplasm (F1)	IM-cyto
Integral inner membrane (B)	IM-integral
Peripheral inner membrane protein facing periplasm (F1)	IM-peri
Inner membrane lipoproteins (E)	IMLP
Periplasmic (G), peripheral outer membrane protein facing periplasm (F3)	Periplasm
Outer membrane lipoprotein (I)	OMLP
Outer membrane β-barrel protein (H)	OMP
Peripheral outer membrane protein facing the extra-cellular space (F4), Extracellular (X)	Secreted

*For proteins with dual (or more) localizations, membrane localization was used, such as if a protein has A and F1 localization codes, it is assigned as IM-cyto.

Table 1: Strains

Strain number	Genotype	Name used	Publication and construction
NT15001	MG1655 Δ*(argF-lac)U169 rprA::lacZ*	WT	Wildtype (Cho et al., 2014; Majdalani et al., 2002)
NT15097	MG1655 Δ*(argF-lac)U169 rprA::lacZ bamA101::Kan*	*bamA101*	P1 transduction to wildtype (Aoki et al., 2008)
NT15174 / NT15175	MG1655 Δ*(argF-lac)U169 rprA::lacZ* Δ*bamE::Kan*	Δ*bamE*	P1 transduction from KEIO mutant to wildtype, NT15174 found to have lost LacZ production
NT15153	MG1655 Δ*(argF-lac)U169 rprA::lacZ bamA6::Kan*	*bamA6*	P1 transduction to wildtype from *bamA6* (formally *yaeT6*) mutant (Ruiz et al., 2006)
NT15048	MG1655 Δ(*argF-lac)U169 rprA::lacZ* Δ*rcsF::Km*	Δ*rcsF*	P1 transduction from KEIO mutant to wildtype
NT15178	MG1655 Δ*(argF-lac)U169 rprA::lacZ* Δ*rcsF::Kan* Δ*rcsB::Cm*	Δ*rcsF* Δ*rcsB*	P1 transduction from ASKA mutant to Δ rcsF
NT15259	MG1655 Δ*(argF-lac)U169 rprA::lacZ* Δ*yedD::Kan*	Δ*yedD*	P1 transduction from KEIO mutant to wildtype
NT15268	MG1655 Δ*(argF-lac)U169 rprA::lacZ* Δ*acrA::Kan*	Δ*acrA*	P1 transduction from KEIO mutant to wildtype
NT15212	MG1655 Δ(*argF-lac)U169 rprA::lacZ rcsF::Flag::Kan*		Flag-tagged strain
NT15214	MG1655 Δ*(argF-lac)U169 rprA::lacZ surA::Flag::Kan*		Flag-tagged strain
NT15218	MG1655 Δ*(argF-lac)U169 rprA::lacZ yedD::Flag::Kan*		Flag-tagged strain
NT15220	MG1655 Δ*(argF-lac)U169 rprA::lacZ mlaA::Flag::Kan*		Flag-tagged strain
NT15221	MG1655 Δ*(argF-lac)U169 rprA::lacZ metQ::Flag::Kan*		Flag-tagged strain
NT15224	MG1655 Δ*(argF-lac)U169 rprA::lacZ yiaD::Flag::Kan*		Flag-tagged strain
NT15233	MG1655 Δ*(argF-lac)U169 rprA::lacZ osmE::Flag::Kan*		Flag-tagged strain
NT15234	MG1655 Δ*(argF-lac)U169 rprA::lacZ yraP::Flag::Kan*		Flag-tagged strain
NT15237	MG1655 Δ*(argF-lac)U169 rprA::lacZ dcrB::Flag::Kan*		Flag-tagged strain

Table 2: Plasmids

Name	Antibiotic resistance	Publication
pKD46	Ampicillin	(Datsenko and Wanner, 2000)
pJPS1	Kanamycin	Joel Selkrig (Typas Lab), Unpublished
pNG162-Empty	Spectinomycin	(Cho et al., 2014)
pNG162-RcsF-WT	Spectinomycin	(Cho et al., 2014)
pNG162-RcsF-IM	Spectinomycin	Matylda Zietek (Typas lab), Unpublished
pTrcHis2A	Ampicillin	Invitrogen
pJV300-Empty	Ampicillin	(Bouvier et al., 2008; Sittka et al., 2007)
pFM1-1-rybB	Ampicillin	(Bouvier et al., 2008)
pMZ13	Chloramphenicol	Matylda Zietek (Typas lab), Unpublished

Table 3: Antibodies

Antigen	Secondary	Dilution used for Western blot (ELISA)	Source
AcrA	Rabbit	200,000 (10,000 for ELISA)	Gifted from Pos Lab (Goethe University, Germany) (Brochado et al., 2018)
BamA	Rabbit	10,000	Gifted from Lithgow Lab (Monash University, Australia)
Flag	Mouse	1,000 (500 for ELISA)	Sigma (F1804-50UG)
FtsZ	Rabbit	1,000	Acris (AS10715)
Mouse	HRP conjugated	5,000 (2,000 for ELISA)	Sigma (A4416)
OmpC	Rabbit	10,000	Gifted from Lithgow Lab (Monash University, Australia)
Rabbit	HRP conjugated	10,000 (2,000 for ELISA)	GE healthcare (NA934)
RcsF	Rabbit	2,000	Lab stock (Cho et al., 2014)
RecA	Rabbit	1,000	Abcam (ab63797)
SecG	Rabbit	6,000	Gifted from Tokuda Lab (University of Morioka, Japan)
Streptavidin	HRP conjugated	10,000	Rockland (S000-03)
YedD	Rabbit	1,000 for ELISA	Gifted from Collet Lab (UCL, Belgium)

Table 4: Chemicals

Common name	Product name	Company	Catalog number
1-step ABTS	1-Step™ ABTS Substrate Solution	Thermo	37615
Acryamide	30% Acrylamide/Bis Solution	Biorad	161-0159
Biotin	EZ-Link™ NHS-LC-LC-Biotin	Thermo	21343
Brilliant Blue R-250	Coomassie® Brilliant Blue R-250	AppliChem	A1092-0025
DDM	Dodecyl-β-D-maltosid	Roth	CN26.3
ECL	ECL™ Western Blotting Reagents	GE Healthcare	RPN2106
ECL sensitive	SuperSignal™ West Pico PLUS Chemiluminescent Substrate	Thermo	34580
Electroporation cuvettes	Gene Pulser®/MicroPulser™ Electroporation Cuvettes, 0.2 cm gap	Biorad	1652082
ELISA plate	MICROPLATE, 96 WELL, PS, F-BOTTOM, CLEAR, MICROLON®, HIGH BINDING	Greiner	655061
Elution buffer	10x Buffer E; Strep-Tactin® Elution Buffer with Desthiobiotin	IBA	2-1000-025
Gel extraction kit	QIAquick Gel Extraction Kit	Qiagen	28706
Gradient SDS-PAGE gel (12 wells)	RunBlue™ TEO-Tricine SDS Gels 4-20% 12 well	Expedeon	NXG42012
Gradient SDS-PAGE gel (17 wells)	RunBlue™ TEO-Tricine SDS Gels 4-20% 17 well	Expedeon	NXG42027
IPTG	Isopropyl β-D-1-thiogalactopyranoside	VWR	730-1497
Lysozyme	Lysozyme from chicken egg white	Sigma	L6876
Miniprep kit	QIAprep Spin Miniprep Kit	Qiagen	27106
Nitrocellulose membrane	Whatman Protran	PerkinElmer	NBA085C001EA
ONPG	2-Nitrophenyl β-D-galactopyranoside	Sigma	73660-25G
Protease inhibitor	cOmplete™, EDTA-free Protease Inhibitor Cocktail	Roche	COEDTAF-RO
Protein ladder	Color Prestained Protein Standard, Broad Range	NEB	P7712
Protein ladder dual color	Precision Plus Protein™ Dual Color Standards	Biorad	1610374
Proteinase K	Proteinase K (from Tritirachium album)	Merck	1.24568.0100
PVDF membrane	Immobilon-P PVDF Membrane	Merck	IPVH00010
Q5 polymerase	Q5® Hot Start High-Fidelity DNA Polymerase	NEB	M0493S

Strep-Tactin Sepharose	Strep-Tactin® Sepharose® 50% suspension	IBA	2-1201-010
Sucrose	Sucrose	Sigma	S7903
TEMED	N,N,N′,N′-Tetramethylethylenediamine	Sigma	T9281
TMT 10plex	TMT10plex™ Isobaric Label Reagent Set	Thermo	90406
TMT11-131C	TMT11-131C Label Reagent	Thermo	A34807
Trypsin	Sequencing Grade Modified Trypsin	Promega	V5111
Tween-20	TWEEN® 20	Sigma	P1379-1L
Ultra-Clear centrifuge tubes for SW40	Ultra-Clear Tubes (14 x 95 mm)	Beckman	344060
Western blot film	LucentBlue X-ray film	advansta	L-07013-100

Table 5: Buffers and media

Name	Recipe
2xYT	10 g Bactotryptone, 10 g Yeast Extraxt, 5 g NaCl per 1 L (pH = 7.2)
Coomassie staining buffer	50% methanol, 40 % H_2O, 10% acetic acid, 1 g Brillient Blue R250 per 1 L
Destaining buffer	40% ethanol, 10% acetic acid, 50% H_2O
Laemmli loading buffer (x4)	200 mM Tris-HCl at pH = 6.8, 8% SDS, 40% glycerol, 400 mM DTT, 0.02% bromophenol blue
LB-agar	LB-lennox with 15 g agar per 1 L
LB-Lennox	10 g Bactotyptone, 5 g Yeast Extract, 5 g NaCl per 1 L
LT buffer (for SDS-PAGE)	1.5 M Tris-HCl at pH = 8.8, 0.4% SDS
Lysis buffer	1.5 mM EDTA, pH = 7.5
Mixed buffer (for biotinylation)	1:2 ratio of spheroplasting to lysis buffer
Permeabilizing buffer	PBS with 20 mM EDTA and 1 mg/mL lysozyme
Resuspending buffer (for P1 transduction)	10 mM of $MgSO_4$ and 5 mM $CaCl_2$
Run-Blue running buffer	0.8 M Tricine, 1.2 M Triethanolamine, 2% SDS
SDS-PAGE running buffer	3 g Tris, 14.42 g glycine, 0.5 g SDS per 1 L
Spheroplasting buffer	0.75 M sucrose in PBS
TBS	20 mM Tris, 10 mM NaCl
TBST	TBS with 0.1% Tween-20, or otherwise stated in text
TSS	20% (w/v) polyethyleneglycol, 100 mM $MgSO_4$, 10% DMSO
UT buffer (for SDS-PAGE)	0.5 M Tris-HCl at pH = 6.8, 0.4% SDS
Western blot transfer buffer	3.03 g Tris, 14.4 g Glycine, 200 mL methanol per 1 L
Z buffer	60 mM $Na_2HPO_4.2H_2O$, 40 mM $NaH_2PO_4.H_2O$, 10 mM KCl, 1 mM $MgSO_4.7H_2O$

Table 6: Primers

Primer name	Sequence
RcsF_H1P1	GCTATCGTCAGGCTGTATGTATCGGTTCTGCGCTTAACATTACGGCGAAAGGCGGCGAC TACAAGGATGACGATGACAAGTAGTGACCTGAGGGGAGATATCG
RcsF_H2P2	TTCTTTATAGGGCGAGCGAATAACGCCTATTTGCTCGAACTGGAAACTGCCATATGAAT ATCCTCCTTAG
SurA_H1P1	GCTGGATGCAGGAACAACGTGCCAGCGCCTACGTTAAAATCCTGAGCAACGGCGGCGA CTACAAGGATGACGATGACAAGTAGTGACCTGAGGGGAGATATCG
SurA_H2P2	AATCCCGGCGGGCTCGCCGGGAGTGATCACAACACGTTGGGTTTTAACCACATATGAAT ATCCTCCTTAG
YedD_H1P1	CAGAGTGTGCCGCCGCGCTGGATAAAGCGCCGTTACCGACGCCACTGCCGGGCGGCGA CTACAAGGATGACGATGACAAGTAGTGACCTGAGGGGAGATATCG
YedD_H2P2	TCGAAGAGGTGATTTAAATTCATCCCCGGCGGCAAGCCGGGGAGATTTCACATATGAAT ATCCTCCTTAG
YiaD_H1P1	CAGAAGGTAAGGCGCAAAACCGCCGTGTAGAAATTACCTTAAGCCCGCTGGGCGGCGA CTACAAGGATGACGATGACAAGTAGTGACCTGAGGGGAGATATCG
YiaD_H2P2	GAAAACTGAAGAATGCCAGGTGATTTCTGCATCACCTGGCATGAAAGGGACATATGAA TATCCTCCTTAG
MlaA_H1P1	ATCCGAACGCACAAGCGATTCAGGATGATTTAAAAGATATTGATTCTGAAGGCGGCGAC TACAAGGATGACGATGACAAGTAGTGACCTGAGGGGAGATATCG
MlaA_H2P2	GATAACAACAAAAAAGGTGAGTTTTGCGACTCACCTTTTTTATTTGTTTCCATATGAATA TCCTCCTTAG
MetQ_H1P1	TTTACGAAGCAGCAAACAAAGTGTTTAACGGCGGAGCTGTTAAAGGCTGGGGCGGCGA CTACAAGGATGACGATGACAAGTAGTGACCTGAGGGGAGATATCG
MetQ_H2P2	GCAAGCGCCCGTCCTGAATGATATTACAAATTGTGGAAACAGCCTAAAAACATATGAAT ATCCTCCTTAG
OsmE_H1P1	CCGGTTATCAGACCTGTGCTGAATACGACACTGATCCACAGGCTGCGAAGGGCGGCGAC TACAAGGATGACGATGACAAGTAGTGACCTGAGGGGAGATATCG
OsmE_H2P2	CCGGCGCAATGGCCGGTTTCCGTTGTTACTCAAGCTTTCAGACGAATTGACATATGAAT ATCCTCCTTAG
YraP_H1P1	GCCGGGTGAGCGGCGTGAAGCGGGTAACTACGGCGTTTACGTTTATTAAAGGCGGCGA CTACAAGGATGACGATGACAAGTAGTGACCTGAGGGGAGATATCG
YraP_H2P2	TAGGCCTGATAAGACGCGGCAAGCGTCGCATCAGGCATTACAAGGGGCTGCATATGAA TATCCTCCTTAG
MdtE_H1P1	CACGAGCAATTTCCTCCAGCCAGGAAAACGCCAGCACCGAATCGAAACAAGGCGGCGA CTACAAGGATGACGATGACAAGTAGTGACCTGAGGGGAGATATCG
MdtE_H2P2	GGGCGATCAATAAAATAGTTAGCCATGAAAGTCCCCTTAAGCCTGCAACGCATATGAAT ATCCTCCTTAG
DcrB_H1P1	AAAAAGCGCAGACCACCGCAGAAAACATCATTAATACGCTGGTTATTCAGGGCGGCGA CTACAAGGATGACGATGACAAGTAGTGACCTGAGGGGAGATATCG
DcrB_H2P2	AAAAACGGCCTCCGGAACACGTCCCTGAGGCCGCCTCATCATCTTAAAACCATATGAAT ATCCTCCTTAG

Table 7: Summary of mass spectrometry samples

MS run ID	Strain	Biotinylation state	Growth phase	Salt	Replicate
P0413	WT	Intact	Exponential	–	1
	WT	Lysate	Exponential		1
	WT	No biotin	Exponential		1
	bamA101	Intact	Exponential		1
	bamA101	Lysate	Exponential		1
	bamA101	No biotin	Exponential		1
	bamA6	Intact	Exponential		1
	bamA6	Lysate	Exponential	+	1
	WT	Intact	Exponential		1
	WT	Lysate	Exponential		1
P0444	WT	Intact	Exponential		2
	WT	Lysate	Exponential		2
	bamA101	Intact	Exponential		1
	bamA101	Lysate	Exponential		1
	WT	Intact	Stationary		1
	WT	Lysate	Stationary		1
	bamA101	Intact	Stationary		1
	bamA101	Lysate	Stationary		1
	bamA101	Intact	Exponential		2
	bamA101	Lysate	Exponential		2
P0483	WT	Intact	Exponential		3
	WT	Lysate	Exponential		3
	bamA101	Intact	Exponential		2
	bamA101	Lysate	Exponential		2
	WT	Intact	Stationary		2
	WT	Lysate	Stationary		2
	bamA101	Intact	Stationary		2
	bamA101	Lysate	Stationary		2
	Δ*bamE*	Intact	Exponential		1
	Δ*bamE*	Lysate	Exponential		1
P0919	WT	Intact	Exponential		4
	WT	Lysate	Exponential		4

	Δ*bamE*	Intact	Exponential		2
	Δ*bamE*	Lysate	Exponential		2
	Δ*bamE*	Intact	Exponential		3
	Δ*bamE*	Lysate	Exponential		3
	WT with pJV300 (control)	Intact	Exponential		1
	WT with pJV300 (control)	Lysate	Exponential		1
	WT with pFM1-1 (rybB)	Intact	Exponential		1
	WT with pFM1-1 (rybB)	Lysate	Exponential		1

Chapter 3: Lipoprotein surface exposure in *Escherichia coli*

3.1 Background

Despite the increasing number of identified surface lipoproteins in *E. coli* and other Gram-negative bacteria, systematic studies for lipoprotein surface exposure are limited. It is therefore still unknown how prevalent surface exposure is for lipoproteins in *E. coli*. Surface protein labelling is a commonly used method to identify proteins on the cell surface. Using a specific crosslinker that does not penetrate the outer-most membrane, proteins on the cell surface can be labelled in the extracellular environment, purified and studied. Such approaches have been established in eukaryotes and Gram-positive bacteria, both of which have single non-polar membranes at their surface (Dreisbach et al., 2011; Esbelin et al., 2018; Roesli et al., 2006; Rugg-Gunn et al., 2012). For example, proteome-wide cell surface proteome studies, including surface labelling methods, have been performed for *Staphylococcus aureus* and *Listeria monocytogenes* (Dreisbach et al., 2011; Esbelin et al., 2018; Hempel et al., 2010). However, for Gram-negative bacteria, the same labelling materials and methods do not work: due to the unique properties of the OM, such labels often enter the cell via passive diffusion through the β-barrel lumen of porins (Monteiro et al., 2018).

In a study by Monteiro et al. in 2018, two different cross-linkable biotin probes, Sulfo-NHS-SS-biotin and Sulfo-NHS-PEG4-bismannose-SS-biotin, with different molecular sizes, were used to explore and compare labelling patterns in *E. coli*. They observed that the surface specificity of these probes were not enough, and led to the detection of non-OM proteins, including cytoplasmic proteins (Monteiro et al., 2018). Thus, a systematic and quantitative assay for surface exposed proteins in Gram-negative bacteria is still missing.

To this end, I developed a quantitative cell-surface labelling method for *E. coli* which I named surface proteome quantification (SPQ). NHS-LC-LC-Biotin (Thermo) was previously shown to be a cross-linkable biotin that cannot enter *E. coli* and can thus label surface proteins specifically (Cowles et al., 2011). Lately, this NHS-LC-LC-Biotin has been used to identify different lipoproteins for their surface localization, including Pal in *E. coli* (Surendran et al., 2015), and it has been used in a proteome-wide study in *Bacteroides*, another Gram-negative bacterium (Wilson et al., 2015). I used the same cross-linkable biotin to label surface proteins,

and developed a method that quantitatively measures the level of surface exposure of proteins through mass spectrometry (MS). Using this method, I managed to investigate surface exposure of proteins in *E. coli*, including lipoproteins, in a systematic and quantitative manner.

3.2 Optimization of surface labelling with NHS-LC-LC-Biotin

Firstly, I tested whether the NHS-LC-LC-Biotin (later referred to as biotin) can specifically label surface proteins as shown before (Cowles et al., 2011). As the surface of *E. coli* is the OM, the biotinylation level of membrane proteins pelleted by ultracentrifugation, containing both the OM and IM, was compared with that of the soluble fraction (supernatant of the ultracentrifugation step), containing cytoplasmic and periplasmic proteins. The level of biotinylation was quantified from Westerm blots against biotin, using Sterptavidin-HRP, revealing the enrichment of biotinylation in the membrane fraction relative to the soluble fraction when the cells were biotinylated in intact state (Figure 3.1A). The lack of a similar enrichment for membrane proteins when using the probe in the cellular lysate suggests that the biotinylation probe specifically labels surface exposed proteins in intact cells because it cannot enter the cells rather than having any spurious preference to membrane proteins (Figure 3.1A).

Next, I optimized the labelling duration, to see if longer/shorter incubation times improve the labelling. The results indicated that the membrane fraction was strongly labelled after 15 min incubation time, which was the minimal incubation time for handling the samples (Figure 3.1A). Longer incubation times (30 and 90 min) led to higher biotinylation of the soluble fraction in intact cell biotinylation, suggesting increased penetration of the biotin through the OM (Figure 3.1A). From the clear difference in the biotinylation patterns of intact cells and cellular lysates, I concluded that the biotin has selectivity for membrane proteins when applied to intact cells, presumably due to its difficulty to cross the OM. Furthermore, shorter incubation times decreased the residual penetration of the biotin through the OM. In contrast to a previous report (Cowles et al., 2011), NHS-LC-LC-Biotin entered the cell to a small extent, but still retained its preferential labelling to membrane (likely surface) proteins, thus encouraged me to use it further as a surface labelling probe.

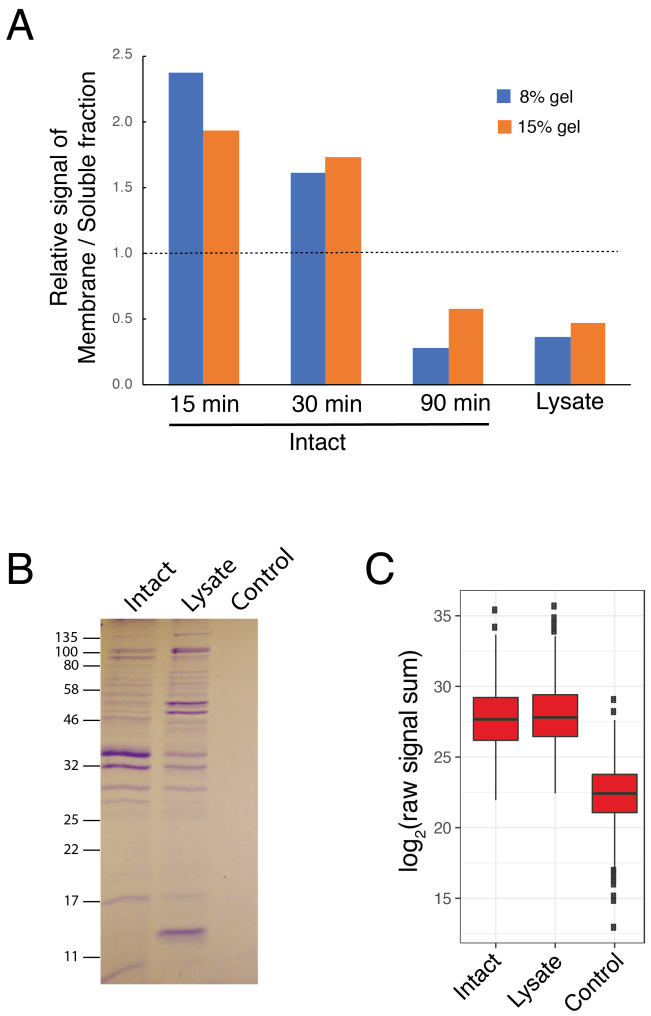


Figure 3.1: Optimizing biotinylation conditions. A) Level of protein biotinylation quantified from Western blot against biotin (using Streptavidin-HRP conjugate). Four different biotinylation conditions (15, 30 and 90 min intact cell biotinylation and cell lysate biotinylation) are quantified and compared for relative signal of membrane/soluble fraction. Fractions were separated by ultracentrifugation. Western blots were performed using 8% and 15% SDS-PAGE gels, and quantified separately (biological replicate n=1, which was analyzed on two separate gels (8% and 15%) as technical replicates). Original blots are found in Appendix. B) Coomassie staining of Strep-Tactin pull-downs from intact and lysate biotinylation samples, with negative control without biotinylation (biological replicate n=1). C) Log_2 values of raw signal sum (sum of signals of reporter ions in MS quantification) values of same three conditions as B) is shown (biological replicate n=1).

As I observed limited but detectable biotinylation of intracellular proteins, I decided to quantitatively compare the surface labelling with lysate labelling. With this approach, even if the biotin probe is partially entering the cell and thus labelling intracellular proteins, I can distinguish them from the surface proteins as the level of biotinylation relative to the lysate should be different. The comparison with lysate biotinylation also allows the measurement of the protein surface exposure while taking into account the protein abundance, labelling bias, and pull-down background. Coomassie staining of pulled-down samples showed the different pattern of enriched proteins between intact and lysate biotinylated samples, confirming the difference of the two labelling conditions (Figure 3.1B).

Then, I tested whether the pull-down step using Strep-Tactin Sepharose beads (IBA), originally developed to pull-down Strep-tagged proteins, specifically enriched biotinylated proteins. A non-biotinylated control sample showed no visible band after the pull-down step (Figure 3.1B), suggesting negligible cross reactivity of the beads to non-biotinylated proteins. A non-biotinylated sample was also measured in a trial MS run (with the support of Proteomics Core Facility at EMBL), showing significantly lowered signal compared to both intact and lysate biotinylated samples (Figure 3.1C). Furthermore, I tested if addition of salt (1M NaCl in washing buffer of beads) during the pull-down could reduce the background enrichment of non-biotinylated proteins. NaCl can disturb residual protein-protein interactions which may have been kept during membrane solubilization with DDM, a non-ionic detergent. I performed one quantitative MS run comparing the two methods (with and without NaCl). Many cytoplasmic proteins, including ribosomal proteins (Rps and Rpl proteins), showed decreased signals in the sample treated with NaCl compared to without NaCl treatment, in both intact and lysate samples (Figure 3.2). Overall, most proteins showed little difference between the two washing methods, suggesting that the effect of the salt is minor (Figure 3.2). As the addition of NaCl was shown to be effective at removing some non-membrane proteins, it was used in all the following sample preparations (Table 7).

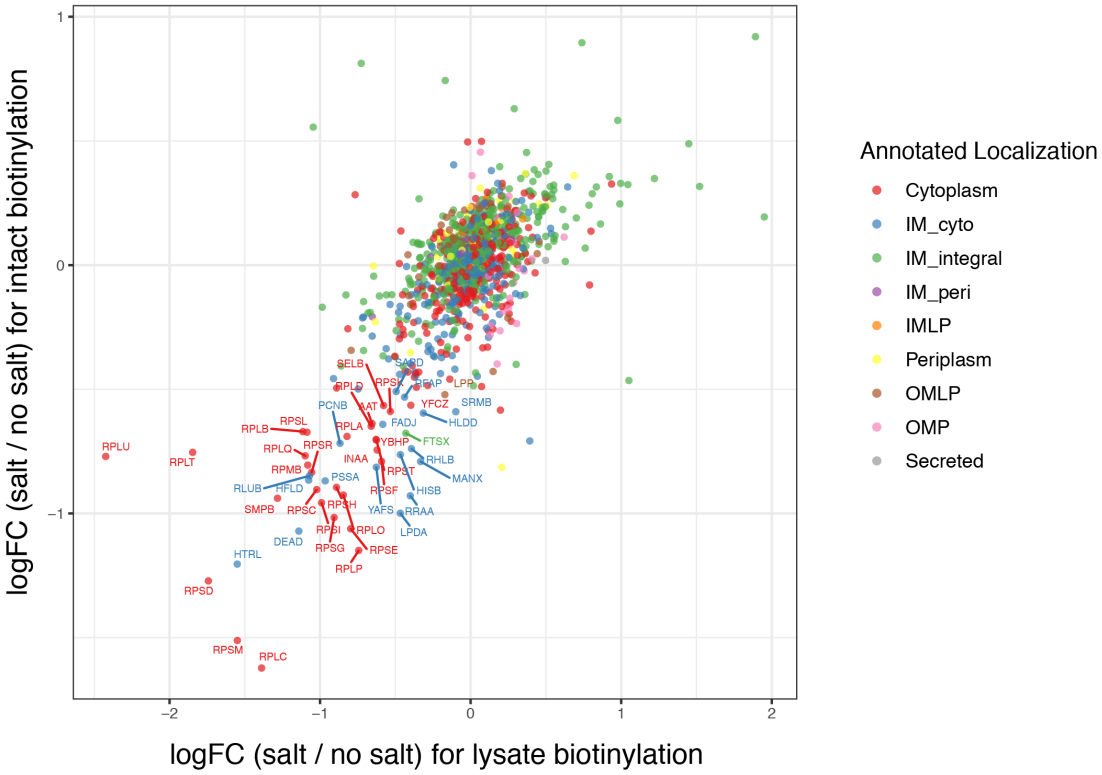


Figure 3.2: Comparison of the pull-downs performed with and without salt. Scatter plot comparing log fold-change (logFC = $\log_2$) of the MS signal ratio of each protein between salt (NaCl) and no salt samples for lysate biotinylation (x-axis) and intact biotinylation (y-axis), data from biological replicate n=1. Proteins showing x < 0 and y < -0.5 are annotated with protein names, and dots are colored according to the annotated localization from STEPdb (colors in the legend).

At the end, I combined the surface biotinylation method using NHS-LC-LC-Biotin with tandem mass tag (TMT) labelling quantitative MS which allows proteome-wide and quantitative comparison of the two samples: surface biotinylation of intact cells and cellular lysate biotinylation. The detailed protocol of this SPQ method is described in the Materials and Method section, and summarized as an illustration in Figure 3.3.

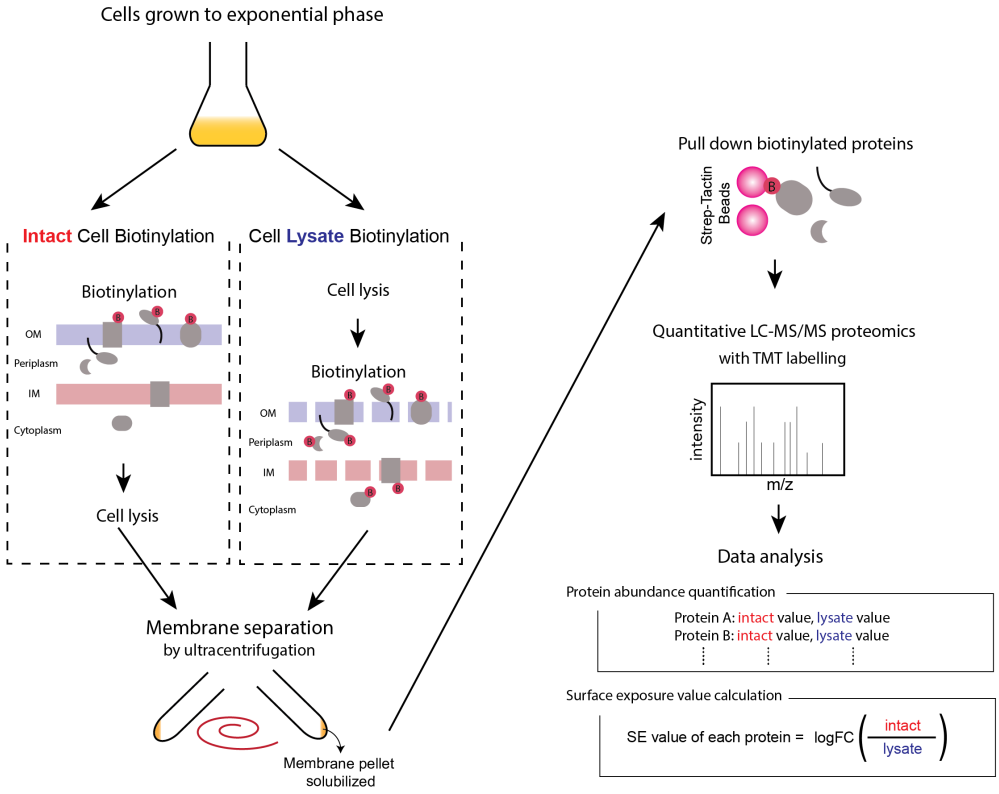


Figure 3.3: Illustration of experimental workflow. Schematic illustration for the sample preparation protocol. Cells were grown until mid-log exponential phase and separated into two fractions: for intact and lysate biotinylation. After biotinylation and cell lysis, membrane pellets were collected by ultracentrifugation and biotinylated proteins were pulled down by Strep-Tactin Sepharose beads. Samples eluted from the beads were quantified using TMT-labelling quantitative MS, and surface exposure (SE) values were calculated for each protein by log fold- change of (intact/lysate) ratio.

3.3 Protein surface exposure level quantified by TMT-labelling mass spectrometry

In order to identify and quantify the proteins biotinylated, I (with the support of Proteomics Core Facility at EMBL) performed TMT-labelling quantitative LC-MS/MS. Pulled-down protein samples, enriched for biotinylated proteins, were prepared and run for LC-MS/MS by Mandy Rettel at the Proteomics Core Facility at EMBL, as described in the Material and Method section and in MS sample list (Table 7). Four biological replicates were prepared in separate experiments, and quantified in different MS runs on different days. Intact and lysate pairs of each sample were run in the same MS run to facilitate quantitative abundance comparisons using TMT-label MS/MS values. The majority of proteins were detected across

all runs (Figure 3.4A) and 1194 proteins were found in at least 3 out of 4 runs with at least two quantified unique peptides, and these were used for further analysis (Figure 3.4A, details in Materials and Methods).

MS data was normalized to compare and merge replicates for the further analysis by Frank Stein at Proteomics Core Facility at EMBL. Firstly, batch effects were removed using the Limma package (Ritchie et al., 2015) and data were normalized using the "vsn" R/bioconductor package (Huber et al., 2002). Then, a data imputation step was carried out for proteins with missing values in one out of the four experiments by using k-nearest neighbors. The PCA plot comparing before and after the data normalization step confirmed the successful normalization which resulted >80% of sample variability originating from the two cellular states of biotinylation (intact and lysate biotinylation) (Figure 3.4B).

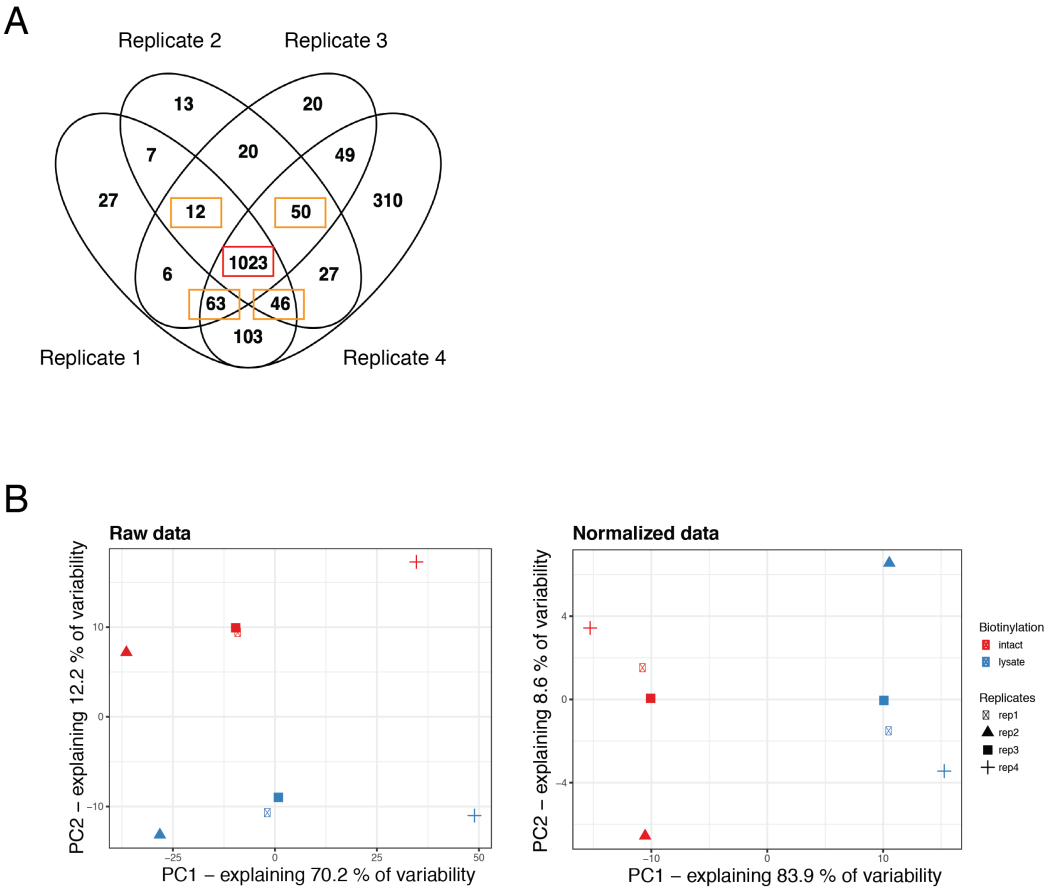


Figure 3.4: MS protein identification and normalization. A) Commonly identified proteins across four MS runs, the Venn diagram was produced using Venny2.1 (Oliveros, 2015). Red – 1023 proteins identified in all 4 replicates, yellow – 12, 50, 63, and 46 proteins identified in 3 out of 4 replicates, in total 1194 proteins were analyzed further. B) PCA plots showing data variability in raw signal sum value of MS runs (left) and after data normalization steps (right). Color shows type of samples (intact biotinylation or lysate biotinylation), and shape of points represents each replicate.

Surface exposure level of each protein was assessed by quantitative comparison of the two samples, intact and lysate biotinylated samples. For each protein and replicate, surface exposure values (SE values) were calculated as the log fold-change of the intact to lysate sample signal ratio (logFC (intact / lysate)). SE values were calculated and averaged between replicates using the Limma method (Ritchie et al., 2015) with p-values corrected for multiple testing using the Benjamini and Hochberg method (Benjamini and Hochberg, 1995). Four biological replicates showed overall good correlations in their SE values between pairs of biological replicates (Figure 3.5A; average $R = 0.77$, lowest: $R = 0.68$ and highest: $R = 0.86$) of all replicate pairs by Pearson correlation. Figure 3.5B shows the SE value distribution for all the proteins, revealing a slightly skewed distribution towards high SE values, implying that most proteins are not surface exposed with some being surface exposed showing high SE values. The SE values for selected well-studied cell envelope proteins related well to their localization: I detected low SE values for the Sec complex (SecYEG) in the IM, and high SE values for the BAM complex (BamABCD) in the OM (Figure 3.5C). For the two complexes harboring proteins across the cell envelope, Lpt and Tam complex, the SE values differ clearly between the components localized in the OM (LptD, LptE and TamA) and in the IM (LptB, LptC, LptG, LptF and TamB) (Figure 3.5C). Validity of the SE values for protein localization identification is further investigated and discussed in the next section.

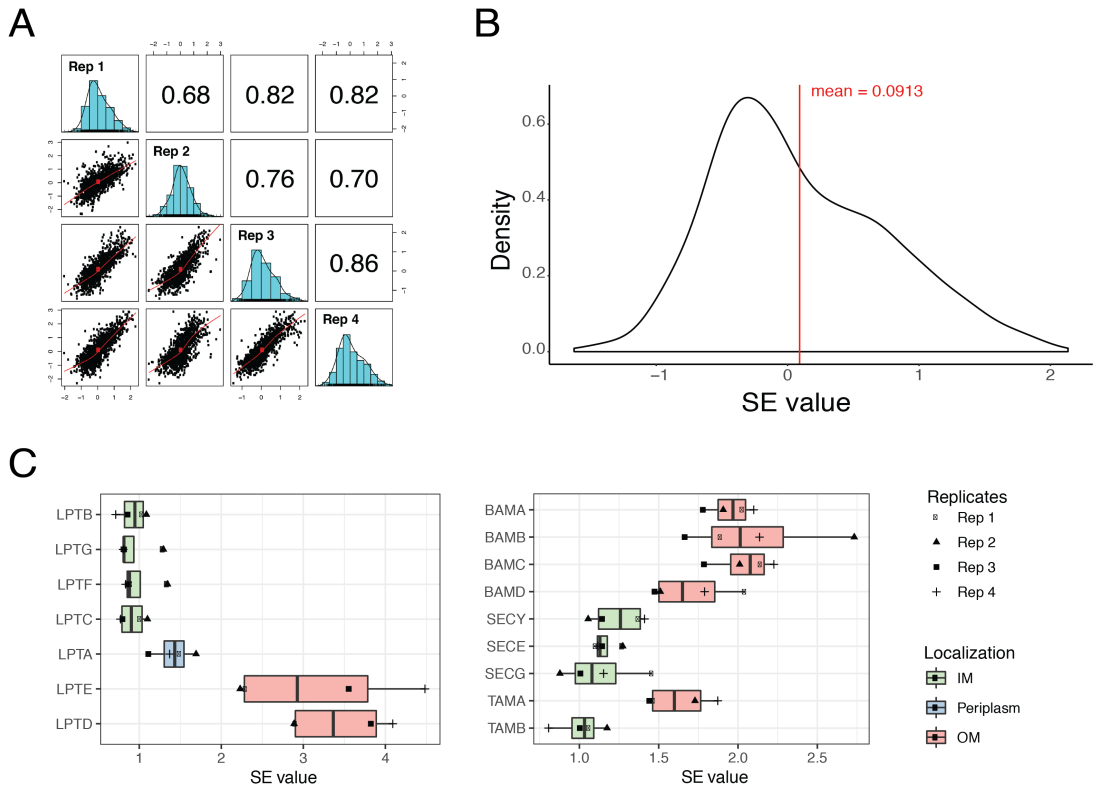


Figure 3.5: Surface exposure value quantification. A) Replicate correlations of surface exposure (SE) values were calculated for each replicate pairs. The numbers are Pearson correlations *R* value, histogram shows distribution of SE values in each replicate. B) Distribution of averaged SE values plotted as density plot, with the mean value (mean = 0.0913) shown as red vertical line. C) Examples of SE values, with all 4 replicates shown as box plot for each protein of selected cell envelope proteins - Left: Lpt complex (LptA–G), Right: Bam complex (BamABCD), Sec complex (SecYEG), and Tam complex (TamA and B). Box plots are colored according to the protein localization from STEPdb database (Orfanoudaki and Economou, 2014).

3.4 Protein localization dictates biotinylation level difference between intact cells and cellular lysate

To assess more systematically whether cell-surface proteins are enriched in intact samples compared to the lysate sample, i.e. whether they have high SE values, I compared the data with annotations from STEPdb (Sub-cellular Topologies of *E. coli* Polypeptides), a database that assembles protein localization information, both from prediction and manual curation (Orfanoudaki and Economou, 2014). All proteins quantified were divided into 9 categories, using the database annotation with some modifications (Materials and Methods) (Figure 3.6A). The SE value of each protein is plotted against average top3 values, that is average signal of three peptides with the highest detected signal levels, which can be used as a proxy for protein abundance in the MS sample (Figure 3.6A). As expected, soluble proteins are less abundant, represented by lower average top3 values, due to the membrane separation step by ultracentrifugation (Figure 3.6A, y-axis). Surface exposure (SE) values, shown as logFC

(intact/lysate) on the x-axis, revealed a gradient, which is reflective of the annotated localization: cytoplasmic proteins have the lowest SE values and SE values increase towards proteins known to be located in the OM, such as OMPs and lipoproteins (OMLPs) (Figure 3.6A (x-axis), Figure 3.6B). OMPs are surface exposed proteins based on their folding and conformation. They display β-barrel structure, where β-strands span through the OM lipids, and the connecting loops are being surface exposed (Galdiero et al., 2007). Interestingly, the higher the protein abundance (average top3 values) in the MS sample, the more distinct the localization gradient becomes (Figure 3.6A). This suggests that we can make more accurate assessments on surface exposure if the protein abundance is higher. Overall, the SE values quantified by the SPQ method represented the localization of proteins from the annotation database (Orfanoudaki and Economou, 2014), showing the method's capability to identify the localization of proteins, including surface proteins.

In contrast to what was previously reported (Cowles et al., 2011), I showed that *E. coli* OM is partially permeable to the biotin reagent, NHS-LC-LC-Biotin, thus also labelling and enriching intracellular proteins upon intact cell biotinylation. However, quantitative proteomics comparing intact cells and cellular lysate, and the short biotin incubation time allowed for identification of known surface exposed OMPs. In addition, I showed that biotinylation happens in a gradient manner from the outside towards inside of the cells, and thus the SPQ method is likely capable of quantifying not only surface exposure but also protein localization within *E. coli* cells.

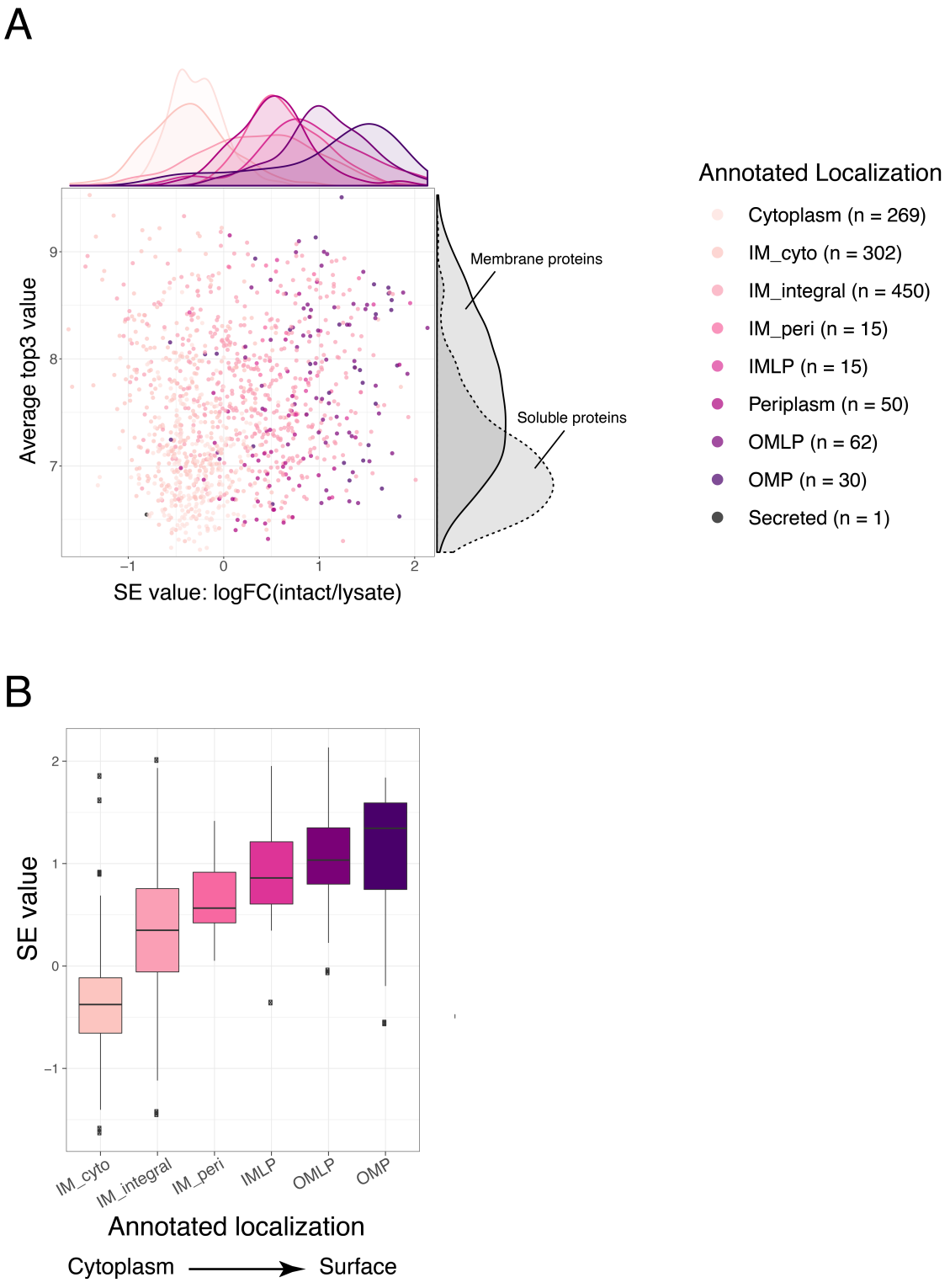


Figure 3.6: SE values compared with protein localization. A) All proteins are colored according to the localization from STEPdb database (Orfanoudaki and Economou, 2014). The scatter plot shows SE values (log fold-change (logFC) of intact/lysate ratio, x-axis) and average top3 value (average signal of three peptides with the highest detected signal levels representing the protein abundance in the MS sample, y-axis) with corresponding density plot on the sides, x-axis for localization and y-axis for solubility (dotted line for grouped soluble cytoplasmic and periplasmic proteins, and solid line for membrane bound proteins). B) Box plot represents the SE values for membrane proteins of each localization category, ordered from cytoplasm towards surface. Same colors used as in panel A are used.

3.5 OMPs are robustly identified to be surface exposed

The performance of SE values to distinguish the surface proteins from non-surface proteins was evaluated using a receiver operating characteristic (ROC) curve. To calculate the ability of the SE values to capture surface exposed proteins, OM beta-barrel proteins (OMPs) were used as positive controls of surface exposed proteins, whereas IM integral proteins were used as non-surface exposed category of proteins. Figure 3.7A displays the distribution of SE values for these two categories of proteins, suggesting there is a distinctive SE value range for each protein localization with some overlap in between. For the ROC curves, the true positive rate (TPR: identified OMPs / all OMPs quantified) and false positive rate (FPR: identified IM integral proteins / all IM integral proteins quantified) are calculated across SE values (from high to low, interval = 0.02), and plotted against each other (Figure 3.7B). The ROC curve indicates that the SE value can robustly capture most surface proteins before detecting many non-surface proteins, for example, at SE value =1.15, I detect > 60% of annotated OMPs (true-positives) with 10% of annotated IM integral proteins (false-positives) (Figure 3.7B). From this ROC curve result, I chose 80 and 90 percentile of IM integral protein SE values, thus 20% and 10% false-positive rate (FPR), to be the cutoffs to annotate the hits and candidates of surface exposed lipoproteins, corresponding to SE values of 0.87 and 1.15 (Figure 3.7B). These cutoffs were used to identify lipoproteins that are surface exposed.

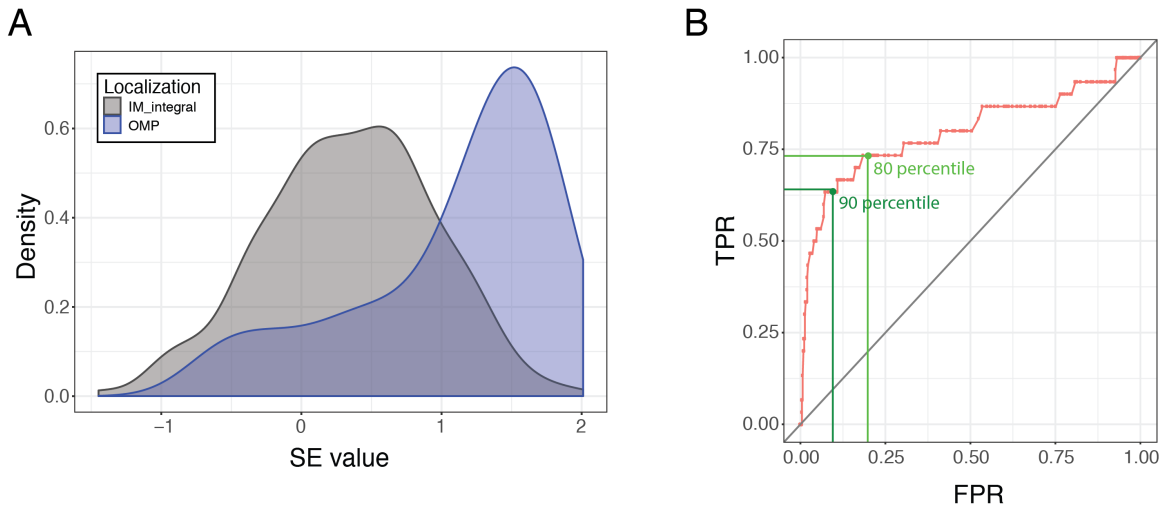


Figure 3.7: ROC curve of SE value for surface protein identification. A) Density plot showing the distribution of the SE values for OMP and IM integral proteins, IM integral proteins in grey, and OMPs in blue. B) Receiver operating characteristic (ROC) curve calculated across the SE values (increment of 0.02) showing identified OMPs / all OMPs (true positive rate, TPR) and identified IM-integral proteins / all IM-integral proteins (false positive rate, FPR). 10% FPR (90 percentile) and 20% FPR (80 percentile) are annotated.

3.6 Common occurrence of lipoprotein surface exposure is revealed

Figure 3.8A shows where the surface identification cutoffs discussed in the previous sections are for OMLPs. The peak of OMLP SE value distribution is between 80 and 90 percentiles, indicating that many lipoproteins are above and within these cutoffs. To assess the localization and SE values of each OM lipoprotein, I made a volcano plot (Figure 3.8B) comparing the SE value (x-axis) with its associated p-value (y-axis). Many OMPs show high SE values with high significance (Figure 3.8B), supporting the method's ability to identify surface exposed proteins. Surprisingly, many lipoproteins have similar or higher SE values with high significance relative to OMPs, and higher than those of known surface lipoproteins such as Lpp and RcsF (Figure 3.8B, labelled in red). This suggests that many more lipoproteins are on the surface than previously reported, and to the extent that lipoprotein surface exposure can be considered common. With the p-value cutoffs of < 0.01 for hit and < 0.05 for candidate, there were 23 OMLPs above the 90 percentile (identified as hits), and another 20 OMLPs above the 80 percentile (identified as candidates) (Figure 3.8B and Table 8). As there are in total only 62 OMLPs identified in the assay, this implies that the surface exposure of lipoproteins is a general phenomenon, with possibly more than 50% of lipoproteins being surface exposed.

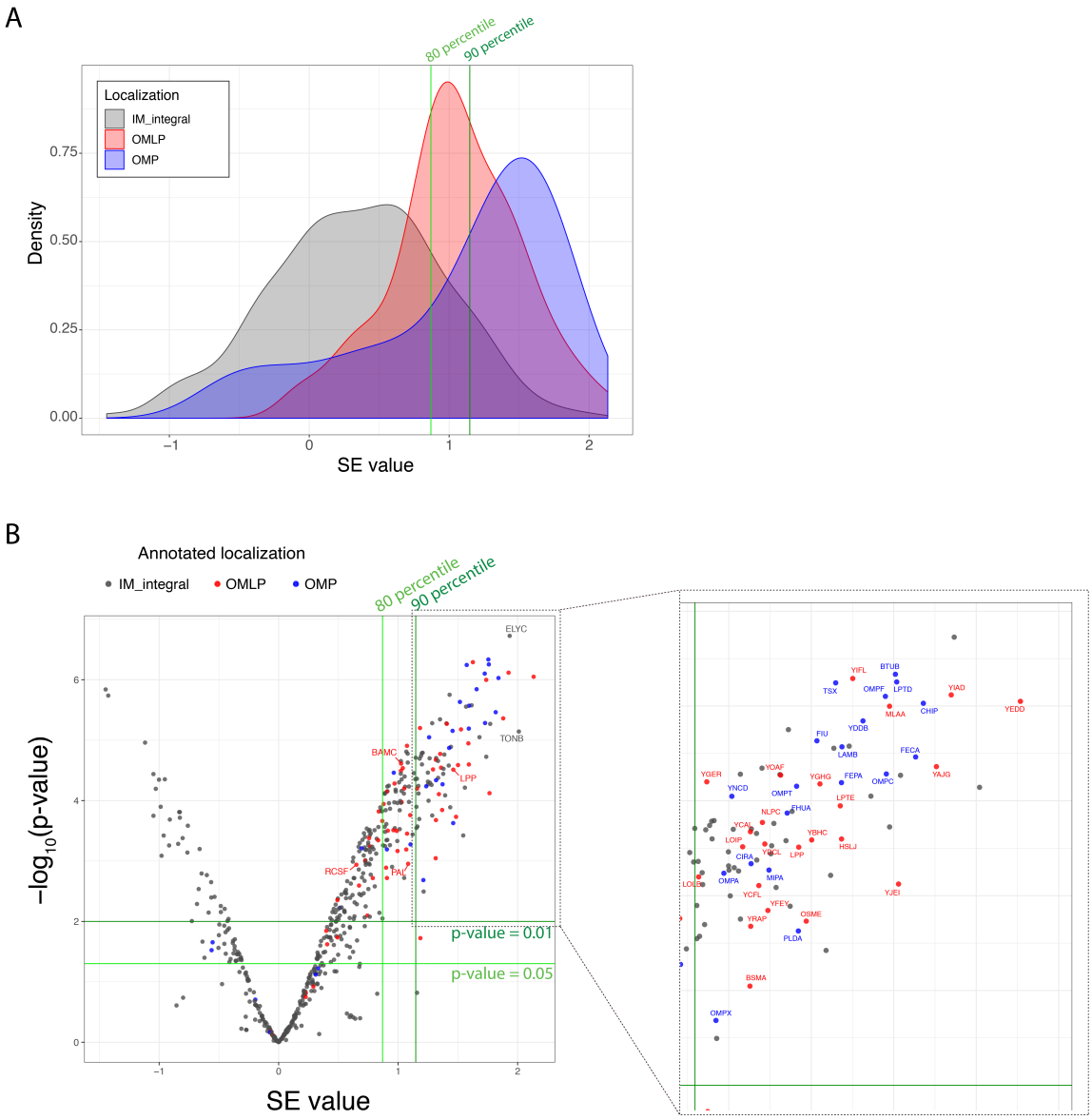


Figure 3.8: Identification of surface exposed lipoproteins. A) Distribution of SE values plotted as a density plot for OMLPs, OMPs and IM integral protein, IM 90 and 80 percentile cutoffs are annotated (as green vertical lines), B) Volcano plot showing SE value (x-axis) and p-value (y-axis). Proteins annotated with the following categories were plotted: IM integral (grey), OMLP (red), OMP (blue). Four previously reported surface exposed lipoproteins are also annotated in the plot (Lpp, BamC, Pal and RcsF). Above 90 percentile of IM integral distribution and p-value < 0.01 are zoomed in on the right top, annotating gene names for OM localized proteins (OMLPs are listed in Table 8).

3.7 Inner membrane proteins with high surface exposure identified

As showed in the density plot (Figure 3.8A) and the volcano plot (Figure 3.8B), some IM integral proteins have high SE values. These are likely false positives, i.e. identified as surface exposed by error. However, there is also a possibility that this assumption is partly wrong and some IM proteins are misannotated as IM or annotated correctly but can reach the surface from the IM. Regarding lipoproteins, two IM lipoproteins, MetQ and DcrB show high surface exposure values (MetQ = 1.95, DcrB = 1.57, both above 90 percentile). These proteins are thus included in the further analysis alongside with other OM proteins. Some other non-OM proteins, such as YgaU (IM-cyto, SE value = 1.86), ElyC and TonB (IM integral proteins, Figure 3.8B) also show high surface exposure values, above 90 percentile. These proteins are also analyzed and discussed later in Chapter 5 and in the Discussion Chapter.

3.8 Independent methods validated surface localization of five novel surface lipoproteins

I sought to validate the surface exposure of lipoproteins using independent methods. In order to investigate each lipoprotein for surface exposure, I tagged the C-terminus of lipoproteins identified as surface exposed, as well as a control proteins, with two glycine residues followed by a single Flag epitope tag (GG-DYKDDDDK) in their chromosomal locus (see Materials and Method). Nine proteins (8 lipoproteins and 1 periplasmic control protein, SurA) were successfully tagged with identifiable level of expression on Western blot by α-Flag antibody (data not shown). Flag-tagged strains were investigated using following two methods: dot blot and whole-cell ELISA (enzyme-linked immunosorbent assay), both aiming to detect α-Flag antibody binding on the surface of the cells.

Dot blot has been previously used as a method to show surface exposure of RcsF (Cho et al., 2014; Konovalova et al., 2014). Here, mid-log cultures were harvested and spotted onto a nitrocellulose membrane, with and without incubation in a buffer permeabilizing the OM to let the antibody through the OM. Intact cells and permeabilized cells on the membrane were incubated with antibodies. WT cells without Flag tagged proteins were tested as background controls of the α-Flag antibody for dot blot (Figure 3.9). A weak signal was observed for permeabilized cells, showing that there is an unspecific binding of the antibody (background),

but for intact cells, the background was very little or not detectable (Figure 3.9). The dot blot confirmed the surface exposure of three lipoproteins with Flag tag: MetQ, OsmE and DcrB, through their signals in intact cells (Figure 3.9). Antibody against AcrA was used as a control to show the surface specificity of the assay, as AcrA is a bona-fide IM lipoprotein and consistently a non-hit in the SPQ assay. AcrA signal was detected well upon cell permeabilization, but not when cells were intact (Figure 3.9). YedD, the OMLP with the highest SE value, was not shown to be surface exposed in the dot blot assay using antibody against Flag tag (Figure 3.9). However, Jean-Francois Collet (UCL, Belgium) and his colleagues have preliminary data of successful dot blot for YedD using an antibody raised for YedD (unpublished data, from personal communication), supporting the surface exposure of YedD. RcsF surface exposure has been previously shown by dot blot using an α-RcsF antibody, but I failed to detect a C-terminally Flag tagged RcsF in my experiment using the α-Flag antibody. This is consistent with a previous report showing that C-terminally Flag tagged RcsF cannot be detected by dot blot, presumably due to occlusion (Konovalova et al., 2014). These observations emphasize the limitation of the dot blot approach using a C-terminus tag. Despite the major limitation of the assay, i.e. that the tag needs to be surface exposed and accessible for the antibody to reach it, these three lipoproteins (MetQ, OsmE and DcrB) were verified to be surface exposed.

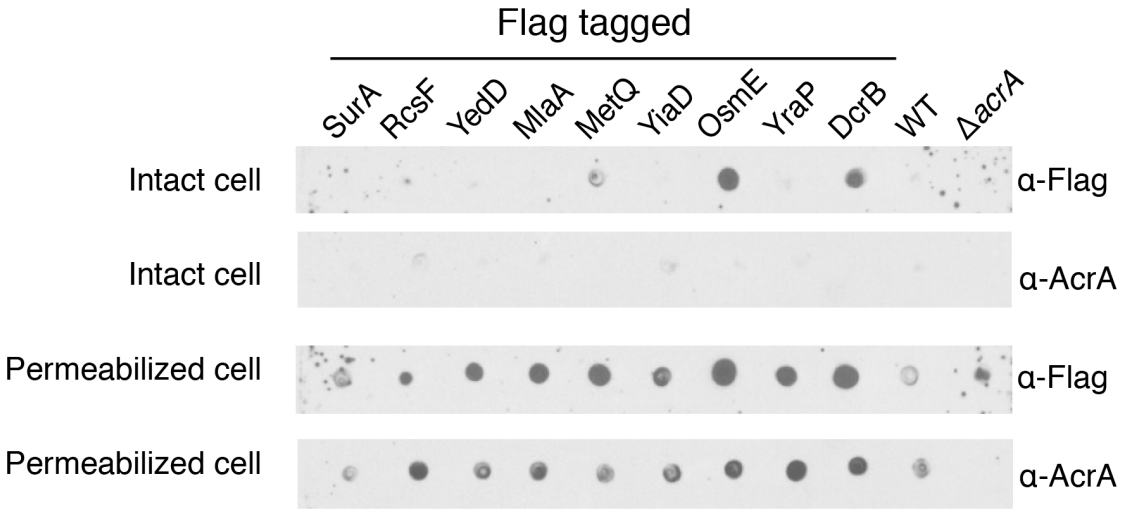


Figure 3.9: Validation of lipoprotein surface exposure by dot blot. 2 µL of cells (concentrated to OD_{578} = 1) was spotted onto nitrocellulose membrane, with and without incubation with permeabilizing buffer (with lysozyme and EDTA). Dot blot assay was performed using α-Flag and α-AcrA antibody (IM control). Studied Flag tagged proteins are: SurA, RcsF, YedD, MlaA, MetQ, YiaD, OsmE, YraP, and DcrB. WT strain without Flag tag and Δ*acrA* strain were used as background controls for respective antibodies. One technical replicate (out of 3) of one biological replicate (out of 3) is depicted. Other replicates are found in Appendix.

Whole-cell ELISA uses the same principle; measuring antibody binding to the surface of an intact cell. This method has been used to show surface exposure of BamC (Webb et al., 2012), and I used a modified version of this method. Wells of 96-well plates coated with cells were incubated with a primary antibody and a secondary antibody conjugated with HRP. The method was modified slightly, and I measured and analyzed the kinetics of absorbance at 405 nm for up-to 30 minutes. WT cells not containing a Flag tag, and a strain with Flag-tagged periplasmic protein SurA were used as negative controls, showing very little reactivity (Figure 3.10A). Similarly, low peroxidase reactions were observed for wells incubated with α-mouse secondary antibody only, i.e. without α-Flag primary antibody. The whole-cell ELISA method showed significant increase in HRP reaction for two strains with Flag-tagged lipoproteins: MlaA and OsmE, validating their surface exposure (Figure 3.10A). These two strains displayed clear signal in all three biological replicates tested. In addition, YedD surface exposure was shown by ELISA using α-YedD antibody, by comparing WT cells and Δ*yedD* cells (Figure 3.10B). α-AcrA antibody was used as a control to show that there is no difference between WT and Δ*acrA,* as AcrA is not surface exposed. Both WT and Δ*acrA* showed high background reactivity, further confirming the surface specificity of the method and YedD antibody (Figure 3.10C). Thus three lipoproteins, MlaA, OsmE and YedD were shown to be surface exposed using whole-cell ELISA.

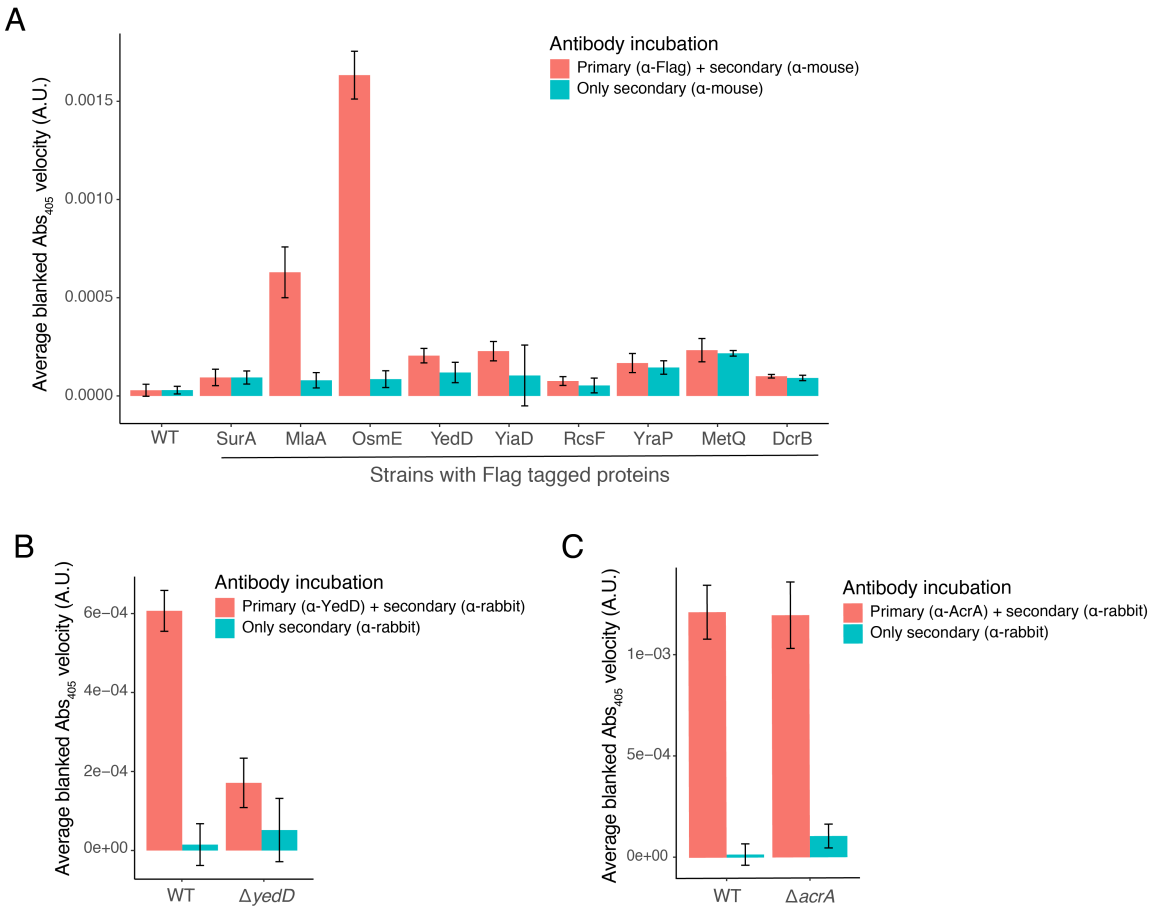


Figure 3.10: Validation of lipoprotein surface exposure by whole-cell ELISA. ELISA plates were coated with mid-log cells, incubated with primary and secondary antibody or only with secondary antibody (control). Non-coated wells were used as blank. A) ELISA for Flag antibody detection. Tested strains are WT (negative control without Flag) and strains with different proteins being tagged by a single Flag tag (SurA, MlaA, OsmE, YedD, YiaD, RcsF, YraP, MetQ, DcrB). Cells incubated with both primary α-Flag and α-mouse-HRP secondary (orange) and only with secondary antibody (green). Bar height indicates mean velocity of Abs405 from which baseline values from blank wells without cells was subtracted. Error bars denote one standard deviation from the mean from three technical replicates (one biological replicate). For OsmE and MlaA, the high velocity relative to background (secondary antibody only) and WT / Flag-tagged SurA strain, was shown in another two independent experiments (Appendix). B) Same as A) using α-YedD antibody and α-rabbit-HRP antibody for WT and Δ*yedD* strain, C) using α-AcrA antibody and α-rabbit antibody for WT and Δ*acrA*, for negative control for panel B. B) and C) were only performed for one replicate. Overall, ELISA revealed surface localization of three lipoproteins, MlaA, OsmE (A) and YedD (B).

Another common method of showing surface exposure is surface shaving using proteases, such as Proteinase K, including a recent study showing extensive surface exposure of lipoproteins in *Borrelia burgdorferi* (Dowdell et al., 2017). I have tried this method, and optimized the method to obtain successful surface shaving of OMPs, BamA and OmpC (Figure 3.11). However, none of the lipoproteins tested showed reproducible significant degradation (data not shown). The reason behind the unsuccessful Proteinase K shaving of *E. coli* surface lipoproteins will be discussed in Discussion Chapter.

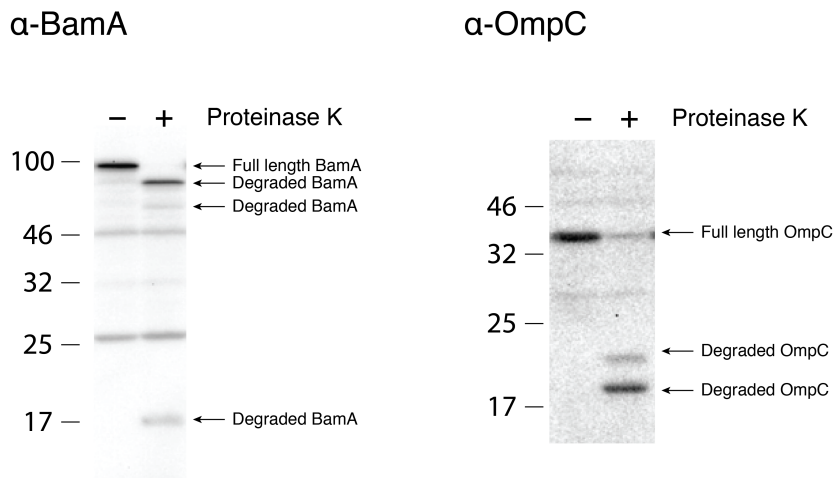


Figure 3.11: Proteinase K surface shaving of OMPs. Mid-log culture of *E. coli* cells were sampled, concentrated to OD_{578} = 10, and 80 µL used for treatment with and without Proteinase K (PK). PK was added and incubated for 1 hour at 37 degrees Celsius. PK reaction was stopped by TCA precipitation, and precipitated pellets were solubilized prior to SDS-PAGE analysis (using 15% gels). Left: Western blot for α-BamA, Right: Western blot for α-OmpC. Biological replicate n=1.

Overall, the SPQ method using surface biotinylation described here could successfully identify intracellular protein localization, including surface localization of control OMPs and also yielding similarly high SE values for many lipoproteins. The result obtained by this method indicates lipoprotein surface exposure to be a common phenomenon in *E. coli*, with 5 (DcrB, MetQ, MlaA, OsmE and YedD) out of 8 putative surface identified lipoproteins confirmed in at least one of the two independent methods.

Table 8: Hits and candidates of surface exposed lipoproteins by the SPQ method

SE value and its associated p-value is shown for hit and candidate OMLPs by the SPQ method. Proteins are categorized into SE class (above 90 percentile: > 90 , or above 80 percentile: > 80) and p-value class (< 0.01 or < 0.05). Hit class shows if the protein is identified as a hit (>90 in SE class and < 0.01 for p-value) or candidate (>80 in SE class and <0.05 for p-value). List is ordered by SE value (from high to low). OMLPs tagged by a Flag tag and thus used in dot blot and whole-cell ELISA are annotated with (*).

	Name	SE value	p-value	SE class	p-value class	Hit class
1	YEDD*	2.134	8.93E-07	> 90	< 0.01	hit
2	YIAD*	1.924	7.64E-07	> 90	< 0.01	hit
3	YAJG	1.879	4.36E-06	> 90	< 0.01	hit
4	YJEI	1.765	7.54E-05	> 90	< 0.01	hit
5	MLAA*	1.737	1.01E-06	> 90	< 0.01	hit
6	YIFL	1.626	5.13E-07	> 90	< 0.01	hit
7	HSLJ	1.592	2.53E-05	> 90	< 0.01	hit
8	LPTE	1.588	1.13E-05	> 90	< 0.01	hit
9	YGHG	1.526	6.63E-06	> 90	< 0.01	hit
10	YBHC	1.501	2.58E-05	> 90	< 0.01	hit
11	OSME*	1.485	1.86E-04	> 90	< 0.01	hit
12	LPP	1.462	3.08E-05	> 90	< 0.01	hit
13	YOAF	1.407	5.36E-06	> 90	< 0.01	hit
14	YFEY	1.368	1.43E-04	> 90	< 0.01	hit
15	YDCL	1.359	2.85E-05	> 90	< 0.01	hit
16	NLPC	1.352	1.69E-05	> 90	< 0.01	hit
17	YCFL	1.341	7.83E-05	> 90	< 0.01	hit
18	YRAP*	1.316	2.10E-04	> 90	< 0.01	hit
19	YCAL	1.316	2.12E-05	> 90	< 0.01	hit
20	BSMA	1.314	8.99E-04	> 90	< 0.01	hit
21	LOIP	1.292	3.05E-05	> 90	< 0.01	hit
22	YGDI	1.186	1.89E-02	> 90	< 0.05	candidate
23	YGER	1.183	6.30E-06	> 90	< 0.01	hit
24	LOLB	1.159	6.34E-05	> 90	< 0.01	hit
25	YBAY	1.101	1.74E-04	> 80	< 0.01	candidate
26	PAL	1.084	1.11E-03	> 80	< 0.01	candidate
27	YFIL	1.075	3.53E-04	> 80	< 0.01	candidate
28	YCEB	1.074	1.24E-05	> 80	< 0.01	candidate
29	YAJI	1.068	6.44E-04	> 80	< 0.01	candidate

30	BAMB	1.050	6.33E-05	> 80	< 0.01	candidate
31	YCJN	1.039	2.91E-05	> 80	< 0.01	candidate
32	YEGR	1.029	1.06E-04	> 80	< 0.01	candidate
33	BAMC	1.025	2.44E-05	> 80	< 0.01	candidate
34	YEHR	1.023	3.22E-05	> 80	< 0.01	candidate
35	MLIC	0.995	6.83E-04	> 80	< 0.01	candidate
36	LPOB	0.988	3.18E-04	> 80	< 0.01	candidate
37	YDDW	0.970	5.23E-05	> 80	< 0.01	candidate
38	NLPI	0.960	3.09E-04	> 80	< 0.01	candidate
39	MLTA	0.914	3.15E-04	> 80	< 0.01	candidate
40	BLC	0.906	1.91E-03	> 80	< 0.01	candidate
41	EMTA	0.905	7.15E-05	> 80	< 0.01	candidate
42	RLPA	0.899	1.29E-03	> 80	< 0.01	candidate
43	YEAY	0.879	1.14E-04	> 80	< 0.01	candidate

Chapter 4: Investigating the mechanisms behind lipoprotein surface exposure

4.1 Background

In the Chapter 3, I showed that the SPQ (surface proteome quantification) method, surface biotinylation combined with quantitative MS, can identify surface localization of proteins, including surface exposed lipoproteins. Upon the observation of the high occurrence of lipoprotein surface exposure in *E. coli*, the next question to address was how those lipoproteins reach the cell-surface. To look into the molecular mechanisms of lipoprotein surface exposure, I applied the SPQ method to examine lipoprotein surface exposure of *E. coli* in different conditions, such as genetic mutants and different growth phases.

To date, there are no known machineries for lipoprotein transport to the cell surface in *E. coli*, except for the OMLP RcsF (Cho et al., 2014). BamA is crucial for RcsF surface translocation, and also some major OMPs, such as OmpA, OmpC and OmpF, were shown to be needed for the surface translocation of RcsF, as it forms a complex with OMPs where RcsF is embedded in the lumen of OMPs (Cho et al., 2014; Konovalova et al., 2014). Additionally, in *Borrelia burgdorferi*, deficiency of BamA homolog has been shown to lower lipoprotein surface exposure (Lenhart and Akins, 2010; Zückert, 2014). Therefore, I hypothesized that BamA is required for lipoprotein surface exposure in *E. coli* for many surface lipoproteins.

As BamA functions as a part of the BAM complex, essential for OM biogenesis, I assessed the impact of mutating members of the BAM complex or BAM-related factors, such as the downstream stress response when BAM malfunctions, on lipoprotein surface exposure (Figure 4.1). In addition to exponentially growing cells used in Chapter 3, I investigated stationary phase cells to examine the validity of the SPQ method in a different growth phase, and to examine whether lipoprotein surface exposure changes in a different growth phase. This chapter explores the results of the SPQ method applied to different mutants and conditions, enhancing our understanding of lipoprotein surface exposure in *E. coli*.

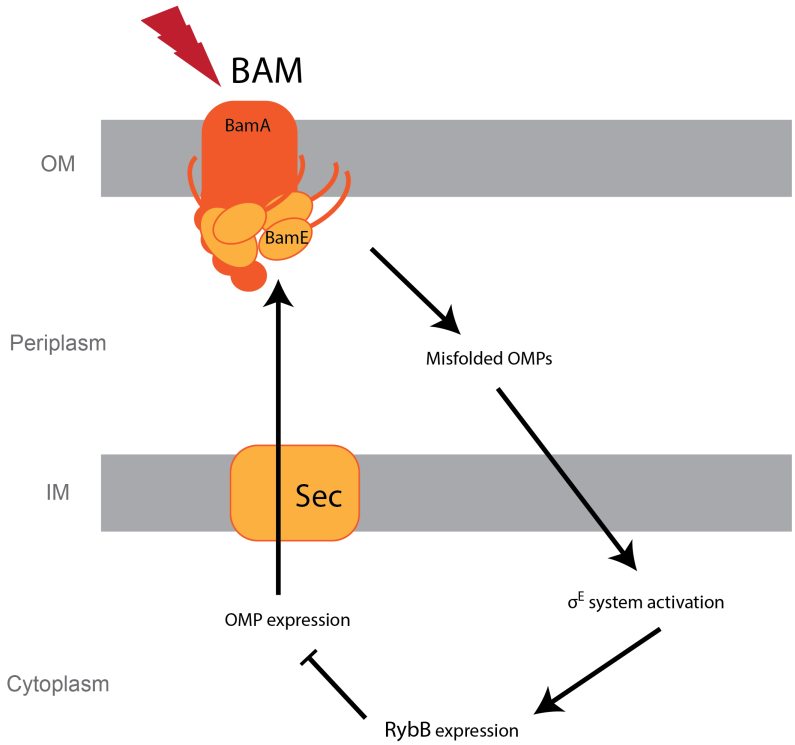


Figure 4.1: Illustration of mutants used in the experiment. BAM machinery components, BamA and BamE in the OM. Upon defects in BAM functionality, misfolded or unfolded OMPs accumulate in the periplasm, and the σ^E system is activated, leading to expression of RybB (a small RNA) which inhibits the expression of OMPs, lowering the load for the BAM machinery.

4.2 Lipoprotein are less surface exposed in *bamA* mutants

As BamA is an essential protein, I used mutants of BamA to investigate its role in lipoprotein surface exposure. Firstly, *bamA101,* a strain expressing BamA at ~10-fold lower level (Aoki et al., 2008), and secondly *bamA6,* a strain expressing a BamA mutant with a two amino acid insertion previously shown to slightly lower OMP levels (Ruiz et al., 2006), were investigated. In the *bamA101* strain, the level of DegP is increased as a result of OMP folding defects (Aoki et al., 2008), and it has also been shown to activate the Rcs signaling system, suggesting RcsF surface translocation is defective in this strain (Cho et al., 2014).

The *bamA101* strain is known to have increased sensitivity to antibiotics, such as vancomycin, due to its increased OM permeability (Ricci et al., 2012). Thus, I examined the biotinylation pattern by comparing the membrane and soluble fractions, as performed in Chapter 3 for the WT strain, to check if increased permeability affects surface specificity of the label. A Western blot against biotin showed no increase in biotinylation of soluble proteins in the *bamA101* strain compared to WT (Figure 4.2). Furthermore, there is less biotinylation of the membrane fraction, suggesting there are fewer proteins on the surface to be labelled by the biotin (Figure 4.2). As

membrane specificity of the label was maintained in the *bamA101 strain*, showing its compatibility for the surface biotinylation method, the *bamA101* strain was assayed using the SPQ method for systematic identification of surface exposed proteins.

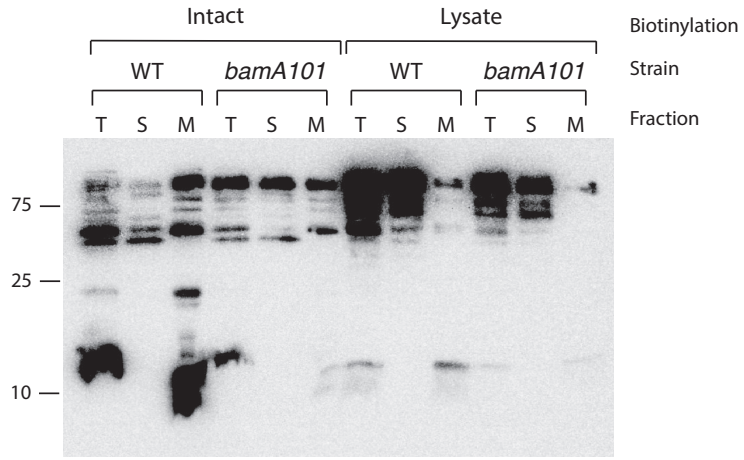


Figure 4.2: Western blot result comparing biotinylation pattern of WT and *bamA101.* Streptavidin-HRP (against biotin) was used. For fractions, T: total cell lysate, S: soluble fraction, M: membrane fraction, with two different biotinylation conditions (intact and lysate). For lysate samples, x10 diluted samples compared to intact biotinylation were loaded. 15% SDS-PAGE gel was used, biological replicate n=1.

The SPQ method was applied for *bamA101* and *bamE* deletion strain (three biological replicates for each) as described for WT in Chapter 3 and Materials and Methods. Together with the results of the SPQ assay performed for WT (Chapter 3), the MS results were normalized and quantified for surface exposure values (SE values, logFC (intact / lysate biotinylation)). I used the localization annotation from STEPdb database (Orfanoudaki and Economou, 2014), and three major categories of membrane proteins, IM integral, OMP and OMLP were compared for their SE values (Figure 4.3A). Firstly, I compared the *bamA101* strain to WT. The density plot of SE value distribution shows that in contrast to WT, the distribution of SE values for all the three categories (including the OM and IM) overlap more with closer density curve peaks (Figure 4.3A). Especially for OMLP, the peak of SE value distribution is almost overlapping with IM integral protein, showing decreased lipoprotein surface exposure in *bamA101* (Figure 4.3A). In the case of OMPs, although I observed changes in distribution of SE values, some OMPs keep their high SE values (Figure 4.3A). One possible explanation is that the cells have strong homeostatic mechanisms which adjust the levels of OMPs in response to lowered level of the BamA. Accumulation of unfolded OMPs in the periplasm is toxic, thus the cells either down-regulates their expression or degrades them by periplasmic proteases (Mitchell and Silhavy, 2019). This results in the SE value remaining relatively unchanged, as the SE value takes the lysate value into account and OMPs are reduced in both: on the surface and in the

lysate. For OMPs showing changes in SE values, this buffering mechanism is presumably not enough, or other factors, such as protein conformations may also be altered (Figure 4.3A). In conclusion, I observed a significant effect of BamA knock-down in the *bamA101* strain on lipoprotein surface exposure, supporting the hypothesis that BamA is generally required for lipoprotein transport to the surface.

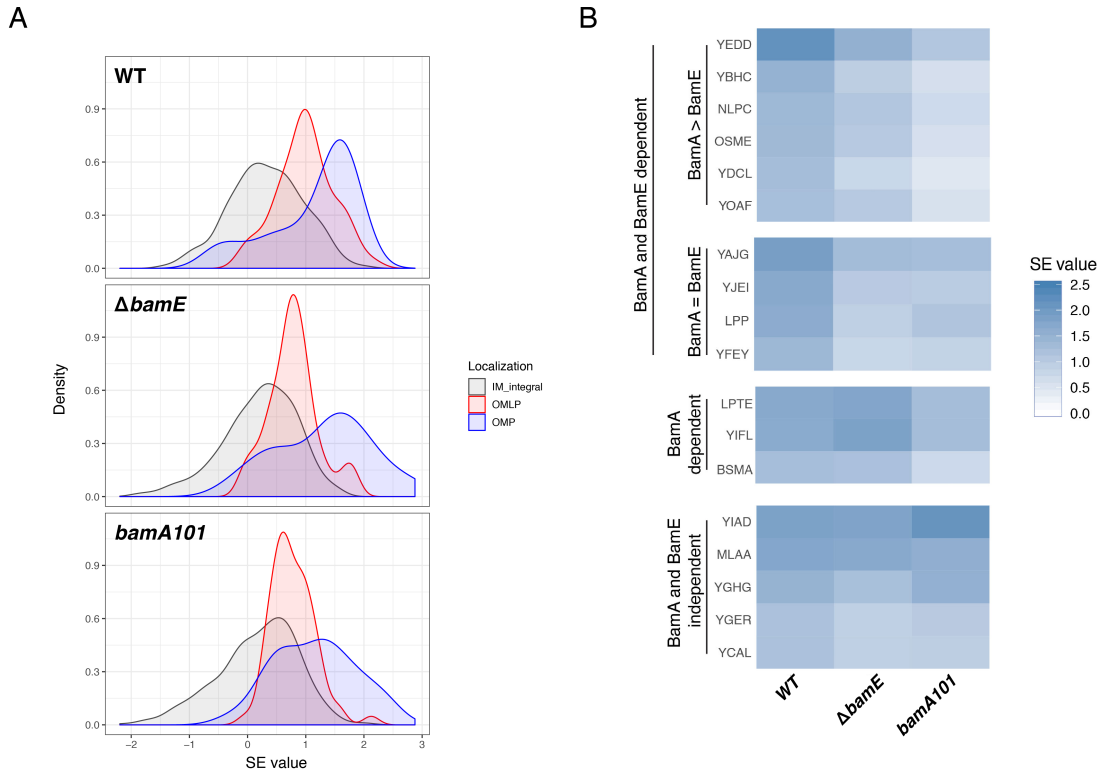


Figure 4.3: SE values of membrane proteins in BAM mutants compared to WT. A) Distribution of SE values of proteins in three protein localization categories (IM integral, OMLP and OMP), of three strains shown as density plot, top: WT, middle: Δ*bamE*, and bottom: *bamA101*. B) Heatmap of SE values of OMLPs across the three strains (shown in panel A) for 18 proteins (from 20 most-surface exposed lipoproteins in WT) which clustered for BAM dependency, only BamA, both BamA and BamE, or BamA and BamE independent. Clustering and full heatmap are found in Appendix. The data plotted for all three strains tested denotes the average SE value calculated from three independent biological replicates.

bamA6 (formally referred as *yaeT6*), harbors a two amino acid duplication mutation within BamA, originally isolated from a suppressor mutation in an LptD mutant (Ruiz et al., 2006). A chemical genomics study has shown increased sensitivity of *bamA6* for various chemicals, such as vancomycin, suggesting the a possible defect in the OM and cell envelope (Nichols et al., 2011). In terms of OMP biogenesis, the observed effect was rather mild, an approximate 10 % reduction in folding of the two OMPs (LamB and OmpA), and thus, I used *bamA6* as a less severe BAM defect mutant compared to that of *bamA101* (Ruiz et al., 2006). Comparing the

surface exposure of OMLPs, OMPs and IM integral proteins, *bamA6* exhibit similar patterns to WT (Figure 4.4). Therefore, I conclude that the BamA mutation in *bamA6* is insufficient to affect surface exposure difference in OM proteins (both OMPs and OMLPs). Thus, only one replicate was performed and removed from further analysis.

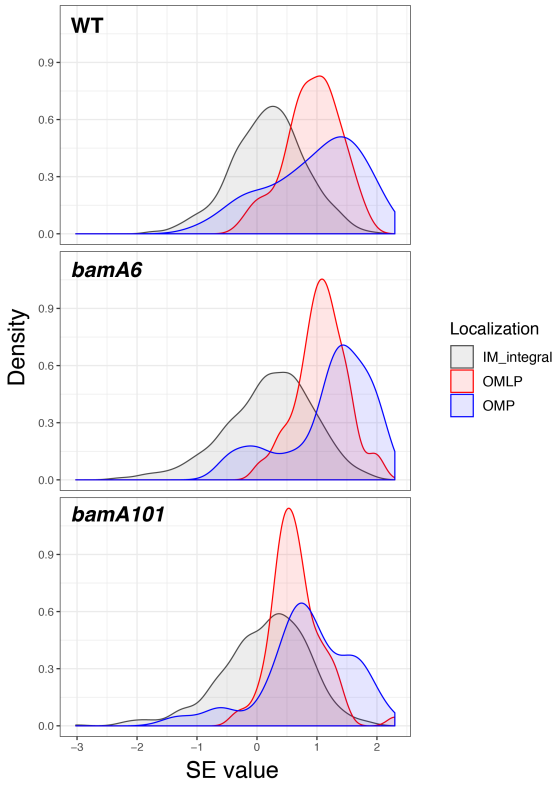


Figure 4.4: SE values of membrane proteins in 2 *bamA* mutants compared to WT. Distribution of SE values of proteins in three protein localization categories (IM integral, OMLP and OMP), of three strains shown as density plots. Top: WT, middle: *bamA6*, bottom: *bamA101*. The result of all three strains are from a different MS run from Figure 4.3, biological replicate n=1.

4.3 Lipoproteins are less surface exposed in *bamE* deletion mutant

Although BamA is the central component in the BAM complex, the complex is comprised of four more components, one essential (BamD) and three non-essential (BamB, BamC, BamE) lipoproteins (Konovalova et al., 2017). Recent studies regarding RcsF translocation have revealed that BamE, an accessory lipoprotein of BAM machinery, is key for RcsF translocation to the surface (Hart et al., 2019; Konovalova et al., 2016; Tata and Konovalova, 2019). The functionality of BamE has not been fully revealed, mainly because deletion of *bamE* does not create a severe growth nor OMP folding defects (Rigel et al., 2012). However, an increase in the RcsF-BamA interaction was observed upon *bamE* deletion, through cross-linking

experiments, while the RcsF-OmpA interaction decreased in the same mutant (Konovalova et al., 2016). Additionally, very recent studies suggested RcsF being "jammed" in the BAM complex in cells lacking BamE, leading to a defect in BAM functionality which creates synthetic lethal phenotypes with other BAM mutants (Hart et al., 2019; Tata and Konovalova, 2019). As RcsF is surface exposed forming a complex with an OMP, these observations suggest that deletion of *bamE* perturbs the surface exposure of RcsF (Hart et al., 2019; Tata and Konovalova, 2019). Thus, I hypothesized that BamE may be involved in surface transport of many lipoproteins, as I observed for BamA using *bamA101*. Since BamE is not essential, a chromosomal deletion mutant (Δ*bamE*) was investigated using the SPQ method.

Comparing the SE values of IM integral, OMP and OMLP obtained for the Δ*bamE* strain with WT and *bamA101*, I could show that Δ*bamE* has the intermediate behavior of lipoprotein surface exposure between WT and *bamA101* (Figure 4.3A). Many OMLPs show SE values similar to that of IM integral proteins, suggesting that they are no longer surface exposed in the Δ*bamE* strain. Whereas some OMLPs seem to maintain their WT level of surface exposure, depicted as a shoulder of the density plot on the high end of SE value, thus suggested to be surface translocated in a BamE independent manner (Figure 4.3A). Deletion of *bamE* did not drastically affect the surface exposure of OMPs (Figure 4.3A), consistent with previous studies on BamE that its role in OMP assembly is limited (Rigel et al., 2012). However, I observed a general and strong effect of the *bamE* deletion on OMLP surface exposure. These observations suggest a potentially distinct role of BamE in BAM functionality, relating to lipoprotein transport to the surface as suggested for RcsF (Hart et al., 2019; Tata and Konovalova, 2019).

To further investigate the roles of these BAM components for lipoprotein surface exposure, SE values of each lipoprotein were investigated across the three strains: WT, *bamA101* and Δ*bamE.* The 20 lipoproteins with the highest SE values in WT cells were clustered for changes in SE values in the *bamA101* and Δ*bamE* strains compared to the WT. Figure 4.3B depicts the SE values for 18 lipoproteins among the 20 lipoproteins which showed qualitatively distinct clustering (see Appendix for clustering and full heatmap). There were two clusters, for 10 lipoproteins in total including YedD and YajG, that showed lowered SE values in both *bamA101* and Δ*bamE,* revealing their surface exposure is dependent on both BamA and BamE (Figure 4.3B). In contrast, 3 lipoproteins (LptE, YifA, BsmA) were only dependent on BamA but not BamE, implying their translocation happens through the BAM machinery, but in a BamE independent manner (Figure 4.3B). Interestingly, 5 out of the 20 lipoproteins, such as

YiaD, showed little or no decrease in SE values in *bamA101* and Δ*bamE*, indicating their BAM-independent surface exposure. A previous study also showed that Wza and CsgG, two OMLPs, form a channel through the OM in a barrel structure, having part of their domains surface exposed, in a BamA-independent manner (Dunstan et al., 2015). However, these proteins were not detected in my assay, most likely due to a low abundance in the condition used and/or accessibility of the biotin. Overall, the involvement of the BAM complex members, not just BamA but also BamE, were shown for lipoprotein surface exposure. Additionally, my data suggests that BamE is more important for lipoprotein surface exposure than for OMP folding, the primarily and most-studied function of the BAM.

4.4 OMP depleted cells overexpressing RybB was tested for lipoprotein surface exposure

Upon the observation of OMLPs being less surface exposed in BAM mutants, I investigated how BAM is involved in lipoprotein surface transport. I had two hypotheses explaining the decrease in lipoprotein surface exposure observed in *bamA101*: (i) a direct effect of BamA depletion or (ii) an indirect effect mediated through other components. In the first hypothesis, BamA directly interacts with its substrate lipoproteins and transports them to the surface, therefore, the abundance of BamA modulates the level of lipoprotein surface exposure accordingly. This direct interaction of BamA to a lipoprotein facilitating lipoprotein surface exposure has been shown for RcsF (Cho et al., 2014; Konovalova et al., 2014). In the second hypothesis, the effect of BamA depletion is more indirect, i.e. there are other proteins defective in *bamA101* which are required for the surface exposure of lipoproteins. Obvious such proteins are OMPs, folded and inserted in the OM by BAM, and in the case of RcsF, OMPs are used by RcsF to reach the surface. Furthermore, the *bamA101* mutant is known to activate cell-envelope stress response systems, including the σ^E response system which lowers levels of many OMPs upon the system activation (Gogol et al., 2011; Mahoney et al., 2016). When the σ^E response system is activated, cells express among others two small RNAs (RybB and MicA) which inhibit the translation of multiple proteins, including several OMPs (Gogol et al., 2011). RybB has been shown to bind to mRNAs of some OMPs including OmpA, OmpC and OmpF, the most abundant OMPs (Gogol et al., 2011). Therefore, I aimed to test whether an artificial decrease of OMP expression would result in decreased lipoprotein surface translocation, even when the BAM machinery is intact. To address this, I used the cells overexpressing RybB on a

plasmid (pFM1-1, (Bouvier et al., 2008)) upon 1 mM IPTG addition, and also used the empty vector plasmid (pJV300, (Pfeiffer et al., 2007)) as control. Once the cells were harvested, the SPQ method was employed to investigate lipoprotein surface exposure.

The SPQ results revealed that the control sample with empty plasmid (pJV300) showed a different pattern compared to previous measurements of WT, with OM proteins not showing a clear increase of SE values, relative to IM proteins (Figure 4.5A). Furthermore, there is no clear difference of the SE value distributions between the control and RybB overexpression (Figure 4.5A). Although the precise reason is unclear, unsuccessful identification of surface proteins may be due to the presence of ampicillin and IPTG during the cell growth. Comparison of lysate samples between the empty and RybB overexpressing plasmid showed reduced levels of major OMPs, such as OmpA, OmpC and OmpF (Figure 4.5A), confirming the expected effect of RybB expression as previously reported (Gogol et al., 2011). Since this result represents preliminary data from one replicate, additional replicates and further technical optimization are required to fully investigate the effect of RybB overexpression, i.e. OMP depletion, on lipoprotein surface exposure.

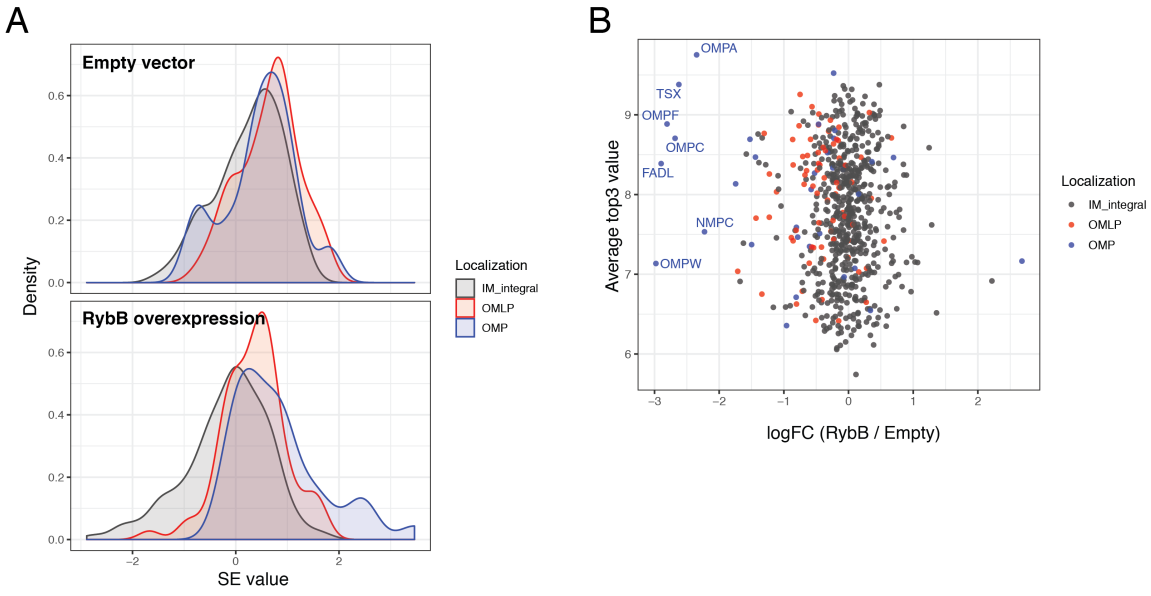


Figure 4.5: SE value comparison of strains with empty vector and RybB overexpressing plasmid. A) Distribution of SE values of proteins in three protein localization categories (IM integral, OMLP and OMP), with induction of two plasmids shown as density plot, top: empty vector (pJV300), low: RybB overexpression (pFM1-1). B) LogFC of the lysate biotinylation measurement of the two plasmids in A). Proteins below logFC < -2, thus lowered in abundance by RybB overexpression, are annotated (7 OMPs). Biological replicate n=1.

4.6 Lipoproteins are more surface exposed in stationary phase

Exponential phase cells, which are fast growing and actively dividing, are used more often to study the cellular physiology of *E. coli*. However, bacterial cells found in their natural environment, such as soil or inside host organisms, are often limited in nutrients and thus slower in growth, more similar to the stationary phase when grown in laboratory conditions. Furthermore, in stationary phase, the σ^E response system is known to be upregulated (Costanzo and Ades, 2006), suggesting that the functionality of BAM machinery may differ in the stationary phase. Additionally, the Rcs phosphorelay system, which is activated via perturbed transportation of RcsF to the surface, is also activated in stationary phase (unpublished data from our lab), yet changes in RcsF surface exposure have not been quantified until now. Therefore, I used the SPQ method to investigate its adaptability to cells in stationary phase, and to examine lipoprotein surface exposure in a different growth phase.

I performed the SPQ assay for cells grown until they reached stationary phase (OD_{578} = 2 to 3), for both WT and *bamA101* strains (two biological replicates each). Using the same analysis described in Chapter 3, SE values were compared for IM integral proteins, OMPs and OMLPs. For both WT and *bamA101*, OMLPs and OMPs showed a largely overlapping SE value distribution in stationary phase (Figure 4.6A). Thus, in stationary phase, OMLPs seem to be more surface exposed than in exponential phase (Figure 4.6A). In addition, when compared to WT, both OMPs and OMLPs were still less surface exposed in the *bamA101* strain, the same trend that I observed in exponential phase (Figure 4.3 and Figure 4.6), but to a smaller degree. One explanation is that in stationary phase, there may be less folding and insertion of OMPs and OMLPs. Thus, unassembled OMPs and surface-directed OMLPs may have been degraded in the periplasm, resulting in higher SE values even in *bamA101* strain. Overall, I identified that lipoproteins are surface exposed to a higher degree in stationary phase, both in WT and in *bamA101*.

To investigate why and which lipoproteins are more surface exposed in stationary phase, surface exposure levels were compared together with the protein abundance change between the two growth phases. The SE values of OMLPs in exponential phase and stationary phase are compared, while depicting the lysate level change in a color gradient, as a measure of abundance change (Figure 4.6B). Within the lipoproteins colored in red, i.e. higher protein abundance in stationary phase sample, I identified some lipoproteins that are known to be

induced in stationary phase, such as Slp, also known as "starvation lipoprotein" (Keseler et al., 2017), and OsmE which is activated under the σ^S regulon in stationary phase (Shimada et al., 2004). Some lipoproteins that are more abundant in stationary phase, were more surface exposed, such as YfiL and OsmE (Figure 4.6B). However, many lipoproteins not showing increased abundance in stationary phase (blue) still showed higher SE values in stationary phase compared to exponential phase (Figure 4.6B). Thus, increased lipoprotein surface exposure is not necessarily dependent on increased expression levels of those lipoproteins in stationary phase. Therefore, this suggests there is another layer of regulation of lipoprotein surface exposure which is independent of lipoprotein abundance.

Overall, the SPQ method developed and discussed in Chapter 3 was successfully applied to other mutants and growth conditions to examine the lipoprotein surface exposure. From the results described in the current chapter, a strong BAM dependency of lipoprotein surface exposure has been revealed. Furthermore, cells in stationary phase were shown to have more surface exposed lipoproteins. Thus, I conclude that the surface topology of lipoproteins is not a static phenomenon and that their degree of surface exposure can vary in different conditions.

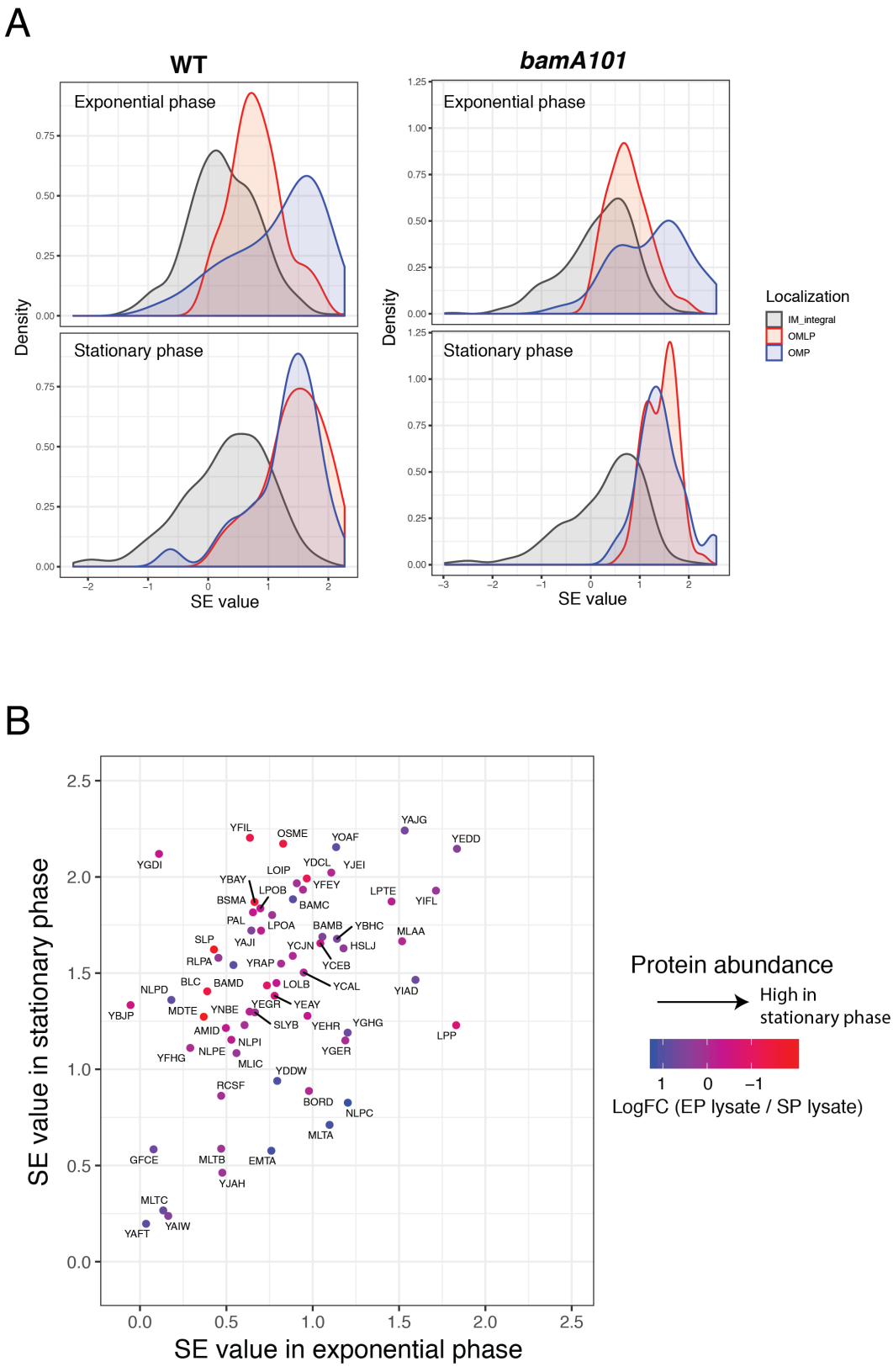


Figure 4.6: SE value comparison of exponential and stationary phase cells, for WT and *bamA101*. A) Distribution of SE values, of proteins in three protein localization categories (IM integral, OMLP and OMP), of two strains (left: WT, right: *bamA101*), of the two growth phases (top: exponential phase, low: stationary phase). B) SE value comparison of lipoproteins between exponential phase (x-axis) and stationary phase (y-axis) in WT, the dot color represents the log fold-change (logFC) of lysate biotinylation measurements as a measure of protein abundance change, red: more abundant in exponential phase (EP), blur: more abundant in stationary phase (SP).

Chapter 5: Systematically mapping protein localization across the bacterial cell envelope

5.1 Background

Cell fractionation is a traditionally used method to elucidate the cellular localization of proteins and other cellular components. The two membranes of Gram-negative bacteria possess different buoyant densities, thus enabling their separation by sucrose density gradients. The sucrose gradient is the most widely used method to separate OM and IM vesicles, in order to examine the localization of a given membrane protein.

By combining the sucrose gradient method with quantitative mass spectroscopy (MS) using TMT-labelling, I generated the first quantitative proteomic dataset of *E. coli* across different sucrose gradient fractions. Using this data, I investigated membrane protein localization in a systematic manner to further advance our knowledge on protein localization, and to compare the protein localization with the SPQ assay from Chapter 3.

5.2 Sucrose gradient fractionation separated IM and OM proteins

WT cells grown in LB media were harvested in exponential phase, lysed and total membrane vesicles (IM and OMs) were isolated as described in the Materials and Methods section. Then, the membrane vesicles were separated by ultracentrifugation across the sucrose gradient, spanning a sucrose concentration from 30% to 60%. Fractionated samples were first analyzed by SDS-PAGE. Coomassie staining showed a gradient separation of the membrane proteome across the different fractions (Figure 5.1). Western blot results of control proteins, BamA for the OM and SecG for the IM, confirmed that the two membranes were successfully separated, with BamA being found at a higher sucrose concentration, and SecG at a lower sucrose concentration, as expected (Figure 5.1). 10 times diluted membrane fraction before sucrose gradient fractionation (total membrane sample) and fractions 2 to 11 (in the increasing order of sucrose percentage) were TMT-labelled and measured by 11-plex TMT-labelling quantitative MS.

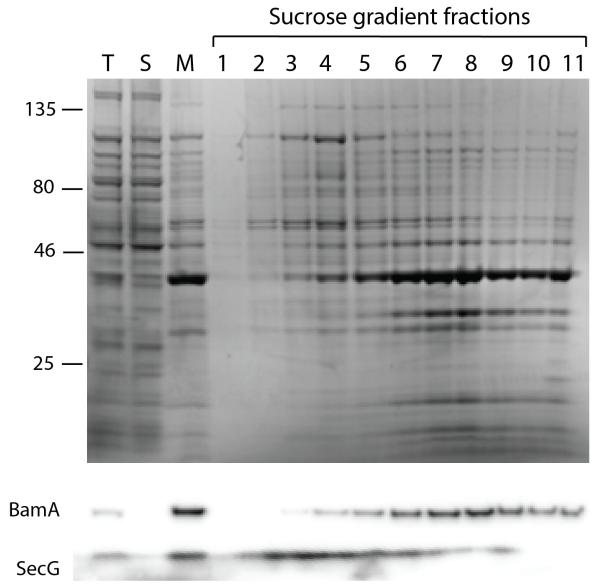


Figure 5.1: Result of sucrose gradient fractionation analyzed on SDS-PAGE gels. Top gel: Coomassie staining of membrane fractionation (T – total cell lysate, S – soluble fraction, M – membrane fraction prior to sucrose gradient) and sucrose gradient fractions from fraction 1 to 11. M (membrane fraction) was diluted 10 times. Lower panel shows Western blots of the corresponding samples as in the above gel, against BamA and SecG. All SDS-PAGE analysis was done using 4-20% gradient gels, for both biological replicates of membrane fractionation, which were later both labelled and measured by MS. Data is representative from two independent biological experiments.

5.3 TMT-labelling MS quantified sucrose gradient fractions

I quantified and analyzed the two biological replicates of sucrose gradient with the assistance of the Proteomics Core Facility at EMBL Heidelberg (Mandy Rettel helped in sample analysis and Frank Stein in data analysis). Overall, 1605 proteins were commonly identified in the two biological replicates with at least two unique peptides and were thus analyzed further. Details of MS, protein identification and data normalization are described in the Materials and Methods section. Figure 5.2A shows the signal distribution of the samples before and after the data normalization, and fractions are annotated and referred to as f02 (fraction 2) to f11 (fraction 11), where the increasing numbers corresponds to increasing sucrose concentration. After data normalization (as described in Materials and Methods), I calculated the log fold change (logFC) of the ratio of signal in each sucrose fraction relative to the total membrane, logFC(each fraction/total membrane). The replicate correlation of these ratio values, between two experiments, was $R = 0.77$, with a p-value of < 2.2e-16 (Pearson correlation) (Figure 5.2B). These analyses confirmed high experimental reproducibility, even after the normalization to the total membrane sample. The results of the two replicates were merged, and one logFC value for each protein for each fraction was assigned using Limma statistical analysis method (Ritchie

et al., 2015). To further analyze and discuss the results, I will from here on use this ratio value of each fraction to the total membrane sample.

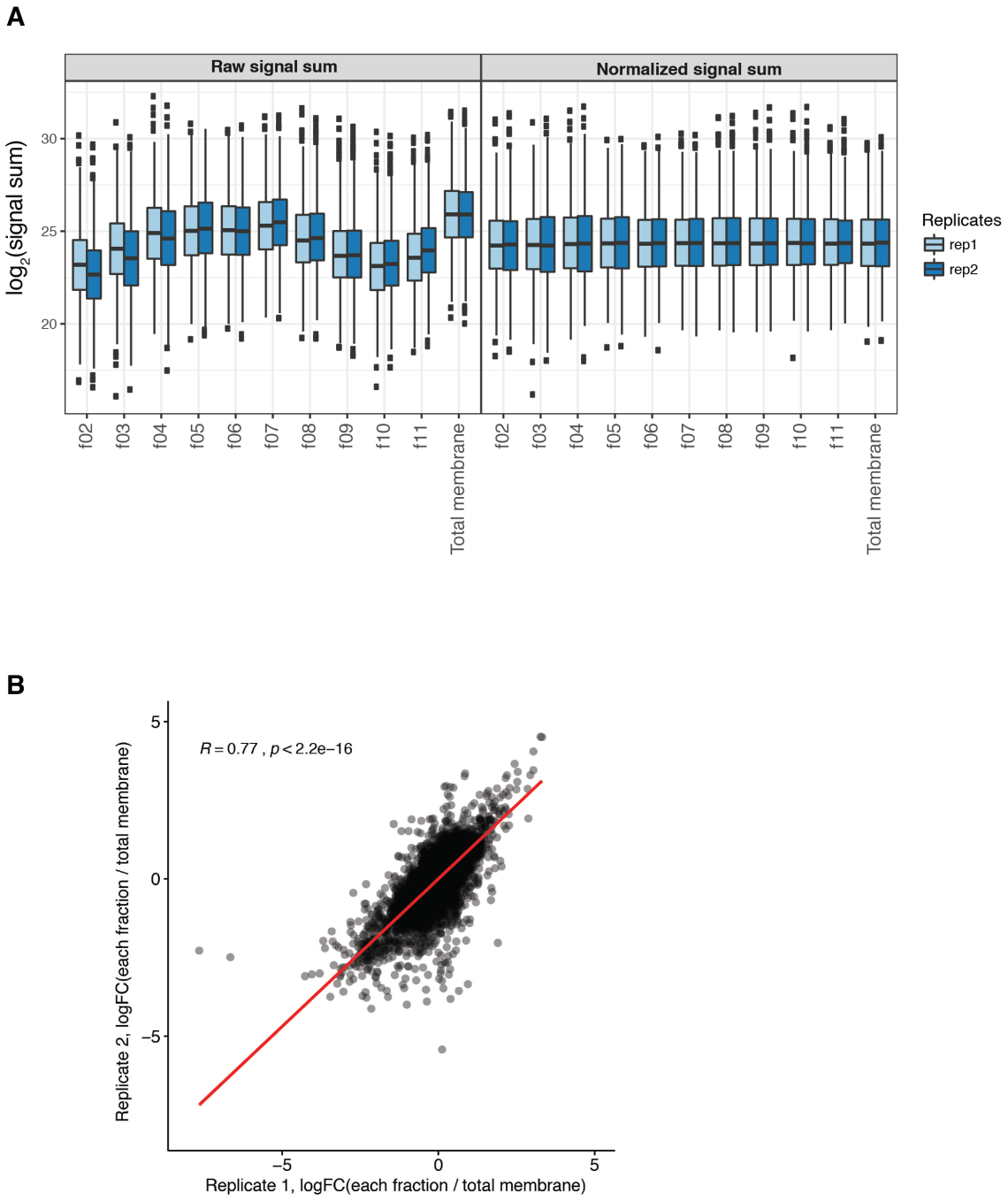


Figure 5.2: Summary of MS measurement, normalization and ratio calculation. A) MS measurement and normalization, left: row signal sum (sum of signals of reporter ions) of each of the sucrose gradient fractions, right: after normalization of signals in the left panel; shown are two biological replicates, from two independent experiments (rep1: light blue and rep2: dark blue). B) Replicate correlation of log fold change (logFC) of the ratio of each of the 10 fractions (f02 to f11 in panel A) / total membrane for each protein identified (1605 proteins identified, thus 16050 data points in total). Values from replicate 1 on x-axis and replicate 2 on y-axis. Pearson correlation of R = 0.77 with p-value < 2.2e-16 was obtained.

5.4 K-means clustering identified OM and IM clusters

Next, I performed k-means clustering using the ratio to the total membrane, setting four as the number of cluster groups, and four was chosen after observing the data distribution in PCA plot (data not shown). The four clusters that resulted contained the following number of proteins: cluster 1: 111, cluster 2: 650, cluster 3: 680, and cluster 4: 164, and the pattern of each cluster is shown in Figure 5.3. Although there were some outliers, most of the proteins in each cluster follow a similar pattern as the median (Figure 5.3A). Lower sucrose density fractions should contain more IM proteins, and higher sucrose density fractions should contain more OM proteins, as seen in Figure 5.1. From Figure 5.3A and B, cluster 1 was hypothesized to contain mainly OM proteins, with higher logFC values in higher fractions (f08 to f11), whereas cluster 2 was hypothesized to be enriched in IM proteins, with higher logFC in lower fractions (f03 to f05). Clusters 3 and 4 displayed patterns that were not matching with either OM or IM proteins and in fact contained many soluble proteins which will be discussed below.

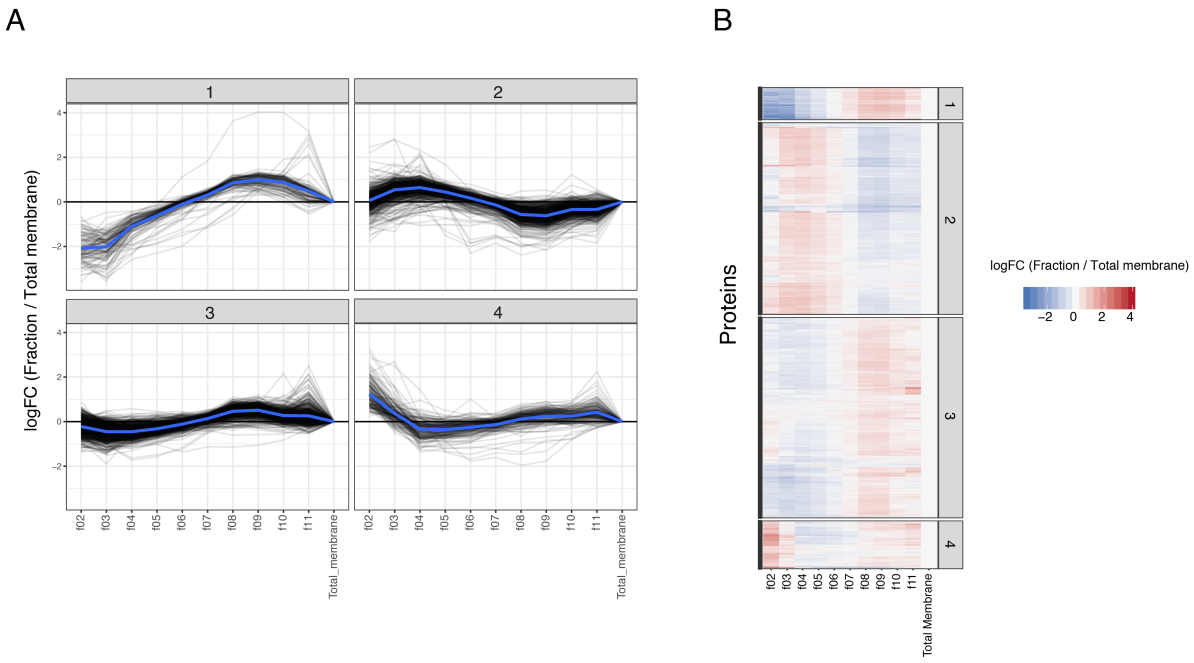


Figure 5.3: K-means clustering of the log fold-change ratio pattern across the fractions. A) log fold-change ratio value (log fold-change ratio of each fraction relative to the total membrane (logFC (Fraction / Total membrane), y-axis) for each k-means cluster (1 to 4). The pattern of each protein across fractions in an increasing order (f02 to f11) and the total membrane fraction (logFC = 0) is shown for each cluster. The median for each cluster is drawn in blue. B) Heatmap of the cluster in A) and the x-axis is same as in A), in an increasing order of fractions (f02 to f11), and Total membrane (logFC = 0). The color scale indicates the logFC in intensifying shades of red and blue that reflect more or less enrichment respectively. The number of proteins identified, and thus shown in both A) and B), are: cluster 1: 111, cluster 2: 650, cluster 3: 680, and cluster 4: 164.

In order to validate clusters in an unbiased manner, I compared the clustering results with known localization annotation from STEPdb database (Orfanoudaki and Economou, 2014), in a similar manner as for the SPQ method in Chapter 3. Figure 5.4 shows the relative assignment of the clustering categories for each localization annotation. As hypothesized, most OM proteins (i.e. both OMPs and OMLPs) are found within cluster 1 and IM proteins including IM integral, IM periplasm and IMLP proteins are associated with cluster 2 (Figure 5.4). From here on, cluster 1 will be referred to as the OM cluster, and cluster 2 will be referred as the IM cluster. An exception are the IM cytoplasmic (IM-cyto) proteins, which are categorized in both cluster 2 and 3. This may be due to the localization annotation which includes both the proteins in the IM and the ones weakly associated to the IM or IM proteins, and also proteins co-localized in both localizations (Figure 5.4A). Overall, I was able to systematically characterize the *E. coli* membrane proteome using quantitative MS, and to assign clustering patterns to the proteome in an unbiased manner, which was is in a good agreement with the membrane distribution of known OM and IM proteins.

Besides membrane proteins clustering into distinct OM or IM clusters, there were two other clusters, cluster 3 and 4, which were dominated by soluble cytoplasmic, both cytoplasmic and IM-cyto, or periplasmic proteins (Figure 5.4A). During the sample preparation, soluble proteins were mostly removed by ultracentrifugation, thus the soluble proteins identified are considered as either contaminants, or proteins interacting with the membrane itself or with membrane proteins, and were thus identified in the membrane fractions. Most of the soluble proteins were not clustered into OM or IM clusters, suggesting they are not localized specifically to either membrane and could also be non-specifically encapsulated within the lumen of membrane vesicles during the cell lysis process. Thus clusters 3 and 4 are considered to contain predominantly soluble proteins, or non-fractionated clusters, and will be discussed later.

As I observed that the abundance of a protein in MS measurements (top3 value) affects the accuracy of protein localization determination using the SPQ method in Chapter 3, I also checked if it is the case for this sucrose gradient experiment. The top3 value (average signal of three peptides with the highest detected signal levels) reflects the abundance of a given protein within a MS sample. All the proteins identified were divided into four “protein abundance categories” using the top3 value (Figure 5.4B). For each abundance and annotated localization category, k-means clustering results are visualized (Figure 5.4B). I found that the higher the protein abundance, the more accurate the protein clustering, especially for the two membrane

clusters (cluster 1 and 2) (Figure 5.4B). All IM proteins (IM-integral, IM-peri and IMLP) in the highest abundance category are found in cluster 2, and all but one of the highest abundant OM proteins (OMP and OMLP) are found in cluster 1 (Figure 5.4B). This one exception was Blc, annotated to be an OMLP. A study on Blc refer it as OMLP, and Blc has been shown to bind fatty acids and phospholipids (Campanacci et al., 2006). Blc does not have a Lol avoidance signal, and from these information, Blc is thought to be localized in the OM. It however, does not behave like a typical OM protein upon sucrose gradient fractionation, hence possibly a fraction of Blc is kept in the IM or periplasm. The IM-cyto annotated proteins are again not clustered to a specific cluster even when proteins are abundant (Figure 5.4B). Overall, similar to the SPQ method, determination of localization is in general affected by the abundance of the protein in the MS samples, and I can use the abundance information to assess confidence of the data.

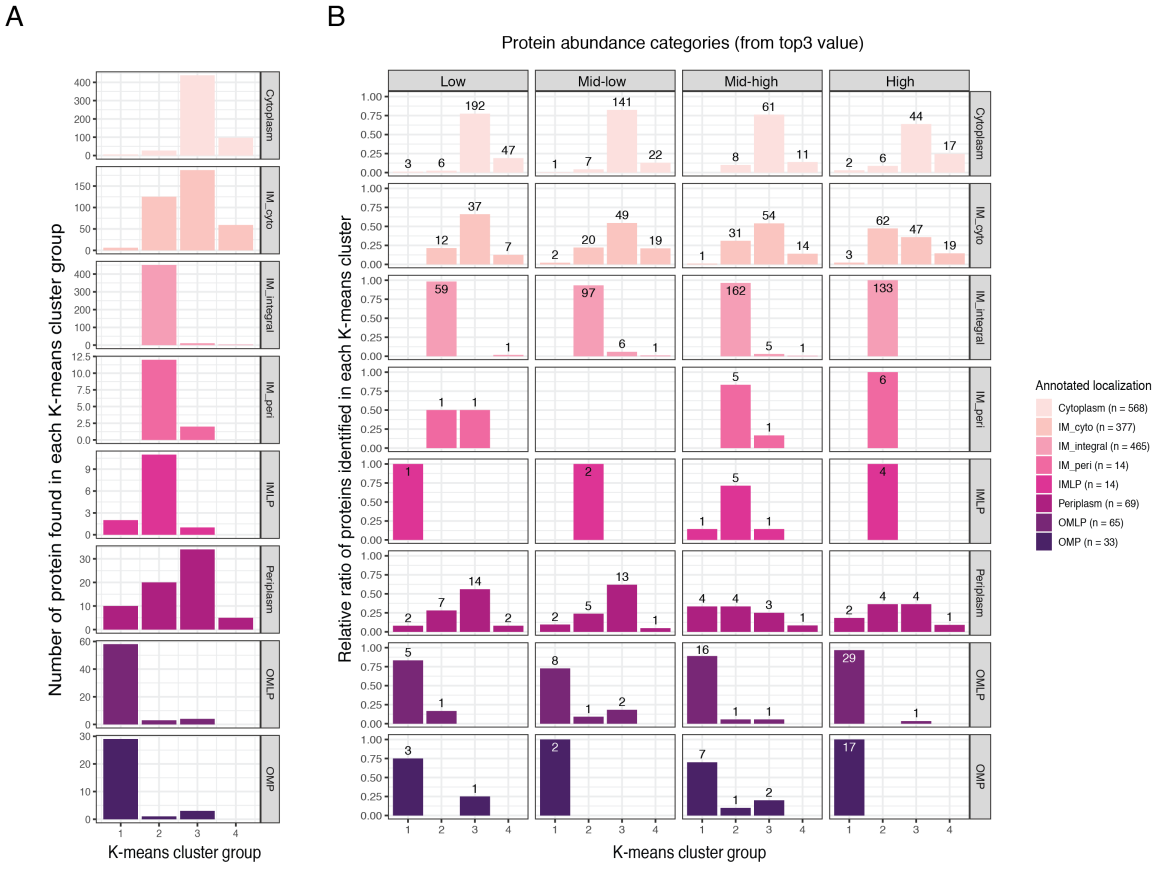


Figure 5.4: Distribution of proteins in each K-means cluster group for each protein localization category. A) The four clusters are displayed on the x-axis and the number of proteins identified in each cluster on the y-axis. Localization categories are ordered (top to bottom) from inside to the surface of *E. coli* cells (cytoplasm, IM-cyto, IM-integral, IM-peri, IMLP, Periplasm, OMLP and OMP), with the number of proteins in each category shown in the legend (same legend as panel B). B) Relative ratio of proteins identified in each k-means cluster group for each annotated localization category (same as panel A), for each protein abundance category. The number at each bar indicates the number of identified proteins. The protein abundance categories were split into 4 categories using top3 value. Top3 value range for each of the categories are: Low – 5.71 to 6.72, Mid-low – 6.72 to 7.22, Mid-high – 7.22 to 7.82, High – 7.82 to 9.56.

5.5 Proteins clustered differently from their annotated localization were identified

5.5.1 Non-OM annotated proteins displaying OM sucrose gradient pattern

Within the OM cluster (k-means cluster 1), there were a number of proteins not annotated as OM proteins (all summarized in Table 9). For example, two lipoproteins annotated to be IM lipoproteins, YidQ and YmbA, were found in the OM cluster (Table 9). YidQ is a protein of unknown function (The UniProt Consortium, 2019). Whereas, YmbA (also known as PqiC) was shown to be involved in the maintenance of the OM, together with other Pqi proteins, and predicted to be in the OM but interacting with its partner IM protein, PqiB (Nakayama and Zhang-Akiyama, 2017). Within the STEPdb database, there is contradicting information regarding YidQ and YmbA, where their IM localization is predicated although they do not appear to encode a classical Lol avoidance signal, which suggests they may be targeted to the OM (Orfanoudaki and Economou, 2014). Thus, as these two IM annotated lipoproteins do not have clear evidence for their IM localization, and from my assay these proteins behave as OM proteins, I conclude that the localization of these proteins should be revised. Therefore, the systematic measurement of sucrose gradient fractionation can be used to find potentially misannotated proteins, and suggest their true localization.

In addition to IM annotated proteins, some soluble proteins were also identified in the OM cluster, including 10 periplasmically-annotated proteins. NanC, although the localization code is for periplasm, some predictions state its OM localization as an OMP within the STEPdb database (Orfanoudaki and Economou, 2014). BcsC is involved in cellulose synthesis, which is abolished in *E. coli* K-12 strains, but in *Salmonella*, BcsC is considered to be localized in the OM (Whitney and Howell, 2013). Besides these proteins, which may be misannotated in the STEPdb database and are actually integrated in the OM, LptA, a well-studied periplasmic protein, was also found in this cluster (Table 9). LptA is a periplasmic component of the Lpt machinery that transports LPS from the IM to the OM, and has been reported to bridge the OM and IM components of the machinery (Sherman et al., 2018). I observed that upon sucrose gradient fractionation, this protein behaves like an OM protein, suggesting a stronger interaction with the OM component, LptD, within the complex. Other components of Lpt machineries were found in clusters of their known localization, LptDE in the OM cluster, LptBCFG in the IM. Therefore, proteome-wide quantitative analysis of membrane fractionation

by sucrose density may be able to demonstrate affinities of periplasmic proteins towards the OM.

Some cytoplasmic proteins were also identified in this cluster, 6 of them annotated to be in IM-cyto, and another 6 only in the cytoplasm. An IM-cyto annotated protein Dps is known to bind DNA, but additionally, its function as ferritin and other roles are also discussed (Calhoun and Kwon, 2011). Despite its known function in the cytoplasm, Dps has been previously reported to be also localized in the OM (Lacqua et al., 2006). This suggests the possibility of other cytoplasmic protein being identified in the OM cluster, due to them having a dual localization. For example, YgaU which is annotated to be IM-cyto, and was previously shown to bind to potassium, and acts as "cytoplasmic potassium sensor" (Ashraf et al., 2016). On the other hand, it was shown to be involved in cell envelope maintenance through regulating peptidoglycan cross-linking, raising the possibility of its localization and functionality in the periplasm (Bernal-Cabas et al., 2015). Overall, although some non-OM proteins may be clustered in the OM cluster by experimental error, for many, it appeared the current annotation may be incorrect, or that consideration of additional localizations is required.

5.5.2 Non-IM annotated proteins displaying IM sucrose gradient pattern

In addition, I identified non-IM annotated proteins within the IM cluster (k-means cluster 2) (Figure 5.4A, Table 9). For example, four OM proteins were identified, three OMLPs and one OMP which clustered together with IM proteins. For one lipoprotein, MdtE, there is conflicting information in the STEPdb database (Orfanoudaki and Economou, 2014) where even though MdtE is annotated as an OMLP, it carries a Lol avoidance signal (lipidated cysteine followed by an aspartate residue) in its N-terminus. My experimental evidence demonstrated MdtE to localize to the IM, which is consistent with the presence of the Lol avoidance signal and conflicts with the OMLP annotation. YhdP is annotated to be an OMP of unknown function and was recently shown to be involved in OM barrier permeability and predicted to localize to the IM, albeit without experimental evidence (Mitchell et al., 2018). Furthermore, within the STEPdb database (Orfanoudaki and Economou, 2014), where predictions and evidence from different sources are compiled, there are some predictions suggesting its IM localization, which is in line with our experimental findings. Thus, MdtE and YhdP are potentially misannotated with respect to their localization and need to be re-annotated as IM proteins.

Some proteins annotated as periplasmic in STEPdb database (Orfanoudaki and Economou, 2014), which are known to carry IM transmembrane domains or being a part of IM protein complex were identified: DegS and RseB involved in the σ^E stress response cascade (Chaba et al., 2011), EnvC is, as a component of the divisome, interacting with and being activated by the IM complex FtsEX, and controlling septal cleavage during division (Yang et al., 2011), and MlaD is an IM component of the Mla transport machinery, confirmed by previous study showing its IM localization by sucrose gradient (Malinverni and Silhavy, 2009). In addition to periplasmic proteins, multiple Rfa proteins, which are annotated to be cytoplasmic proteins known to be involved in LPS synthesis pathways, were also found in this cluster, suggesting their direct association to the IM or to IM proteins (Table 9). These observations revealed the complexities of assigning localization, as many proteins may be annotated to be soluble in the STEPdb database (Orfanoudaki and Economou, 2014), presumably as they function in the soluble environment. However, they may anchor in or interact with membranes (and membrane proteins) and thus are also considered as membrane proteins, as shown from quantitative analysis of the sucrose density gradient.

5.5.3 Non-membrane clustered membrane proteins

While I found most membrane-annotated proteins within OM (cluster 1) or IM (cluster 2) clusters, some membrane-annotated proteins were associated with cluster 3 or 4 that typically contained soluble proteins (Figure 5.4 and Table 9). Of those proteins, two well-studied proteins which are known to form stable functional complexes with OM components were identified: TonB and TamB. TonB is the crucial protein for TonB-dependent transport of compounds through TonB dependent transporters (TBDTs) in the OM, and contains a long periplasmic domain which spans through the periplasm and interacts with OMPs at the OM (Braun et al., 2006). It has been previously shown that TonB is found in both IM and OM fractions of a sucrose gradient experiment (Higgs et al., 2002; Letain and Postle, 1997). Similarly, TamB has its partner OM protein TamA, an OMP, forming a stable protein complex that spans the periplasmic space and functions in autotransporter biogenesis (Selkrig et al., 2012). Thus, these results suggest that some proteins that span the periplasm, or have stable interactions with both membranes, exhibit an atypical fractionation pattern (similar to non-membrane proteins) and thus have been clustered to these two clusters which is where the majority of soluble proteins are found.

5.6 Similarities and differences between the sucrose gradient fractionation and the SPQ method

Finally, I examined how the surface proteome quantification (SPQ) method from Chapter 3 compares to the classical sucrose gradient fractionation. In terms of experimental conditions, cells were harvested from the same growth phase were analyzed by SPQ and sucrose gradient fractionation, the cell lysate and membrane fractions were prepared in the same way, and the strain used was also the same. Thus, the results of the two separate datasets should have minimal biological differences, and therefore any differences should be inherent to the methodologies. Traditionally, sucrose gradient fractionation experiments are analyzed through Western blots as shown in Figure 5.1. As membrane proteins, both IM and OM proteins, are found highly abundant across several fractions, I therefore compared the average values of logFC (each fraction/total membrane) of pooled high sucrose fractions (Fraction 8, 9 and 10) and pooled low sucrose fractions (Fraction 2, 3 and 4). Then the difference of the average logFC (high fraction logFC – low fraction logFC), was calculated to assign a single value for each protein reflecting the sucrose gradient result, and the value is referred to as the "sucrose gradient ratio" (Figure 5.5A). Figure 5.5 shows that this sucrose gradient ratio captured the distinct localization of proteins in a more refined manner than k-means clustering. From this analysis, OM proteins displayed high sucrose gradient ratios, soluble proteins being in the middle range, and IM having low values for IM proteins. I recapitulated the unusual behavior of IM-cyto proteins, which show a bimodal distribution in this quantification (Figure 5.5A), matching with the k-means clustering result that this protein group contain both IM proteins and soluble proteins.

As I found the sucrose gradient ratio reflects the membrane localization of proteins, I used this value to predict membrane localization. From the comparison of IM-integral proteins to OM proteins (OMPs and OMLPs), the ratio = 1 was selected as the cutoff to separate OM and IM proteins (Figure 5.5B). As expected from k-means clustering, all non-OM annotated proteins in cluster 1 (OM cluster) have the sucrose gradient ratio above 1 (Table 9). Similarly, non-IM annotate proteins in cluster 2 (IM cluster) all have the sucrose gradient ratio below 1 (Table 9). This further supports the identification of a protein in an unexpected cluster based on its annotated localization was not an artifact of the clustering method, as it was entirely congruent with the calculated sucrose gradient ratio.

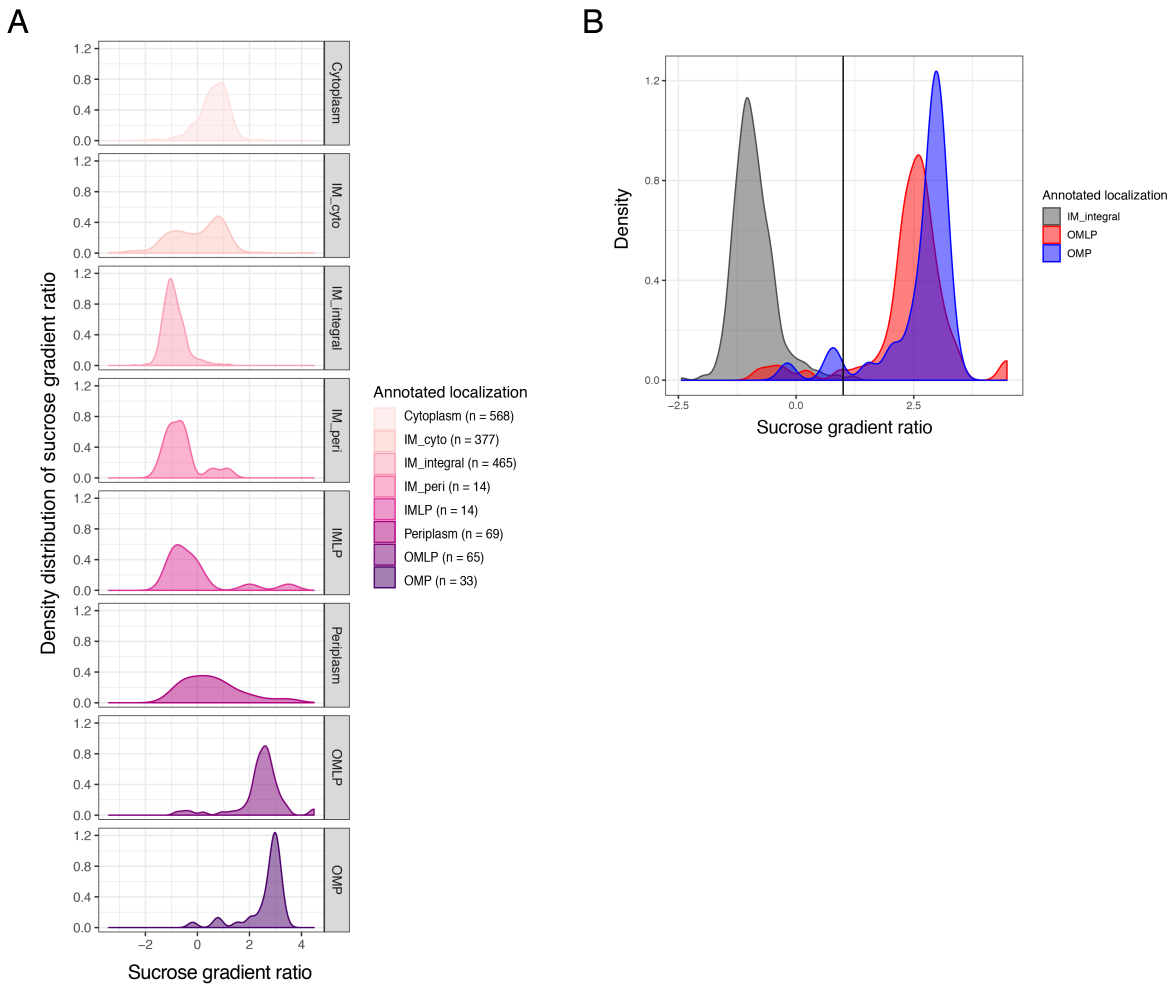


Figure 5.5: Density plots of sucrose gradient ratio for each annotated protein localization. Sucrose gradient ratio was calculated as the difference of average logFC of high sucrose fractions (fraction 8, 9, 10) and low sucrose fractions (2, 3, and 4). Density plots are plotted separately for each annotated protein localization groups. A) for all 8 annotated localizations separately, B) for IM-integral, OMLP and OMP only, with the vertical line for sucrose gradient ratio = 1. Note, higher or lower sucrose gradient ratios reflects OM or IM protein localization, respectively.

After obtaining a single value for each protein that reflects its localization based on the sucrose density gradient fractionation, I then directly compared this with the SE values acquired using the SPQ method described in Chapter 3 (Figure 5.6). In general, IM proteins displayed lower SE values and low sucrose gradient ratio, and they are thus found in bottom-left side of Figure 5.6. On the other hand, OM proteins, both OMPs and OMLPs are found on the top-right corner of Figure 5.6 due to a high sucrose gradient ratio and high SE value. Therefore, both SE values from the SPQ method and sucrose gradient ratios were able to differentiate protein localizations between inner and outer membranes of *E. coli*. However, within the proteins localized in each membrane, there was no clear correlation between SE values and sucrose ratio values (Figure 5.6), showing that the two methods provide partially orthogonal information.

I then checked the clustering and sucrose gradient ratio of the non-OM proteins showing high surface exposure values (SE values) in the SPQ method. MetQ and DcrB, the two IM-annotated lipoproteins, validated to be surface exposed by dot blot in Chapter 3, were clustered to the IM k-means cluster and displayed low sucrose gradient ratios (MetQ: -0.934, DcrB: -0.939) (Figure 5.6). Thus, these two proteins were concluded to be in the IM based on the sucrose gradient but surface exposed from the SPQ method. A possible explanation for this discrepancy is that these proteins may have some surface exposed domains albeit anchored to the IM. Alternatively, a very small fraction of these proteins may be transported to the OM and surface exposed, while when the sucrose-gradient experiment is performed they behave as IM proteins as the majority of the protein is localized in the IM. However, further investigation is required to conclude the topology of these proteins.

Other IM proteins that showed high SE values were TonB, ElyC and YgaU (Chapter 3). As discussed above, TonB was not clustered with other IM proteins, with the sucrose gradient ratio value of 1.245 being above the threshold of 1 (Figure 5.6). It also showed high SE values, consistent with its unique confirmation of spanning the cell envelope (Braun et al., 2006) (Figure 5.6). Whereas YgaU, which is annotated to be IM-cyto also showed high SE values and behaved as an OM protein by sucrose gradient (sucrose gradient ratio = 1.844). Therefore, from these two assays, YgaU was concluded to show properties of an OM protein. On the other hand, even though ElyC showed high SE values by the SPQ, in the sucrose gradient it was clustered with other IM proteins (sucrose gradient ratio = -0.204). Thus, similarly to MetQ and DcrB discussed above, ElyC is most likely anchored to the IM while displaying a portion of the protein at the cell surface.

Overall, I showed that the quantitative analysis of sucrose gradient using TMT-labelling MS can systematically separate and identify the localization of membrane proteins. Additionally, the result suggested non-conventional or mis-annotated localization for some proteins, including soluble proteins possibly associated with membrane proteins. I compared the sucrose gradient results with the SPQ method and confirmed their common ability to identify membrane protein localization, while emphasizing differences in the information provided by the two methods. Whilst quantitative measurements of protein enrichment across the sucrose density gradient can robustly separate known IM and OM proteins, the SPQ method provides distinct information relating to protein topology within the cell envelope. Thus, the combination of these two methods reveals an additional dimension of membrane protein topology in *E. coli*.

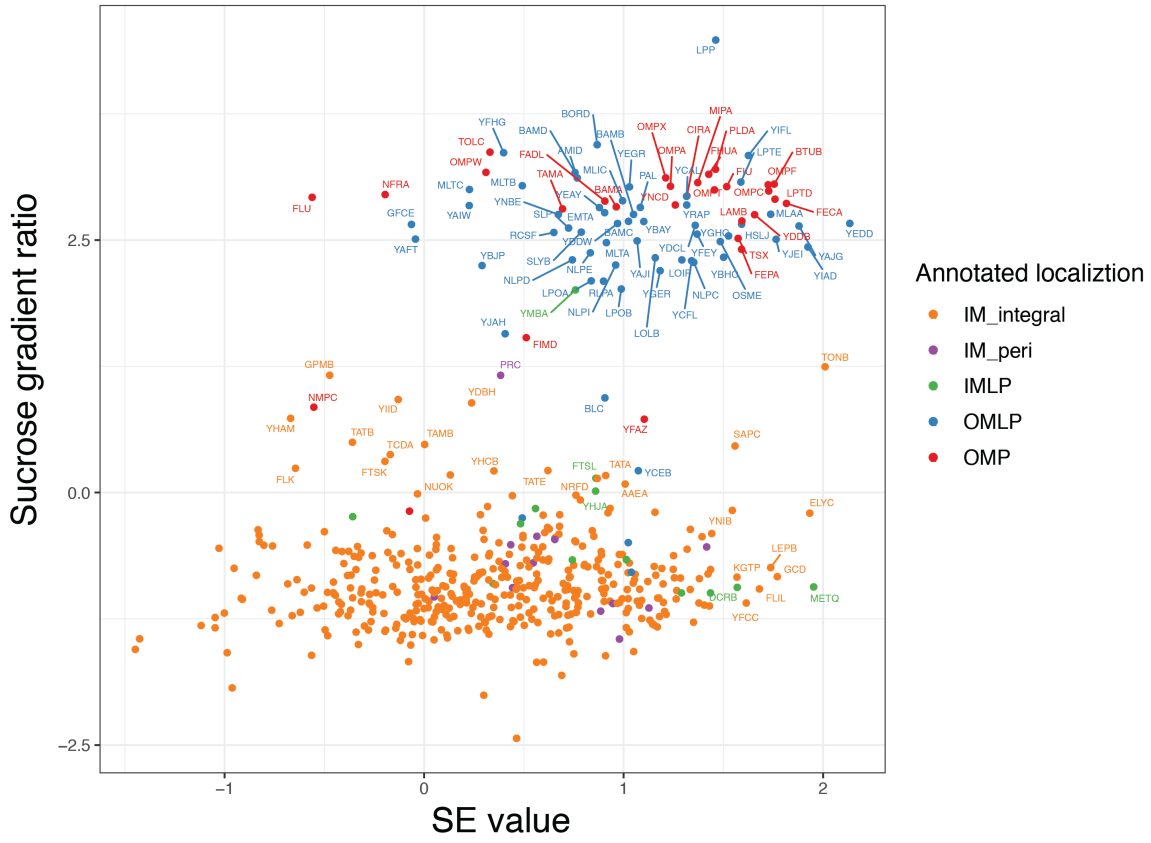


Figure 5.6: SE value compared to sucrose gradient ratio. SE value (surface exposure value, x-axis) from Chapter 3 compared with sucrose gradient ratio (y-axis). Only membrane proteins in the five localization categories (IM integral, IM peri, IMLP, OMLP and OMP) are shown and colored as indicated in the figure. Proteins with either sucrose gradient ratio > 0 or SE value > 1.5 are annotated for their protein names.

Table 9: Proteins annotated to be in different localization from k-means cluster groups

For each protein the following information is included, (1) K means annotation (1 – OM, 2 – IM, 3 – soluble, 4 – soluble (4)), (2) abundance category calculated from top3 value as shown in Figure 5.4B, (3) sucrose gradient ratio, either above 1 or below 1, (4) SE value from Chapter 3, NA for proteins not quantified in the SPQ method in Chapter 3, (5) annotated localization from STEPdb databse (Orfanoudaki and Economou, 2014).

Protein Name	K means cluster	Abundance category	Sucrose gradient ratio	SE value	STEPdb annotation
BCSC	OM	Mid-low	Above 1	0.045	Periplasm
LPTA	OM	Mid-high	Above 1	0.486	Periplasm
MALM	OM	Mid-high	Above 1	0.125	Periplasm
NANC	OM	High	Above 1	0.719	Periplasm
PLIG	OM	Mid-low	Above 1	NA	Periplasm
YBCH	OM	Mid-high	Above 1	-0.094	Periplasm
YCHO	OM	Low	Above 1	NA	Periplasm
YDIY	OM	High	Above 1	0.668	Periplasm
YNFB	OM	Mid-high	Above 1	0.745	Periplasm
YNHG	OM	Low	Above 1	NA	Periplasm
YIDQ	OM	Low	Above 1	NA	IMLP
YMBA	OM	Mid-high	Above 1	0.759	IMLP
DPS	OM	High	Above 1	-0.321	IM-cyto
GLGA	OM	Mid-high	Above 1	-0.304	IM-cyto
GROS	OM	Mid-low	Above 1	NA	IM-cyto
SEQA	OM	High	Above 1	NA	IM-cyto
TRXA	OM	Mid-low	Above 1	NA	IM-cyto
YGAU	OM	High	Above 1	1.855	IM-cyto
ACEB	OM	Low	Above 1	-0.347	Cytoplasm
CSPC	OM	Low	Above 1	NA	Cytoplasm
PANB	OM	Low	Above 1	-0.546	Cytoplasm
RPMC	OM	High	Above 1	NA	Cytoplasm
RPME	OM	Mid-low	Above 1	NA	Cytoplasm
RPMG	OM	High	Above 1	NA	Cytoplasm
BCSE	IM	Mid-low	Below 1	-0.275	Cytoplasm
BFR	IM	Low	Below 1	-0.719	Cytoplasm
CLSC	IM	Mid-low	Below 1	-0.268	Cytoplasm
CYSG	IM	Mid-low	Below 1	NA	Cytoplasm

FEOA	IM	Mid-low	Below 1	NA	Cytoplasm
INAA	IM	Mid-high	Below 1	-0.756	Cytoplasm
ISCR	IM	Low	Below 1	NA	Cytoplasm
LIPA	IM	High	Below 1	-0.287	Cytoplasm
PURL	IM	Mid-low	Below 1	NA	Cytoplasm
RATB	IM	Mid-high	Below 1	NA	Cytoplasm
RFAI	IM	Mid-high	Below 1	-0.456	Cytoplasm
RFAJ	IM	High	Below 1	-0.440	Cytoplasm
RFAS	IM	Mid-high	Below 1	-0.402	Cytoplasm
RFAY	IM	High	Below 1	-0.801	Cytoplasm
RFAZ	IM	Mid-high	Below 1	-0.464	Cytoplasm
RPSB	IM	High	Below 1	-0.617	Cytoplasm
RPST	IM	High	Below 1	-0.689	Cytoplasm
YBGI	IM	Low	Below 1	NA	Cytoplasm
YBHP	IM	Low	Below 1	NA	Cytoplasm
YBJQ	IM	Mid-high	Below 1	NA	Cytoplasm
YFCH	IM	High	Below 1	0.617	Cytoplasm
YGCN	IM	Low	Below 1	-0.256	Cytoplasm
YGIC	IM	Mid-low	Below 1	-0.105	Cytoplasm
YHET	IM	Mid-high	Below 1	NA	Cytoplasm
YKGG	IM	Low	Below 1	-0.618	Cytoplasm
YRBL	IM	Mid-high	Below 1	NA	Cytoplasm
ZAPE	IM	Mid-low	Below 1	-0.385	Cytoplasm
MDTE	IM	Mid-high	Below 1	0.492	OMLP
YCJN	IM	Mid-low	Below 1	1.039	OMLP
YEHR	IM	Low	Below 1	1.023	OMLP
YHDP	IM	Mid-high	Below 1	-0.073	OMP
BGLX	IM	Low	Below 1	0.303	Periplasm
BTUF	IM	Low	Below 1	NA	Periplasm
CDH	IM	High	Below 1	1.841	Periplasm
CREA	IM	Mid-low	Below 1	NA	Periplasm
DEGS	IM	High	Below 1	0.791	Periplasm
DSBA	IM	Low	Below 1	NA	Periplasm
ENVC	IM	Mid-high	Below 1	0.995	Periplasm
FDOG	IM	High	Below 1	0.063	Periplasm
FHUD	IM	Low	Below 1	1.157	Periplasm
GLPQ	IM	Mid-low	Below 1	0.222	Periplasm
MLAD	IM	High	Below 1	0.976	Periplasm
MLTF	IM	Low	Below 1	0.479	Periplasm
NAPG	IM	Low	Below 1	NA	Periplasm

OPPA	IM	Mid-low	Below 1	0.356	Periplasm
PSTS	IM	Low	Below 1	NA	Periplasm
RSEB	IM	Mid-high	Below 1	0.711	Periplasm
RSXG	IM	Mid-high	Below 1	0.681	Periplasm
SAPA	IM	Mid-high	Below 1	1.381	Periplasm
YIBQ	IM	Mid-low	Below 1	0.358	Periplasm
YIIQ	IM	Mid-low	Below 1	NA	Periplasm
BIRA	Soluble	Mid-low	Below 1	NA	IM-integral
CYOC	Soluble	Mid-low	Below 1	NA	IM-integral
FTSK	Soluble	Mid-high	Below 1	-0.196	IM-integral
GPMB	Soluble	Mid-high	Above 1	-0.474	IM-integral
SAPC	Soluble	Mid-high	Below 1	1.559	IM-integral
TAMB	Soluble	Mid-high	Below 1	0.003	IM-integral
TCDA	Soluble	Mid-low	Below 1	-0.170	IM-integral
TONB	Soluble	Mid-high	Above 1	2.010	IM-integral
YDBH	Soluble	Mid-low	Below 1	0.238	IM-integral
YHAM	Soluble	Mid-low	Below 1	-0.669	IM-integral
YIID	Soluble	Mid-low	Below 1	-0.130	IM-integral
PRC	Soluble	Mid-high	Above 1	0.384	IM-peri
YADE	Soluble	Low	Below 1	NA	IM-peri
YHJA	Soluble	Mid-high	Below 1	0.860	IMLP
BLC	Soluble	High	Below 1	0.906	OMLP
YCEB	Soluble	Mid-high	Below 1	1.074	OMLP
YECR	Soluble	Mid-low	Above 1	NA	OMLP
YJAH	Soluble	Mid-low	Above 1	0.406	OMLP
FIMD	Soluble	Low	Above 1	0.512	OMP
NMPC	Soluble	Mid-high	Below 1	-0.553	OMP
YFAZ	Soluble	Mid-high	Below 1	1.104	OMP
GSIC	Soluble (4)	Mid-low	Below 1	0.464	IM-integral
MALP	Soluble (4)	Mid-high	Below 1	-0.911	IM-integral
RSMB	Soluble (4)	Low	Below 1	NA	IM-integral

Chapter 6: Rcs system activation by mislocalized surface lipoprotein RcsF

Disclaimer:

This chapter presents collaborative work with other members of the Typas lab and the Huang Lab (Stanford University, USA). I performed the Rcs activation (Figure 6.2) and the FtsZ level measurements (Figure 6.5). Matylda Zietek (Typas Lab) performed the cell length determination in batch cultures (Figure 6.3), and Amanda Miguel (Huang Lab) performed the single-cell experiments for Rcs activation (Figure 6.4) and FtsZ ring formation (Figure 6.7). This work is part of a manuscript, which is about to be submitted.

6.1 Background

As mentioned throughout my thesis, RcsF is the model surface exposed lipoprotein that its transport to the surface requires the BAM machinery (Cho et al., 2014; Konovalova et al., 2014). These two studies in 2014 revealed not only the surface localization of RcsF, but also that the transport efficiency of RcsF is important since it acts as the sensor for an envelope stress response system, Rcs (Cho et al., 2014; Konovalova et al., 2014). To our knowledge, this is the only surface lipoprotein that its surface transport plays a crucial role in a signaling system in *E. coli*. Thus we investigated further the functionality of RcsF localization and surface transport.

The Rcs system is an envelope stress response system that regulates the expression of many genes, among them those required for cell envelope maintenance and biofilm production (Filippova et al., 2018; Guo and Sun, 2017; Wall et al., 2018). In 2014, the molecular mechanism by which the system is activated was discovered (Cho et al., 2014; Konovalova et al., 2014). In the absence of envelope stress, the Rcs system remains in an inactive state whereby the OM lipoprotein RcsF is transported to the cell surface being plugged and sequestered into the lumen of a β-barrel OMP by the BAM machinery to form an RcsF-OMP complex in the OM. The sequestration of RcsF to the surface within an OMP ensures that the system is off as RcsF cannot interact with the downstream protein of the cascade, IgaA (Figure 6.1, left) (Cho et al., 2014). Upon cell envelope stress, the journey of RcsF to the OM may be perturbed, e.g. defects in Lol-mediated transport of RcsF to the OM, or BamA defects in assembling RcsF into

OMPs. Then RcsF remains trapped within the periplasmic space where it interacts with IgaA and alleviates the IgaA-mediated repression of the downstream signaling proteins (RcsC/D), thus activating the Rcs phosphorelay (Figure 6.1, right) (Cho et al., 2014). This Rcs system activation can be artificially induced by mutating the RcsF to be retained at the IM; there RcsF interacts with IgaA and promotes activation of the Rcs phosphorelay (Figure 6.1, right) (Cho et al., 2014). We chose to artificially induce Rcs activation in this way in the following experiments to monitor the consequences of activating the cascade in the absence of the activating signal (envelope damage) of the Rcs system.

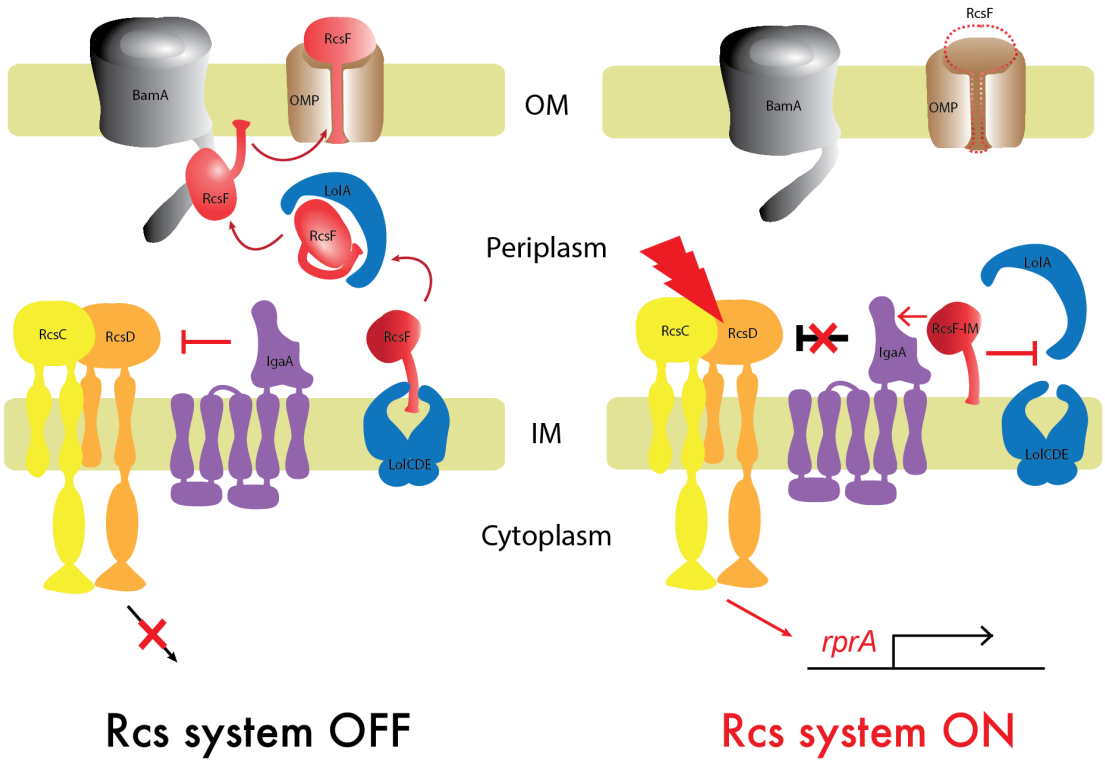


Figure 6.1: Illustration of the Rcs system activation mechanism by RcsF-IM. The Rcs system is turned off in non-inducing conditions (left). RcsF is canonically transported to the OM by the Lol system (LolCDE in the IM and LolA in the periplasm), and inserted inside an OMP by the BamA to reach the cell surface. In these conditions, the phosphorelay cascade by RcsC/D is repressed by the IM protein IgaA. Right panel shows the Rcs system induction by RcsF-IM. RcsF-IM containing a Lol avoidance signal is retained in the IM and has full access to interact with IgaA and relieve its inhibition of the phosphorylation cascade. *rprA* is one of the genes known to be induced upon Rcs activation. Figure adapted from previous work (Cho et al., 2014, and unpublished work by Matylda Zietek).

Beside the evidence of Rcs system involved in regulating a diverse set of genes, how the activation of the system is beneficial for the cells undergoing cell envelope stress is not entirely clear. The Rcs output includes a feed back to fix lipoprotein transport defects, by expressing *lolA* gene encoding for an essential protein in the Lol transport machinery (Tao et al., 2012). Other evidence, such as a chemical genomics screen of *E. coli*, suggests that the Rcs system may be important to maintain cell shape (Nichols et al., 2011). There the deletion of Rcs system genes (*rcsF*, *rcsC*, *rcsD*, *rcsB*) leads to growth defects upon treatments with chemicals that change the cell shape (NaCl, A22 and Mecillinam) or directly affect LPS assembly (Nichols et al., 2011). A22 and Mecillinam are also known to inhibit the activities of proteins crucial for cell elongation, MreB and PBP2 respectively (Bean et al., 2009; Spratt, 1975). Cell shape and cell cycle are also closely related within cellular physiology, and thus we hypothesized that the Rcs system is involved, and possibly taking a major role in bacterial cell shape and cell cycle control upon cell envelope stress.

6.2 Rcs system activation by RcsF-IM slows down growth and shortens cellular length

To induce the Rcs system in a controlled dose-dependent manner and monitor the cellular consequences, we used an IM localized version of RcsF (RcsF-IM). The IM lipoproteins in *E. coli* encode a signal sequence called “Lol avoidance signal” right after the N-terminal lipidated cysteine residue, which controls the retention of lipoproteins at the IM - escaping recruitment by the Lol system (Okuda and Tokuda, 2011). Modifying this amino acid sequence can change the membrane that the lipoproteins are localized, between the IM and OM (Yamaguchi et al., 1988). Using this knowledge, a mutant RcsF carrying the Lol avoidance signal (aspartic acid in position +2 and +3) was previously constructed and used to study the Rcs system (Cho et al., 2014). In our study, the same mutant of RcsF-IM expressed from the IPTG inducible low-copy pNG162 vector was used in strains carrying a chromosomal *rprA::lacZ* fusion to monitor the output of the Rcs system (Cho et al., 2014; Majdalani et al., 2002).

Previous work demonstrated that the Rcs system is rapidly activated upon induction of RcsF-IM (Cho et al., 2014; Farris et al., 2010). However, the dose-dependent response of Rcs system activation and its effect on bacterial cell physiology had not been explored previously. Cells were grown with different levels of RcsF-IM induction for several hours to monitor the Rcs

system activation levels over time. Overnight bacterial cultures were diluted to OD_{578} = 0.001 and grown for 3 hours without IPTG induction in order to dilute out LacZ in the overnight culture. Then different concentrations of IPTG (0 to 20 μM) were added to induce the expression of RcsF-IM. Cell cultures were back-diluted 3 or 4 times, depending on growth rate in order to maintain them in exponential growth phase for a period of 6 to 9 hours (Upper panels of Figure 6.2A and B). I sampled cultures for measuring LacZ production (by β-galactosidase assay) to monitor the Rcs activity using the *rprA::lacZ* fusion. The result shows that the Rcs system was induced already at the first time point (~1 hour after IPTG induction) in an IPTG dose-dependent manner, and the level of activation stabilized within 2-3 hours after the induction (lower panels of Figure 6.2A and B). As shown in Figure 6.2A, there is a step-wise activation of the Rcs system between 0 to 5 μM of IPTG induction, where the system saturates, as testing higher concentrations of IPTG in an independent experiment did not activate the system further (Figure 6.2B). This experiment revealed that, rather than the Rcs system activation existing as a binary on-or-off state, I observed dose-dependent activation demonstrating a dynamic and tunable activation mechanism of the system.

Simultaneously, I observed a slow-down of growth rate upon induction of RcsF-IM expression, which stabilized when cultures reached a steady state after the second dilution, allowing me to calculate growth rates (Upper panels of Figure 6.2A and B). Interestingly, growth rate inversely correlated with RcsF-IM induction: the higher the Rcs activation the slower the growth rate (Upper panels of Figure 6.2A and B). I quantified the average Rcs activation and growth rate, and Figure 6.2C represents the IPTG dose dependency of these two measurements. As expected from Figure 6.2B, until 5 μM, both the activation of the Rcs system and growth rate changed linearly, but after 5 μM both of them saturated (Figure 6.2C). Overall, these results imply that the amount of RcsF localized in the IM, thus the level of the Rcs activation determines the growth rate.

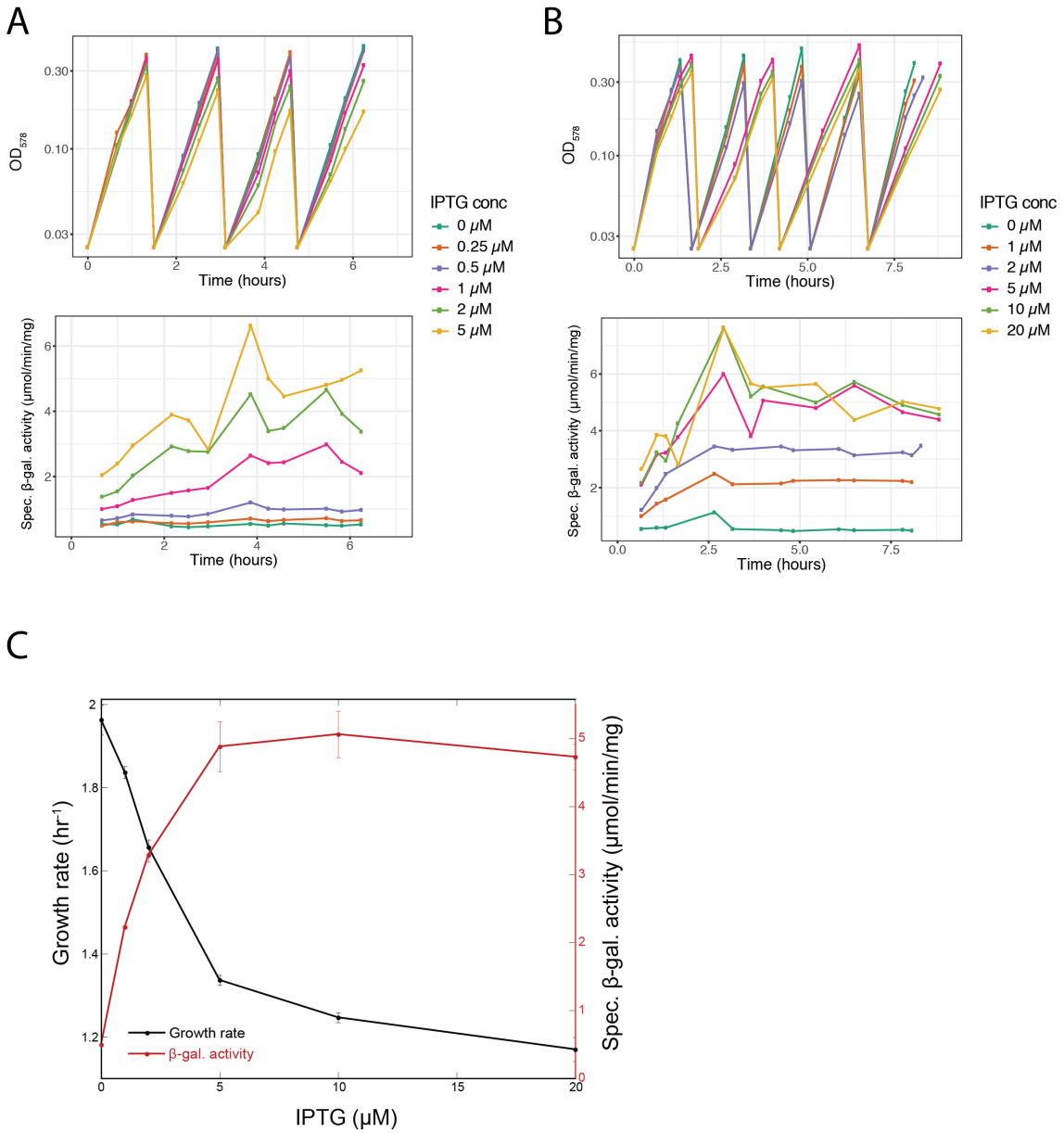


Figure 6.2: The Rcs activation and growth rate change by RcsF-IM induction. A) and B) Top panels show cell growth measured at OD_{578} and lower panels show β-galactosidase activity from LacZ production by *rprA::lacZ* over time. Cells were kept in exponential phase (constant growth rate) in LB by back-diluting with fresh media when reached an OD_{578} of 0.3. Time (x-axis) is normalized to the time point of IPTG addition. A) and B) represent one experiment each. C) Using the result of B), average growth rate (left y-axis, black) and average specific β-gal activity (right y-axis, red) were calculated for each IPTG concentration. Average growth rates quantified by curve fit to exponential curve ($y=ae^{(bx)}$, b: growth rate) from $OD_{578} = 0.025$ from second to fourth dilution. Average specific β-gal activity is the average of the last three timepoints of β-gal measurements. Error bar represents the standard deviation of the three measurements. Figures were produced by KaleidaGraph software (Kirsch and Ekerdt, 2000).

I used optical density to measure bacterial growth (Figure 6.2), however, it is known that other factors, not just the biomass, can affect the measurement of OD_{578}, including that of cell shape (Stevenson et al., 2016). We thus used single cell measurements to further dissect the effect of Rcs on cellular physiology. Cells grown in liquid batch culture with IPTG induction of RcsF-IM (15 μM IPTG) were shorter than the cells with empty vector control (Figure 6.3, results obtained by Matylda Zietek (Typas lab, unpublished)). To be able to link better changes in single-cell growth and cell size, the effect of RcsF-IM induction was examined with time-lapse microscopy in a microfluidic flow cell method (CellASIC) by Amanda Miguel (Huang Lab, Stanford University, unpublished). To visualize and quantify the Rcs system activation at the single cell level, strains carrying a plasmid expressing msfGFP from *rprA* promoter (pMZ13) were used (Figure 6.4A). Figure 6.4B (upper panel) shows the lengths of each cell over time in three conditions: empty vector, RcsF-WT and RcsF-IM, all upon maximal induction with 15 μM IPTG at the indicated time point. Cells with RcsF-IM were shorter compared to empty and RcsF-WT plasmids, with more pronounced difference at later time points at 80 and 100 minutes (Figure 6.4A), recapitulating the batch culture result (Figure 6.3).

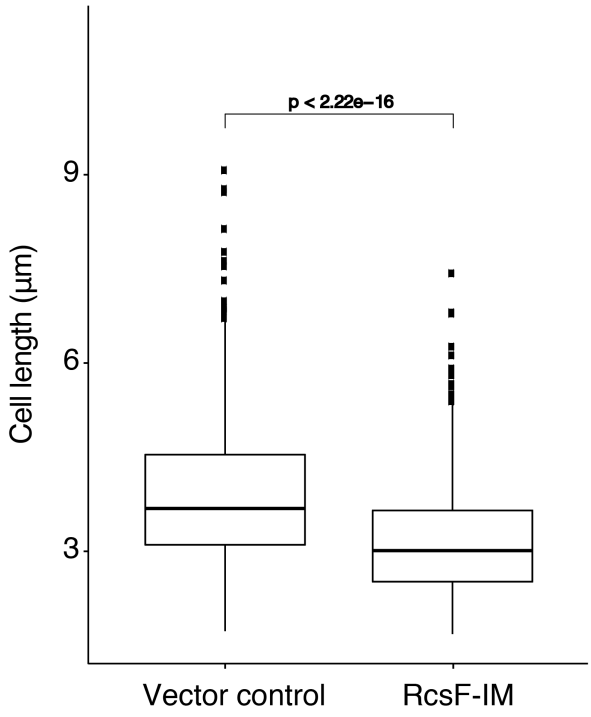


Figure 6.3: Cellular length change by RcsF-IM induction. The strain used lack endogenous RcsF (Δ*rcsF*) with transformed low-copy plasmid (pNG162) either empty vector control or the plasmid containing RcsF-IM. Cells were induced with IPTG 15 μM for 1 hour, fixed by formaldehyde, and imaged with a widefield microscopy. >500 cells were quantified for each sample. The significant difference in cell length between induction of the two plasmids were shown by t-test (p-value < 2.22e-16). The data represented is from one experiment. Experiment and quantification by Matylda Zietek.

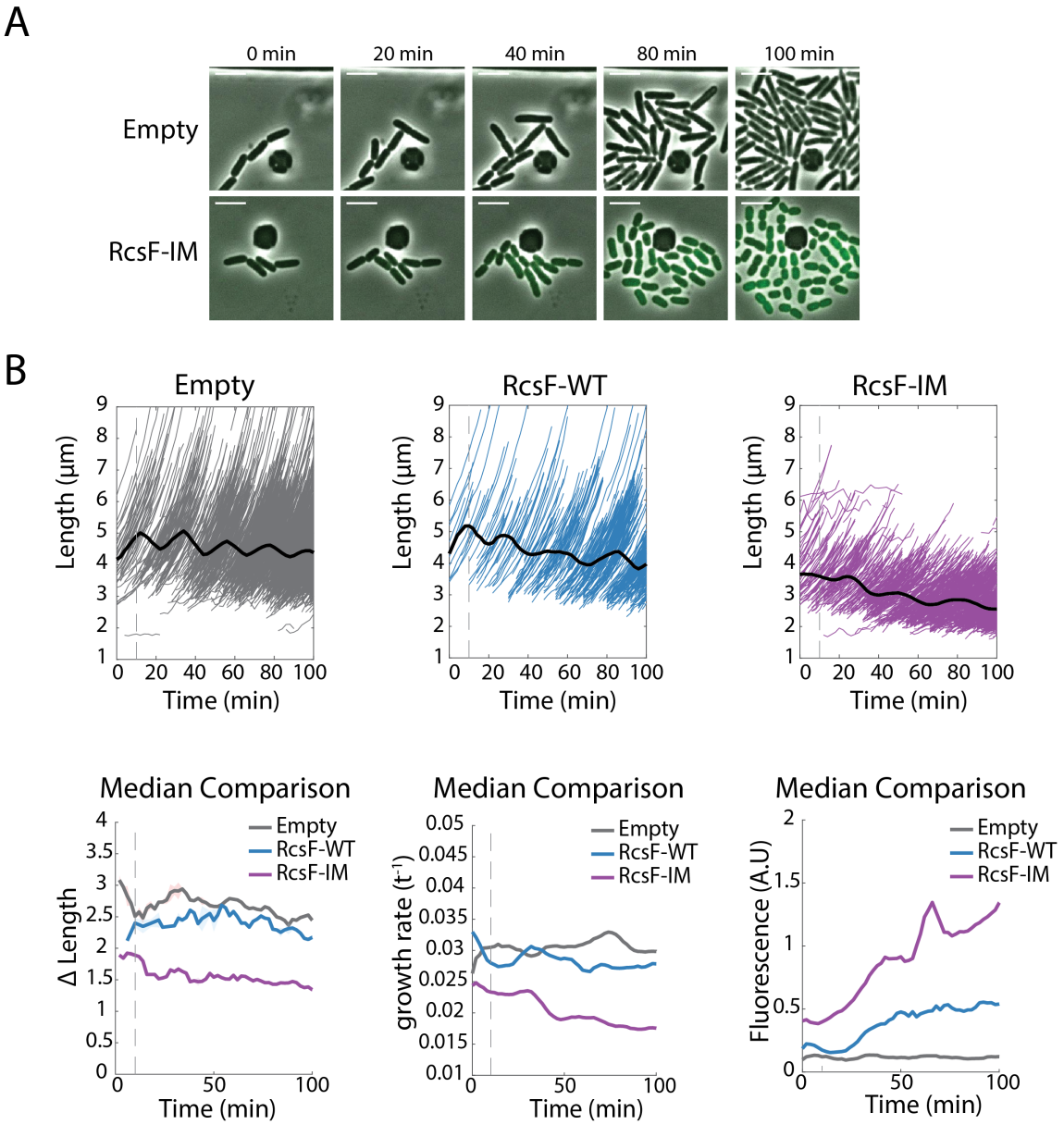


Figure 6.4: CellASIC microfluidic flow cell measurement of cells with RcsF-IM. A) Time-lapse images of cells with empty vector or with RcsF-IM, both induced by 15 µM IPTG. Cells carry a plasmid expressing smfGFP under *rprA* promoter (pMZ13). B) Top panels: each line indicates length change of a single cell, and black lines indicates the median of all cells across time (Empty vector, RcsF-WT and RcsF-IM, all induced by 15 µM IPTG). IPTG was added at the time shown by the grey vertical line. Lower panels: 1. Δ length (increase of cell length from one division to next) 2. Growth rate (increase in the pixel of cell length per minute), 3. Average fluorescence signal for *rprA::smfGFP*. Data is from one experiment. Data acquisition and figure construction by Amanda Miguel.

Using the time-lapse information obtained from the microfluidic experiment, we investigated the possible mechanism how the Rcs system shortens the cell length. RcsF-IM cells showed a decrease in growth rate almost instantaneously upon IPTG induction (Figure 6.4B, lower middle panel), as observed by OD measurement in the batch culture (Figure 6.2A and B). However, the growth rate stabilized around 50 minutes, suggesting that the cells have reached a new steady state (Figure 6.4B, lower middle panel). Whereas the length of the cells shortened throughout the measurement time with RcsF-IM (Figure 6.4A and Figure 6.4B, upper right panel). This phenomenon can be explained using the third measurement: Δ Length (ΔL) (Figure 6.4B, lower left panel). ΔL is the length difference between the two divisions, i.e. how much a cell elongates before the next division, and it is what bacteria keep constant when growing in steady state to maintain a cell size homeostasis (Campos et al., 2014). ΔL dropped right after the IPTG induction of RcsF-IM, and only changes marginally afterwards (Figure 6.4B, lower left panel). Thus, once Rcs system is activated, the cells quickly respond by growing at a slower rate and elongating less per cell-cycle (fixed to a new ΔL), which leads to a gradual decrease in cell length as multiple cell division cycles are needed to fix to a new cell length. Our results together suggest that the Rcs system changes the growth rate at the single cell level, leading to decreased elongation per cell cycle, and hence resulting in shorter cells.

6.3 Rcs system activation by RcsF-IM increases FtsZ level

Knowing that Rcs system activation can modulate cell elongation, together with growth rate and cell length, we hypothesized that the Rcs system may directly impact the counter part of the cell cycle: cell division. RcsB has been shown to regulate the promoter of *ftsA* and *ftsZ,* both central components of cell division (Carballes et al., 1999). Whereas, the levels of FtsZ (and FtsA) have been previously suggested to regulate frequency and timing of cell division (Sekar et al., 2018; Ward and Lutkenhaus, 1985). Thus we hypothesized that the activation of Rcs system may change the level of important players of the cell division machinery, and modulate the frequency of division events.

To test this, I measured FtsZ protein levels upon Rcs activation, using RcsF-IM expression by IPTG induction. I observed a mild but reproducible increase in FtsZ levels upon RcsF-IM induction by 2 and 5 μM IPTG by Western blot (Figure 6.5). Cytoplasmic RecA was used as loading control and to normalize for changes in protein abundance due to changes in cell shape

(Figure 6.5). The empty plasmid control and RcsF-IM induction in a Δ*rcsB* background confirmed that increase of FtsZ is due to activation of the Rcs signaling cascade (Figure 6.5). Comparison of the different Rcs activation levels showed a small increase in FtsZ levels between 0 μM and 2 μM IPTG, but no further increase between 2 μM and 15 μM (Figure 6.5), suggesting that FtsZ induction has saturated at or before 2 μM IPTG. Even without induction (at 0 μM IPTG), cells already expressed more FtsZ compared to the empty plasmid or the Δ*rcsB* background, due to the leaky expression level of RcsF-IM from the plasmid (Figure 6.5). As discussed above, Rcs activation measured through *rprA::lacZ* did not saturate until 5 μM (Figure 6.2C), suggesting that the saturation of Rcs-activation differs between gene promoters, whereby the activation of the *ftsZ* promoter saturates earlier than that of the *rprA* promoter.

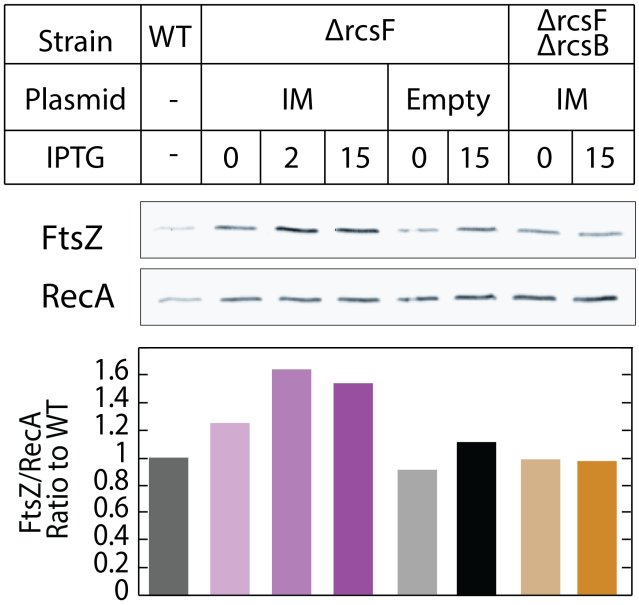


Figure 6.5: FtsZ level change by RcsF-IM induction measured by Western blot. FtsZ levels were measured using α-FtsZ antibody, and α-RecA antibody as a loading control. Cells carrying indicated plasmids were induced by respective IPTG concentration (μM) during exponential phase, after incubating for >3 hours, cultures were sampled and adjusted for OD_{578} prior to loading on a 15% SDS-PAGE gel. Western blot bands were quantified using Fiji software, and the ratio of FtsZ/RecA adjusted to WT=1 is plotted. Data is from one experiment, another blot recapitulating FtsZ/RecA ratio increase with IPTG addition (lane 2-4 in this blot) is in the Appendix.

To further investigate the consequence and functional relevance of the increased FtsZ level, we followed FtsZ localization upon Rcs activation in single cells (experiments by Amanda Miguel, Huang lab, Stanford University, unpublished). The experiments were carried out through measuring the cells on agarose pads, using a strain carrying a functional FtsZ-msfGFP chromosomal fusion to monitor FtsZ levels and localization, with and without IPTG induction

(0 and 15 μM) of RcsF-IM. Slight increase in average fluorescence signal distribution of FtsZ-msfGFP was evident upon Rcs induction (Figure 6.6A), recapitulating the FtsZ level increase observed in batch culture (Figure 6.5). In addition, the intracellular localization of FtsZ changed upon Rcs activation. FtsZ is a tublin-like protein that forms a ring at mid-cell, called the Z-ring, which initiates division events, bearing itself constriction force and recruiting a multi-complex protein machinery at division sites, the divisome (Typas et al., 2012). We observed a sharp increase in FtsZ concentration at rings upon RcsF-IM induction (Figure 6.6B). Interestingly, the effect of RcsF-IM induction by IPTG was much more pronounced for the fluorescence concentration at the rings (Figure 6.6B), compared to the average cellular concentration (Figure 6.6A). This suggests that Rcs induction leads to slight increases in FtsZ levels, and an amplified effect on FtsZ localization at the mid-cell, which is what presumably affects cell division frequencies and thereby cell growth and size.

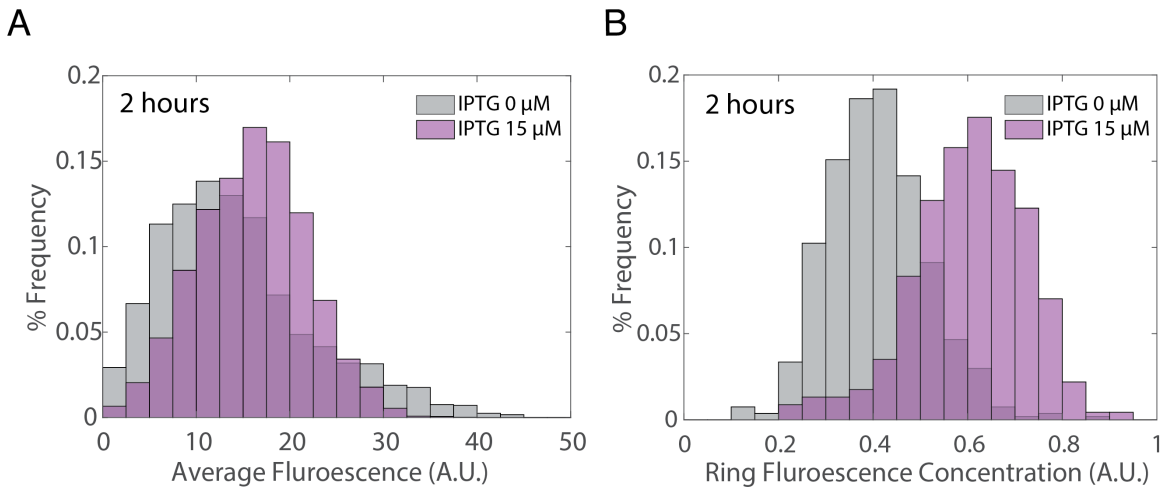


Figure 6.6: FtsZ-smfGFP level change by RcsF-IM induction. Cells with pNG162 plasmid with RcsF-IM were induced by IPTG (0 or 15 μM) for 2 hours before spotted onto agarose pads and imaged for phase contrast and fluorescence signals. Left: average fluorescence per area, Right: ring fluorescence concentration at mid-cell. Data represents one experiment. Data acquisition and figure production by Amanda Miguel.

To confirm the increase in Z-ring formation frequency, we monitored Z-ring formation in cells treated with cephalexin, which prevents cell division while allowing Z-ring formation, thus producing elongated cells with multiple Z-rings per cell (Amanda Miguel, Huang lab, Stanford University, unpublished). Using this system, Z-ring formation frequency was quantified relative to the cell length (Figure 6.7A). Induction of RcsF-IM led to higher number of Z-ring formation per cell length, i.e. Z-rings were more closely spaced (Figure 6.7A and B). This

confirms that Rcs induction increases the levels of FtsZ expression, which concentrates at each Z-ring, while forming more frequent Z-ring per cellular length, and thus leading to shorter cells.

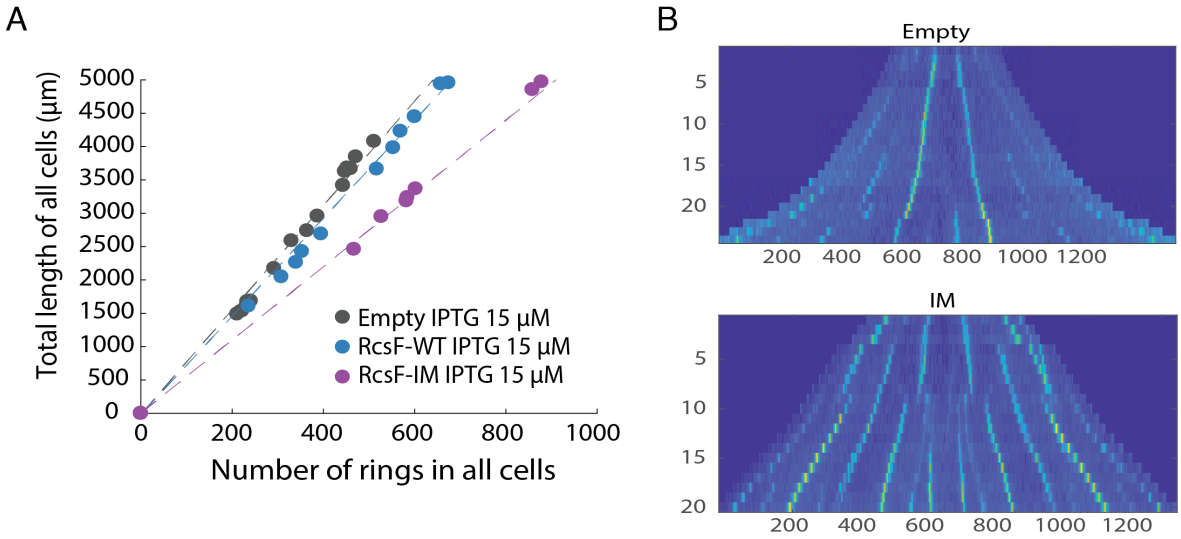


Figure 6.7: Z-ring formation in cells treated with cephalexin. Treatment with cephalexin leads to cell division arrest and filamentation, but allows for Z-ring formation. A) Number of rings observed by fluorescent microscopy for FtsZ-msfGFP in all the cells quantified (x-axis) compared to total length of all cells quantified (y-axis), B) Kymograph of FtsZ-msfGFP of cells treated with cephalexin. Upper panel: empty plasmid, lower panel: RcsF-IM induced with 15 uM IPTG. More Z-rings appear with RcsF-IM induction, with shorter distances between rings. Data acquisition and figure production by Amanda Miguel.

Overall, we, together with other collaborators, identified an unexpected effect of the Rcs system activation upon cell shape and division using mislocalized RcsF, a model surface exposed lipoprotein. Further investigation is needed to understand if these changes in cellular physiology help the cells to survive in the stressed conditions. Furthermore, if activation of the Rcs system changes the localization of RcsF, and any other surface lipoproteins in *E. coli,* is something we would like to investigate in the future.

Chapter 7: Discussion

7.1 Systematic and quantitative method for surface protein identification

Accumulating evidence has demonstrated that lipoprotein exposure at the cell surface is a wide spread phenomenon common to many bacterial species (Konovalova and Silhavy, 2015; Szewczyk and Collet, 2016). However, the number of OM lipoproteins (OMLPs) that are cell-surface exposed and the mechanisms leading to the surface translocation remains poorly understood. I therefore developed a method, SPQ (Surface Proteome Quantification), wherein protein surface exposure is systematically and quantitatively determined using a biotin-conjugated primary-amine cross-linker (NHS-LC-LC-Biotin) coupled with quantitative MS analysis using multiplexed TMT labelling. I established this workflow in *Escherichia coli*, the model organism of Gram-negative bacteria.

Contrary to the original report (Cowles et al., 2011), the NHS-LC-LC-Biotin is not entirely cell-impermeable. From my data, the preference of this biotin to label proteins at the surface of *E. coli* was confirmed as reported (Cowles et al., 2011), however I observed clear leakage of the probe throughout the cell albeit in a gradient manner. Through optimization of the method, multiple factors contributing to enrichment of surface proteins were examined, including label incubation time and buffer composite for labelled protein pull-down (Chapter 3). Several studies in the past have applied different cross-linkable biotin derivatives to label the surface proteins in Gram-negative bacteria (Monteiro et al., 2018; Surendran et al., 2015). The most recent work with surface labelling in *E. coli* was carried out comparing labelling by the two labels: Sulfo-NHS-SS-Biotin and Sulfo-NHS-PEG4-bismannose-SS-Biotin (Monteiro et al., 2018). They also observed some permeating properties of both of these biotin reagents across the OM, and emphasized the importance of short biotin incubation times (Monteiro et al., 2018). Furthermore, they comment on the need for "molecules truly impermeable to the bacterial OM," which I agree from my results that more stringent approaches are needed (Monteiro et al., 2018). Despite this, due to the quantitative nature of the SPQ method I was able to identify surface proteins systematically. Moreover, the SPQ method takes advantage of the gradient labelling throughout the cell, thus providing an alternative method to identify protein localization across the cell, not only at the cell-surface.

The SPQ assay identified surface exposed lipoproteins in *E. coli* which were previously shown as well as those not observed on the cell surface before. I identified ~ 37% of the detected lipoproteins (23 out of 62 quantified OMLPs), including previously shown surface lipoprotein Lpp (Cowles et al., 2011), to be surface exposed by using a 90 percentile of non-surface IM protein as a cutoff for cells in exponential phase (Chapter 3). The proportion increased even further in stationary phase, ~ 49% of OMLPs identified as surface exposed using the same cutoff as exponential phase samples (Chapter 4). Furthermore, another 20 OMLPs, including BamC and Pal reported before to be surface exposed (Surendran et al., 2015; Webb et al., 2012), were identified as putative candidates for surface exposure using an 80 percentile cutoff (Chapter 3). I therefore conclude that the occurrence of lipoprotein surface exposure in *E. coli* is more prevalent than previously appreciated. While this is the first study describing such high level of lipoprotein surface localization in *E. coli*, studies in another Gram-negative, *Bacteroides fragilis* (Wilson et al., 2015), and other diderm bacteria such as *Borrelia burgdorferi* (Schulze and Zückert, 2006), have shown similarly high or even higher level of lipoprotein surface localizations. In *Borrelia*, the surface is even referred as the "default" localization of lipoproteins for the organism (Schulze and Zückert, 2006). Thus lipoprotein surface exposure is hypothesized to be a more conserved phenomenon across species than previously thought and can be systematically interrogated using the SPQ method.

As the reagent is commercially available, and the sample preparation protocol is rather simple, the method displays its future potential to be applied to other Gram-negative bacteria, including pathogens where surface proteins may interact with the host and thus are often ideal candidates as vaccine targets. Comparing surface proteomes of different bacteria can provide more insights into both immunological potential of the surface proteome, and the conservation of lipoprotein surface exposure across species.

7.2 Independent methods of protein surface identification

From the highly surface exposed lipoproteins identified in the SPQ assay, I validated 5 lipoproteins (three OMLPs - MlaA, OsmE, YedD and two IMLPs - MetQ and DcrB) for their surface localization by independent methods of dot blot and whole-cell ELISA (Chapter 3). Upon validation, I faced difficulties that there are no standardized methods to accurately assess OMLP topology in the OM. I used two methods, dot blot and whole-cell ELISA, both

measuring the antibody binding on the cell surface. While both of these methods depend on antibody accessibility to the epitope of target proteins, I managed to validate surface exposure for 5 lipoproteins using a C-terminal Flag tag. It should be noted however, that the proteins not validated in these methods still leave the possibility that the proteins are surface exposed but the Flag tag was not accessible from the extracellular face of the OM. An example of this was my positive control, RcsF, which was previously detected on the surface using an antibody against the full protein, but not with an epitope recognizing a C-terminal Flag tag (Cho et al., 2014; Konovalova et al., 2014). In addition to accessibility of epitope tags, possible interference of the tag upon protein transport, thus the localization also needs to be considered for these approaches.

Besides these two biochemical methods used to orthogonally validate the SPQ results, structural data also provides supporting evidence of lipoprotein surface exposure. Firstly, LptE is the OMLP component of the Lpt machinery that transports LPS to the cell surface, and together with the OMP LptD, it forms a protein complex in the OM (Konovalova et al., 2017). The crystal structure of the LptD-E complex showed that a domain of LptE is embedded within the lumen of LtpD and extends to the cell surface as far as the β-barrel protein LptD (Figure 7.1, PDB: 4RHB). Also for MlaA, the crystal structure of MlaA in complex with OmpF shows its horizontal alignment with OmpF, thus the MlaA domain is mostly embedded in the OM and not in the periplasm (Figure 7.1, PDB: 5NUP from *Klebsiella* homolog). Both of these structural information are from their crystal structure, which leaves the possibility of an alternative conformation may exist *in vivo*, including the surface exposure of those lipoproteins. Currently we lack structural information for most lipoproteins, especially those in complex with their interacting partners in the OM to assess their membrane topology. However, as knowledge of protein structures increase, similar observations as LptE and MlaA discussed here may support surface exposure of other lipoproteins in the future.

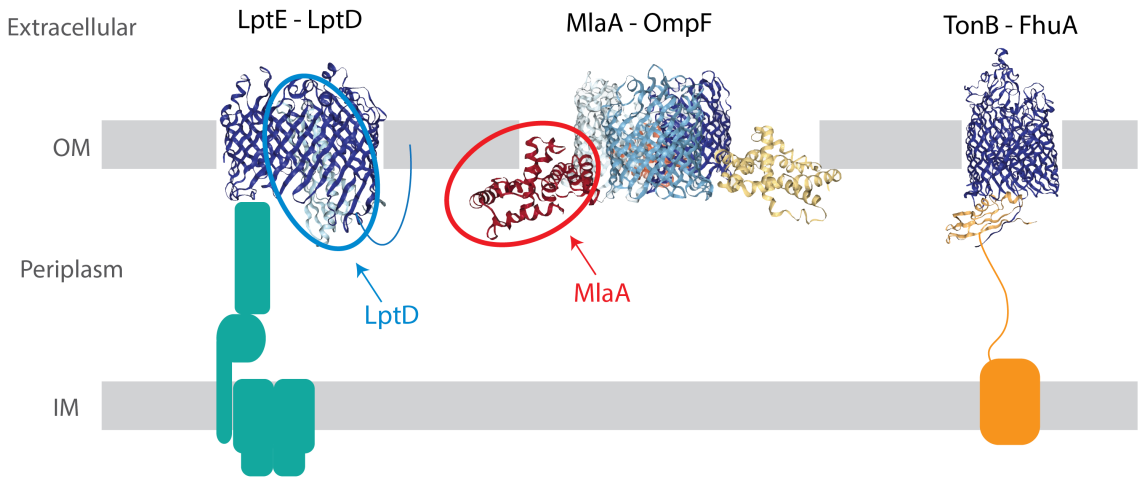


Figure 7.1: Model of protein localization using structural data. Left: LptE-LptD crystal structure (PDB: 4RHB), LptD β-barrel in blue, LptE in cyan embedded in LptD with a superimposed lipid tail connecting it to the OM. Rest of Lpt complex drawn in schematics in green. Middle: MlaA-OmpF crystal structure (PDB: 5NUP) from *Klebsiella pneumoniae* sequence. Homotrimer of heterodimer of MlaA-OmpF complex depicted, with OmpF drawn in blue and cyan colors, and MlaA in red, yellow and orange. Right: TonB-FhuA crystal structure (PDB: 2GRX), FhuA β-barrel in blue, and interacting TonB domain in orange with extended domains drawn as schematics in orange. Crystal structural data from PDB were visualized using NGL viewer (Rose et al., 2018).

Surface shaving using proteinases, such as Proteinase K, is also a commonly used method to investigate protein surface localization (Dowdell et al., 2017; Wilson et al., 2015). I tested Proteinase K treatment and observed cleavage of OMPs from the cell surface, demonstrating the assay was working as expected (Chapter 3). However, I failed to reproducibly demonstrate surface shaving of any lipoproteins tested. Previous studies in other organisms, such as *Borrelia* and *Bacteroides*, Proteinase K shaving was successfully used for identification of lipoproteins on the surface (Dowdell et al., 2017; Wilson et al., 2015). Whilst for *E. coli*, shaving of lipoproteins were not achieved in the conditions that are reflective of their endogenous state, including BamC and Lpp for which their surface shaving was achieved through over-expression and epitope tagging, respectively (Cowles et al., 2011; Webb et al., 2012). Existence and the structure of LPS differ from *E. coli* in the two organisms: *Borrelia* lack LPS, and *Bacteroides* LPS is structurally and functionally different from *E. coli* (Vatanen et al., 2016). This may be the reason of differing Proteinase K activity on surface lipoproteins. Furthermore, it suggests the possibility that the lipoproteins on the cell surface of *E. coli* are highly resistant to Proteinase K, which could be due to their conformation or interference of other cell-surface structures.

Through attempts to validate lipoprotein surface localization, I encountered technical difficulties. In particular, there are no known methods to show a lipoprotein is not surface exposed, which made it difficult to assess for false positives of the SPQ method identifying non-surface lipoprotein as surface exposed lipoproteins. However, the validation of five surface identified lipoproteins, including two IM annotated lipoproteins, provides the confidence that the SPQ method identifies surface exposed lipoproteins.

7.3 BAM dependency of lipoprotein surface exposure

Despite increasing reports of surface exposed lipoproteins in *E. coli* and other Gram-negative bacteria, our knowledge on the mechanism of lipoprotein surface transport is limited (Konovalova and Silhavy, 2015; Szewczyk and Collet, 2016). Slam protein in *Neisseria* was shown to transport lipoproteins to the surface, however, the homolog does not exist in *E. coli* (Hooda et al., 2016). I hypothesized that the BAM machinery is involved in general lipoprotein transport for the following reasons: 1. RcsF was shown to be surface exposed in a BAM dependent manner, directly interacting with BamA (Cho et al., 2014; Konovalova et al., 2014), 2. Many lipoproteins have been reported to function together with OMPs which are folded and inserted to the OM by BAM (Konovalova et al., 2017).

By applying the SPQ assay to BAM mutants, I identified the BAM dependency of lipoprotein surface exposure in *E. coli*; 15 out of the 20 most surface exposed OMLPs (i.e. high SE values) were lowered in their surface exposure upon BamA depletion or BamE deletion (Chapter 4). Out of these 20 most-surface lipoproteins, 10 lipoproteins showed clear decrease in the SE value in both BamA and BamE mutants, such as YajG and YedD (Chapter 4). Our collaborator, Jean-Francois Collet (UCL, Belgium) and his colleagues successfully identified a physical interaction between YedD and the BAM through a pull-down experiment (unpublished data through personal communication), further suggesting a direct role of the BAM in YedD surface exposure. As RcsF was previously shown to be BamA dependent for its surface translocation, I expected RcsF to show high SE values in wildtype *E. coli*, and lowered surface exposure upon BamA depletion (Cho et al., 2014; Konovalova et al., 2014). To my surprise, RcsF did not show the expected pattern; even in wildtype *E. coli*, the SPQ method did not identify RcsF as a highly surface exposed lipoprotein (Chapter 3). Possible explanation is due to the SPQ method requiring lysine residues to be surface exposed at the cell surface, in a conformation that the biotin is accessible. As RcsF is shown to reside inside the lumen of OMPs, biotin accessibility

may be very limited (Cho et al., 2014; Konovalova et al., 2014). Further investigation is required to test if other newly identified BAM dependent surface lipoproteins (e.g. YajG and OsmE) interact directly with BamA or other BAM complex members.

While BamA is the central component of the BAM complex, there are four lipoproteins (BamB -E) that form part of the BAM complex. However, knowledge regarding the function and mechanism for each of these individual BAM lipoprotein components is limited (Konovalova et al., 2017). Consistent with the SPQ data on the Δ*bamE* mutant, BamE was previously shown to be involved in BAM complex conformation although the deletion displays very mild defect in OMP folding capacity and shows no growth defect (Rigel et al., 2012), unlike the *bamA101* (BamA depletion) strain (Aoki et al., 2008). I applied the SPQ method to the Δ*bamE* mutant, and I observed unexpectedly low surface lipoprotein abundances, suggesting its specific relevance for lipoprotein transport. Consistently to previous reports, I observed only mild defects on OMP expression and surface exposure (Chapter 4). Recent studies have shown an important role of BamE for RcsF surface transport (Tata and Konovalova, 2019). They observed increased RcsF-BamA interaction while decreased RcsF-OmpA interaction in Δ*bamE* mutant through cross-linking experiments, showing that RcsF is not efficiently forming a complex with OmpA in Δ*bamE*, thus suggesting that RcsF is not transported to the cell-surface without BamE (Tata and Konovalova, 2019). This is in line with a more general role for BamE in lipoprotein transport that I observed and discussed in Chapter 4. My data shows that surface exposure of many lipoproteins (10 out of 20 top surface exposed lipoproteins in WT) depends on BamE (Chapter 4). As the function of BamE has not been fully characterized, these results suggests its unrevealed functionality in lipoprotein surface transport. Of note, it remains to be seen how other lipoprotein members of BAM complex, BamB, BamC and BamD, impact lipoprotein surface exposure, or whether this is specific to BamE. It would be especially interesting to examine the role of BamB for lipoprotein surface exposure, where the deletion of *bamB* was shown to create a synthetic lethal phenotype with the *bamE* deletion due to "lethal jamming" of RcsF, and is therefore another protein potentially involved in lipoprotein surface transport (Hart et al., 2019).

Overall, these observations support the notion that many lipoproteins, not only RcsF, interact with the BAM and are transported to the bacterial cell-surface. Unfortunately, I was unable to disentangle whether the limiting factor in this process is the BAM machinery or the OMP abundance, using the RybB overexpressing plasmid (Chapter 4). As I performed the experiment

only once, more replicates with more controls are needed to conclude why SPQ was unsuccessful in this experiment. Possible future optimization approaches include the expression of RybB in a method that can avoid the exposure of antibiotics to the cells, and also experimenting with different RybB expression levels.

Although further investigation is needed, some lipoproteins I identified to be surface exposed were shown to have interacting OMPs which are folded by BAM, such as LptE shown in Section 7.2. LptE was also shown to be required for LptD folding and insertion to the OM by BAM (Chimalakonda et al., 2011). These observations suggest that surface exposed lipoproteins, especially the ones identified to be BAM dependent, may be in the confirmation forming a complex with OMPs, as seen for RcsF (Cho et al., 2014; Konovalova et al., 2014). As OMPs and OMLPs are the only two types of protein found in the OM, one can speculate that the translocation of these two types of protein is coupled. Future protein interaction studies of those BAM dependent surface lipoproteins, to test if they directly interact with BAM, will be able to investigate this hypothesis further.

7.4 Functions of surface-exposed lipoproteins

As more lipoproteins are being identified as surface exposed, this raises the question regarding the function of the lipoproteins on the cell surface. Some lipoproteins were studied for their important function on the surface, such as SusD and BtuG in *Bacteroides* (Glenwright et al., 2017; Wexler et al., 2018). BtuG functions as a Vitamin B12 receptor, capturing B12 in the extracellular environment, and essential for the B12 uptake by OMP transporter, BtuB (Wexler et al., 2018). Wexler et al. showed the function of BtuG can be complemented when supplied extracellularly, confirming its function on the surface (Wexler et al., 2018). In *E. coli*, RcsF was originally shown that its transport efficiency to the surface act as the sensor of the downstream system, suggesting its system activating potential is lost once sequestered at the cell surface in an OMP (Cho et al., 2014; Konovalova et al., 2014). However, a recent study suggested a function for RcsF at the cell surface in sensing LPS integrity (Konovalova et al., 2016). However, for most surface lipoproteins, their function at the cell surface remains unclear.

Some surface identified lipoproteins from the SPQ assay play a role in OM biogenesis and homeostasis functions, such as MlaA and LptD as discussed above (Choi and Lee, 2019). For

these two lipoproteins, it is entirely plausible that during the process of transporting substrates to or from the cell surface (i.e. phospholipids for MlaA and lipopolysaccharides for LptE), that they become surface exposed. A similar scenario is possible for other OM biogenesis machinery proteins, such as BAM complex member lipoproteins (BamB, BamC, and BamD), and LolB from Lol lipoprotein transport system. I also identified LolB as a hit (above 90 percentile cutoff), and BamB and BamC as candidates (above 80 percentile cutoff) for surface exposure, supporting this hypothesis. Thus, these proteins may be surface exposed while carrying out their functions at the OM, but not necessarily owning specific functions on the cell surface.

Many lipoproteins I identified to be surface exposed have important and sometimes essential function in the periplasm. For example, Lpp is known to be cross-linked to peptidoglycan (PG), and controls the distance of the OM to the PG, thus the width of the periplasmic space (Asmar et al., 2017). At the same time, studies have shown that not all Lpp in the cell is bound to the PG, and those "free" Lpp can be cell-surface exposed, while its function at the surface remains unknown (Braun and Hantke, 2019; Cowles et al., 2011). Results from stationary phase cultures showed that more lipoproteins are surface exposed in stationary phase (Chapter 4), 30 OMLPs out of 61 quantified (using the same 90 percentile cutoff as exponential phase), demonstrating lipoprotein localization may change depending on the conditions. These observations raise the hypothesis that many OMLPs can have dual localizations, facing the periplasm or cell-surface depending on the cellular state. Bacteria can take advantage of this to control the functionality of the protein, i.e. the protein is only functional at one side of the OM. This led me to speculate that lipoproteins are transported to the surface to be excluded from the periplasm and thereby spatially regulating their function and access to periplasmic substrates. In line with this, accumulation of proteins in the periplasm is toxic for a cell (Mitchell and Silhavy, 2019), thus surface exposure may be a mechanism to remove unnecessary proteins from the periplasm. Once they are needed inside the cell, there may be a mechanism that lipoprotein on the surface can flip-back to facing periplasm. This means that the surface is a temporary storage localization of lipoproteins, and flipping-back of lipoproteins is presumably faster and/or more cost-effective mechanism, compared to synthesis and degradation, to regulate their abundance and function in the periplasm. Additionally, in *Bacteroides*, surface exposed lipoproteins were shown to be enriched in outer membrane vesicles (OMVs) (Valguarnera et al., 2018). In this case, surface lipoproteins are considered to be secreted, thus those lipoproteins do not carry out direct function for their originating cells. If this is also true in *E. coli*, lipoprotein surface

exposure can be a potential mechanism for regulating protein turnover; instead of degradation, bacterial may surface expose and secrete to remove lipoprotein from the cell.

For Pal, its surface localization was shown to be happening in a “all or nothing” pattern, thus only a subpopulation has Pal surface exposure (Surendran et al., 2015). If other surface lipoproteins show this “all or nothing” pattern, the lipoprotein surface exposure can be suggested to provide heterogeneity in a population. In the SPQ assay, I cannot identify this cell-to-cell heterogeneity in the population. I am interested in testing if heterogeneity is found in other lipoproteins in the future, and if so, whether it is coordinated between lipoproteins. There may be a scenario that there is a subpopulation of cells which surface expose many different lipoproteins, and the other population do not. Single-cell level investigation of lipoprotein localization is needed to answer this.

While the increased lipoprotein surface exposure in the stationary phase is intriguing, possibility of more technical and methodological background that could have derived this observation needs to be addressed. For example, differences of biophysical properties of cell surface and the OM in different growth phases may affect the efficiency of surface labelling. The biotin reagent used in the SPQ assay may penetrate the OM less efficiently in stationary phase, and thus provide less intra-cellular labelling. Testing the SPQ method in more conditions, such as different growth media and salt or metal ion concentrations are needed to provide more insights into how different cellular states and/or physical properties at the cell surface affect the SPQ assay results.

7.5 Complexity of protein intracellular localization

Protein localization provides information on where a given protein functions, with which protein it can interact and how it is processed and transported within the cell. Many different databases, such as STEPdb (Orfanoudaki and Economou, 2014), Ecocyc (Keseler et al., 2017) and Uniprot (The UniProt Consortium, 2019), provide protein localization information. Especially STEPdb, the database used as a main reference in my study, compiles sub-cellular localization information from different databases, literature and signal sequence predictions (Orfanoudaki and Economou, 2014). For cell envelope proteins, including lipoproteins, localization annotation is often based on the amino acid sequences, such as Sec and Tat signal

sequences that target client proteins to the respective translocon across the IM, and lipobox and Lol avoidance signals for lipoproteins (Orfanoudaki and Economou, 2014). My work employed two independent methods that quantitatively assess protein localization and added another layer of localization/topology information that reflects physiological protein localization.

In chapter 5, I performed quantitative analysis of a sucrose density gradient that physically separates the IM and OM, and compared the result to the SPQ assay. The sucrose gradient allows the identification of which membrane a protein is associated with (i.e. IM or OM), whereas the SPQ method using surface biotinyaltion quantifies the average distance from the surface. Using the quantitative and systematic analysis of the sucrose gradient data, proteins were clustered into four groups, IM, OM and two soluble protein clusters. Overall, the clustering successfully categorized most of IM and OM proteins to respective clusters (Chapter 5). The accuracy of clustering was especially high for abundant proteins; for top 25% abundant proteins (abundance category = high, with top3 value between 7.82 and 9.56), all IM integral annotated proteins (133 proteins) were found in the IM cluster, all but one OM proteins (46 out of 47 OMPs and OMLPs) were found in the OM cluster (Chapter 5). Thus, protein abundance is an important factor upon quantification of sucrose gradient fractionation that needs to be taken into account.

Through comparing the sucrose gradient fractionation and the SPQ data, I showed that in general, both methods mostly identify the same membrane localization, presumably because most membrane proteins are anchored and localized to either the IM or OM. Comparing the distribution of quantified values of the two methods, sucrose gradient clearly separates IM and OM proteins better (by sucrose gradient ratio values) than the SPQ method (by SE values) (Chapter 5). This is as expected from their methodology, and the more variability captured in SE values confirms that the SPQ assay is depicting another layer of information, presumably the localization of protein domains and not only the anchoring membrane. A protein that stood out from these two assays was TonB, annotated as an IM integral protein. However, as depicted in Figure 7.1, TonB has a long extending domain across the periplasm and interacts with OMPs called TonB-dependent transporters (TBDT), such as FepA and BtuB (Braun et al., 2006; Noinaj et al., 2010). TonB showed high SE values in the SPQ assay, and also did not cluster with neither OM or IM cluster groups upon sucrose gradient analysis. A high SE value might be explained from the previous structural and interaction studies revealing its interaction with OMPs at the OM (Figure 7.1, PDB: 2GRX) (Noinaj et al., 2010). However, whether TonB is

explicitly surface exposed and not only interacting with OMPs on the periplasmic side of the OM as seen in the crystal structures, will require further investigation. For sucrose gradient result, a previous study has shown that TonB enriches in both OM and IM fractions (Higgs et al., 2002; Letain and Postle, 1997), confirming my result. This common identification of TonB as an outlier supported the validity of both methods, and emphasizes the existence of proteins that are not easily annotated in a conventional manner, and the relevance of protein localization to its function.

Lipoproteins form a unique category of proteins in terms of protein localization, as they can be localized on every face of the membrane within a cell. They are anchored to membranes, but also in cytoplasm before crossing the IM and in periplasm during IM to OM transport. Additionally, lipoprotein domains can be in a different compartment from their anchoring membranes. Two IM-annotated lipoproteins, MetQ and DcrB, were identified as surface exposed in the SPQ assay, and were subsequently validated for their surface exposure by dot blot (Chapter 3). Through sucrose gradient quantification, both MetQ and DcrB were clustered to the IM cluster, confirming the annotation in STEPdb database (Chapter 5). Crystal structure of a mutant form of MetQ as a part of MetNIQ complex in the IM has been solved, and thus this IM localization is likely to be a localization state of MetQ (Nguyen et al., 2018). On the other hand, another study has shown that MetQ in *Neisseria,* conserved homolog from *E. coli,* is surface exposed (Semchenko et al., 2017). Thus, I hypothesize that there may be a subpopulation of MetQ not being part of MetNIQ complex in the IM, which is transported to the OM and surface exposed. Further investigation is needed to show the topology of these lipoproteins in *E. coli*.

Although not originally developed to study protein localization, recent proteome-wide methodologies have provided novel insights into cellular compartmentalization in *E. coli*. One such assay is Thermal Proteome Profiling (TPP). TPP measures protein thermal stability, which infers protein confirmation, protein-ligand or protein-protein interaction (Mateus et al., 2018; Savitski et al., 2014). Recently, the first *E. coli* TPP was carried out, revealing localization dependency of thermostability of proteins (Mateus et al., 2018). Within their work, they made comparisons between the following four compartments: cytoplasm, IM, periplasm and OM, showing that higher temperatures were required to melt proteins (i.e. more thermostable) the more distantly located they are from the cytoplasm (Mateus et al., 2018), which was akin to what I observed for the SE value in the SPQ assay (Chapter 3). Thus, I tested if the

thermostability, measured as melting temperature from TPP (Mateus et al., 2018), is correlating with the SE values of OM lipoproteins. The result showed the Pearson correlation of $R = 0.4$ for exponential phase, and $R = 0.68$ for stationary phase for OM lipoproteins, revealing rather high and significant correlation, especially for stationary phase, despite the limited number of OMLPs commonly identified in the two experiments (n = 37) (Figure 7.2). As discussed above, there are more lipoproteins on the surface in stationary phase, and presumably this is behind the improved correlation. Further investigation is needed to explore the mechanism and biology behind this, but the observed correlations further confirm that the SPQ method is capturing the differences in protein topology of lipoproteins. Furthermore, it suggests surface exposed lipoproteins are highly thermo-stable and therefore may reflect a rigid folding state and/or a specific property of the bacterial cell surface, particularly in stationary phase.

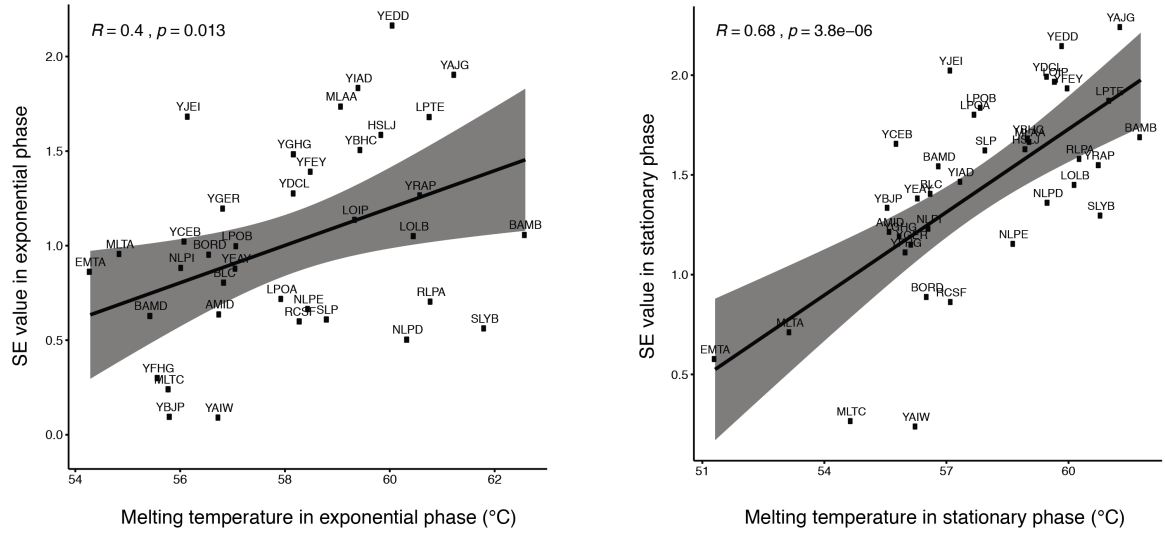


Figure 7.2: Correlation of thermostability and surface exposure of lipoproteins. Melting temperature in exponential and stationary phase (x-axis) from Thermal Proteome Profiling (TPP) data (Mateus et al., 2018), compared to surface exposure values from SPQ assay (Chapter 4, y-axis) for commonly identified 37 OM lipoproteins. Left: for exponential phase, right: for stationary phase. Pearson was used as correlation method. Pearson correlation with its p-value is shown.

7.6 Consequences of lipoprotein mislocalization

Lipoprotein mislocalization has been studied for some lipoproteins, especially for the lipoproteins involved in regulating the activation of signaling systems (Cho et al., 2014; Delhaye et al., 2019). As described in Chapter 6, the IM localized mutant of RcsF (RcsF-IM),

a surface exposed OMLP, showed RcsF-IM expression level dependent effect on the Rcs system activation and growth rate. This revealed the potential of lipoprotein surface transport efficiency, thus its cellular localization, as a mechanism to control complex signaling cascade in a dynamic manner. Using a microfluidics device, we captured the growth rate change at the single-cell level (data obtained by Amanda Miguel). Furthermore, the change in cell length by expressing RcsF-IM was in line with a previously proposed model of cell length homeostasis through adjusting the cell elongation per cell cycle (Campos et al., 2014), further confirming our observation to have physiological relevance.

Additionally, we also showed that RcsF in the IM increases the expression of FtsZ, the central component of cell division, in both average abundance in the cell and at the Z-ring it forms at mid-cell. Further investigation is needed to link this effect on FtsZ to the division events, and how this is beneficial for the cells upon Rcs system activation which is caused by cell envelope defects.

Therefore, we showed that localization of the lipoprotein RcsF can affect central cellular physiology: cell size and growth homeostasis, through a two-component system, Rcs system. RcsF is not the only known lipoprotein involved in signaling systems: NlpE for Cpx system and QseG for QseE/F system have been shown to be involved in activating signaling systems (two-component systems) in *E. coli* (Delhaye et al., 2016; Göpel and Görke, 2018). Furthermore, the localization of both lipoproteins in the periplasm is essential for signaling functions (Delhaye et al., 2016; Göpel and Görke, 2018). Many lipoproteins also remain to be functionally annotated, and these observations suggests that maybe more of them are signaling to IM components of two-component systems, with their localization in IM, OM or potentially at their cell-surface playing an important role.

7.7 Conclusion

Protein localization in Gram-negative bacteria is complex, and especially, the surface localization of lipoprotein remains a poorly defined phenomenon. To study lipoprotein surface exposure in *E. coli*, I developed an assay which allows surface proteome quantification (SPQ) to address long standing questions in the field of cell-surface lipoprotein topology. The SPQ assay revealed that surface localization of lipoproteins is a common phenomenon in *E. coli*, with 37% of quantified OMLPs (23 OMLPs) identified to be surface exposed, with the observation reaching 49% in stationary phase using the same threshold. In verifying my data I was able to confirm novel surface localization for 5 lipoproteins (DcrB, MetQ, MlaA, OsmE and YedD) using orthogonal methods. Further technological development of the SPQ method will be beneficial in the future. For example, improving the assay stringency using probes with less cell permeability will increase the reliability of the thresholds for identifying surface lipoproteins. This will also aid in application of the SPQ method to other organisms where genomes are poorly annotated and not genetically tractable. Furthermore, I was able to provide evidence that the BAM machinery is behind lipoprotein surface exposure in *E. coli*, in a manner that is strongly depending on BamA and BamE. The actual molecular involvement of BAM for lipoprotein surface transport, the precise conformation of lipoproteins at the surface, and the functional relevance of their surface exposure will need to be explored in future studies. In parallel, I performed quantitative proteomics analysis of the sucrose gradient fractionation, unraveling the protein membrane localization landscape across *E. coli* proteome for the first time. There I verified much of the previous and predicted knowledge, while identifying several cases of mis-annotations. By integrating the SPQ data with the quantitative membrane fractionation study, I found complex proteins localizations which cannot be solely explained by membrane anchoring. Lastly, the importance of lipoprotein localization upon bacterial physiology was investigated using the mislocalized model surface lipoprotein, RcsF. The direct effect of RcsF mislocalization upon cell size and growth was revealed. Overall, this study contributed to our understanding of lipoprotein localization in *E. coli*, with a strong focus on their surface exposure. This opens up many questions to be further investigated to reveal both the topology and function of lipoproteins in bacteria in the future.

References

Anwari, K., Poggio, S., Perry, A., Gatsos, X., Ramarathinam, S.H., Williamson, N.A., Noinaj, N., Buchanan, S., Gabriel, K., Purcell, A.W., et al. (2010). A Modular BAM Complex in the Outer Membrane of the α-Proteobacterium Caulobacter crescentus. PLoS One *5*, e8619.

Aoki, S.K., Malinverni, J.C., Jacoby, K., Thomas, B., Pamma, R., Trinh, B.N., Remers, S., Webb, J., Braaten, B.A., Silhavy, T.J., et al. (2008). Contact-dependent growth inhibition requires the essential outer membrane protein BamA (YaeT) as the receptor and the inner membrane transport protein AcrB. Mol. Microbiol. *70*, 323–340.

Arnold, M.F.F., Caro-Hernandez, P., Tan, K., Runti, G., Wehmeier, S., Scocchi, M., Doerrler, W.T., Walker, G.C., and Ferguson, G.P. (2014). Enteric YaiW Is a Surface-Exposed Outer Membrane Lipoprotein That Affects Sensitivity to an Antimicrobial Peptide. J. Bacteriol. *196*, 436–444.

Ashraf, K.U., Josts, I., Mosbahi, K., Kelly, S.M., Byron, O., Smith, B.O., and Walker, D. (2016). The Potassium Binding Protein Kbp Is a Cytoplasmic Potassium Sensor. Structure *24*, 741–749.

Asmar, A.T., and Collet, J.-F. (2018). Lpp, the Braun lipoprotein, turns 50—major achievements and remaining issues. FEMS Microbiol. Lett. *365*, 1–8.

Asmar, A.T., Ferreira, J.L., Cohen, E.J., Cho, S.-H., Beeby, M., Hughes, K.T., and Collet, J.-F. (2017). Communication across the bacterial cell envelope depends on the size of the periplasm. PLOS Biol. *15*, e2004303.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. *2*, 2006.0008.

Bean, G.J., Flickinger, S.T., Westler, W.M., McCully, M.E., Sept, D., Weibel, D.B., and Amann, K.J. (2009). A22 Disrupts the Bacterial Actin Cytoskeleton by Directly Binding and Inducing a Low-Affinity State in MreB. Biochemistry *48*, 4852–4857.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B *57*, 289–300.

Bernal-Cabas, M., Ayala, J.A., and Raivio, T.L. (2015). The Cpx Envelope Stress Response Modifies Peptidoglycan Cross-Linking via the l,d-Transpeptidase LdtD and the Novel Protein YgaU. J. Bacteriol. *197*, 603–614.

Bouveret, E., Derouiche, R., Rigal, A., Lloubes, R., Lazdunski, C., and Benedetti, H. (1995). Peptidoglycan-associated lipoprotein-TolB interaction. A possible key to explaining the formation of contact sites between the inner and outer membranes of Escherichia coli. J. Biol. Chem. *270*, 11071–11077.

Bouvier, M., Sharma, C.M., Mika, F., Nierhaus, K.H., and Vogel, J. (2008). Small RNA Binding to 5′ mRNA Coding Region Inhibits Translational Initiation. Mol. Cell *32*, 827–837.

Braun, V. (2018). The Outer Membrane Took Center Stage. Annu. Rev. Microbiol. *72*, 1–24.

Braun, V., and Hantke, K. (2019). Lipoproteins: Structure, Function, Biosynthesis. A. Kuhn, ed. (Cham: Springer International Publishing), pp. 39–77.

Braun, M., Kadner, R.J., Deisenhofer, J., Heijne, W., Mende, J., Braun, V., Kamp, R.M., Mohanty, a K., Wiener, M.C., Hennecke, F., et al. (2006). Structure of TonB in Complex. Science (80-.). 1399–1402.

Brochado, A.R., Telzerow, A., Bobonis, J., Banzhaf, M., Mateus, A., Selkrig, J., Huth, E., Bassler, S., Zamarreño Beas, J., Zietek, M., et al. (2018). Species-specific activity of antibacterial drug combinations. Nature *559*, 259–263.

Calhoun, L.N., and Kwon, Y.M. (2011). Structure, function and regulation of the DNA-binding protein Dps and its role in acid and oxidative stress resistance in Escherichia coli: a review. J. Appl. Microbiol. *110*, 375–386.

Campanacci, V., Bishop, R.E., Blangy, S., Tegoni, M., and Cambillau, C. (2006). The membrane bound bacterial lipocalin Blc is a functional dimer with binding preference for lysophospholipids. FEBS Lett. *580*, 4877–4883.

Campos, M., Surovtsev, I.V., Kato, S., Paintdakhi, A., Beltran, B., Ebmeier, S.E., and Jacobs-Wagner, C. (2014). A Constant Size Extension Drives Bacterial Cell Size Homeostasis. Cell *159*, 1433–1446.

Carballes, F., Bertrand, C., Bouche, J.-P., and Cam, K. (1999). Regulation of Escherichia coli cell division genes ftsA and ftsZ by the two-component system rcsC-rcsB. Mol. Microbiol. *34*, 442–450.

Cascales, E., Gavioli, M., Sturgis, J.N., and Lloubes, R. (2000). Proton motive force drives the interaction of the inner membrane TolA and outer membrane Pal proteins in Escherichia coli. Mol. Microbiol. *38*, 904–915.

Chaba, R., Alba, B.M., Guo, M.S., Sohn, J., Ahuja, N., Sauer, R.T., and Gross, C.A. (2011). Signal integration by DegS and RseB governs the σ E -mediated envelope stress response in Escherichia coli. Proc. Natl. Acad. Sci. *108*, 2106–2111.

Chimalakonda, G., Ruiz, N., Chng, S.-S., Garner, R.A., Kahne, D., and Silhavy, T.J. (2011). Lipoprotein LptE is required for the assembly of LptD by the -barrel assembly machine in the outer membrane of Escherichia coli. Proc. Natl. Acad. Sci. *108*, 2492–2497.

Cho, S., Szewczyk, J., Pesavento, C., Zietek, M., Banzhaf, M., Roszczenko, P., Asmar, A., Laloux, G., Hov, A., Leverrier, P., et al. (2014). Detecting Envelope Stress by Monitoring β-Barrel Assembly. Cell *159*, 1652–1664.

Choi, U., and Lee, C.-R. (2019). Antimicrobial Agents That Inhibit the Outer Membrane Assembly Machines of Gram-Negative Bacteria. J. Microbiol. Biotechnol. *29*, 1–10.

Chung, C.T., Niemela, S.L., and Miller, R.H. (1989). One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. Proc. Natl. Acad. Sci. *86*, 2172–2175.

Costanzo, A., and Ades, S.E. (2006). Growth Phase-Dependent Regulation of the Extracytoplasmic Stress Factor, E, by Guanosine 3',5'-Bispyrophosphate (ppGpp). J. Bacteriol. *188*, 4627–4634.

Cowles, C.E., Li, Y., Semmelhack, M.F., Cristea, I.M., and Silhavy, T.J. (2011). The free and bound forms of Lpp occupy distinct subcellular locations in Escherichia coli. Mol. Microbiol. *79*, 1168–1181.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl. Acad. Sci. *97*, 6640–6645.

Delhaye, A., Collet, J., and Laloux, G. (2016). Fine-Tuning of the Cpx Envelope Stress Response Is Required for Cell Wall Homeostasis in Escherichia coli. MBio *7*, e00047-16.

Delhaye, A., Laloux, G., and Collet, J.-F. (2019). The Lipoprotein NlpE Is a Cpx Sensor That Serves as a Sentinel for Protein Sorting and Folding Defects in the Escherichia coli Envelope. J. Bacteriol. *201*, 1–12.

Dowdell, A.S., Murphy, M.D., Azodi, C., Swanson, S.K., Florens, L., Chen, S., and Zückert, W.R. (2017). Comprehensive spatial analysis of the Borrelia burgdorferi lipoproteome reveals a compartmentalization bias toward the bacterial surface. J. Bacteriol. *199*, e00658-16.

Dreisbach, A., van Dijl, J.M., and Buist, G. (2011). The cell surface proteome of Staphylococcus aureus. Proteomics *11*, 3154–3168.

Drummelsmith, J. (2000). Translocation of group 1 capsular polysaccharide to the surface of Escherichia coli requires a multimeric complex in the outer membrane. EMBO J. *19*, 57–66.

Dunstan, R.A., Hay, I.D., Wilksch, J.J., Schittenhelm, R.B., Purcell, A.W., Clark, J., Costin, A., Ramm, G., Strugnell, R.A., and Lithgow, T. (2015). Assembly of the secretion pores GspD, Wza and CsgG into bacterial outer membranes does not require the Omp85 proteins BamA or TamA. Mol. Microbiol. *97*, 616–629.

Ekiert, D.C., Bhabha, G., Isom, G.L., Greenan, G., Ovchinnikov, S., Henderson, I.R., Cox, J.S., and Vale, R.D. (2017). Architectures of Lipid Transport Systems for the Bacterial Outer Membrane. Cell *169*, 273-285.e17.

Esbelin, J., Santos, T., Ribière, C., Desvaux, M., Viala, D., Chambon, C., and Hébraud, M. (2018). Comparison of three methods for cell surface proteome extraction of Listeria monocytogenes biofilms. Omi. A J. Integr. Biol. *22*, 779–787.

Farris, C., Sanowar, S., Bader, M.W., Pfuetzner, R., and Miller, S.I. (2010). Antimicrobial peptides activate the Rcs regulon through the outer membrane lipoprotein RcsF. J. Bacteriol. *192*, 4894–4903.

Filippova, E. V., Zemaitaitis, B., Aung, T., Wolfe, A.J., and Anderson, W.F. (2018). Structural Basis for DNA Recognition by the Two-Component Response Regulator RcsB. MBio *9*, e01993-17.

Franken, H., Mathieson, T., Childs, D., Sweetman, G.M.A., Werner, T., Tögel, I., Doce, C., Gade, S., Bantscheff, M., Drewes, G., et al. (2015). Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. Nat. Protoc. *10*, 1567–1593.

Galdiero, S., Galdiero, M., and Pedone, C. (2007). beta-Barrel membrane bacterial proteins: structure, function, assembly and interaction with lipids. Curr. Protein Pept. Sci. *8*, 63–82.

Glenwright, A.J., Pothula, K.R., Bhamidimarri, S.P., Chorev, D.S., Baslé, A., Firbank, S.J., Zheng, H., Robinson, C. V., Winterhalter, M., Kleinekathöfer, U., et al. (2017). Structural basis for nutrient acquisition by dominant members of the human gut microbiota. Nature *541*, 407–411.

Gogol, E.B., Rhodius, V.A., Papenfort, K., Vogel, J., and Gross, C.A. (2011). Small RNAs endow a transcriptional activator with essential repressor functions for single-tier control of a global stress regulon. Proc. Natl. Acad. Sci. *108*, 12875–12880.

Goolab, S., Roth, R.L., van Heerden, H., and Crampton, M.C. (2015). Analyzing the molecular mechanism of lipoprotein localization in Brucella. Front. Microbiol. *6*, 1–20.

Göpel, Y., and Görke, B. (2018). Interaction of lipoprotein QseG with sensor kinase QseE in the periplasm controls the phosphorylation state of the two-component system QseE/QseF in Escherichia coli. PLOS Genet. *14*, e1007547.

Grabowicz, M. (2018). Lipoprotein Transport: Greasing the Machines of Outer Membrane Biogenesis. BioEssays *40*, 1700187.

Guo, X.-P., and Sun, Y.-C. (2017). New Insights into the Non-orthodox Two Component Rcs Phosphorelay System. Front. Microbiol. *8*, 1–11.

Hart, E.M., Gupta, M., Wühr, M., and Silhavy, T.J. (2019). The Synthetic Phenotype of Δ bamB Δ bamE Double Mutants Results from a Lethal Jamming of the Bam Complex by the Lipoprotein RcsF. MBio *10*, 1–12.

Hempel, K., Pané-Farré, J., Otto, A., Sievers, S., Hecker, M., and Becher, D. (2010). Quantitative Cell Surface Proteome Profiling for SigB-Dependent Protein Expression in the Human Pathogen Staphylococcus aureus via Biotinylation Approach. J. Proteome Res. *9*, 1579–1590.

Higgs, P.I., Letain, T.E., Merriam, K.K., Burke, N.S., Park, H., Kang, C., and Postle, K. (2002). TonB Interacts with Nonreceptor Proteins in the Outer Membrane of Escherichia coli. J. Bacteriol. *184*, 1640–1648.

Hoch, J.A. (2000). Two-component and phosphorelay signal transduction. Curr. Opin. Microbiol. *3*, 165–170.

Hooda, Y., Lai, C.C.-L., Judd, A., Buckwalter, C.M., Shin, H.E., Gray-Owen, S.D., and Moraes, T.F. (2016). Slam is an outer membrane protein that is required for the surface display of lipidated virulence factors in Neisseria. Nat. Microbiol. *1*, 16009.

Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A., and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics *18 Suppl 1*, S96-104.

Hughes, C.S., Foehr, S., Garfield, D.A., Furlong, E.E., Steinmetz, L.M., and Krijgsveld, J. (2014). Ultrasensitive proteome analysis using paramagnetic bead technology. Mol. Syst. Biol. *10*, 757–757.

Hughes, C.S., Moggridge, S., Müller, T., Sorensen, P.H., Morin, G.B., and Krijgsveld, J. (2019). Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nat. Protoc. *14*, 68–85.

J. H. Miller (1972). Experiments in Molecular Genetics (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Kenedy, M.R., Lenhart, T.R., and Akins, D.R. (2012). The role of Borrelia burgdorferi outer surface proteins. FEMS Immunol. Med. Microbiol. *66*, 1–19.

Keseler, I.M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martínez, C., Caspi, R., Fulcher, C., Gama-Castro, S., Kothari, A., Krummenacker, M., et al. (2017). The EcoCyc database: reflecting new knowledge about Escherichia coli K-12. Nucleic Acids Res. *45*, D543–D550.

Kirsch, P.D., and Ekerdt, J.G. (2000). KaleidaGraph: Graphing and Data Analysis. Version 3.5 for Windows Synergy Software, 2457 Perkiomen Ave., Reading, PA 19606-2049. www.Synergy.com. $155.00. J. Am. Chem. Soc. *122*, 11755–11755.

Konovalova, A., and Silhavy, T.J. (2015). Outer membrane lipoprotein biogenesis: Lol is not the end. Philos. Trans. R. Soc. B Biol. Sci. *370*, 20150030.

Konovalova, A., Perlman, D.H., Cowles, C.E., and Silhavy, T.J. (2014). Transmembrane domain of surface-exposed outer membrane lipoprotein RcsF is threaded through the lumen of -barrel proteins. Proc. Natl. Acad. Sci. *111*, E4350–E4358.

Konovalova, A., Mitchell, A.M., and Silhavy, T.J. (2016). A lipoprotein/β-barrel complex monitors lipopolysaccharide integrity transducing information across the outer membrane. Elife *5*, 1–17.

Konovalova, A., Kahne, D.E., and Silhavy, T.J. (2017). Outer Membrane Biogenesis. Annu. Rev. Microbiol. *71*, 539–556.

Lacqua, A., Wanner, O., Colangelo, T., Martinotti, M.G., and Landini, P. (2006). Emergence of Biofilm-Forming Subpopulations upon Exposure of Escherichia coli to Environmental Bacteriophages. Appl. Environ. Microbiol. *72*, 956–959.

Lauber, F., Cornelis, G.R., and Renzi, F. (2016). Identification of a New Lipoprotein Export Signal in Gram-Negative Bacteria. MBio *7*, e01232-16.

Lenhart, T.R., and Akins, D.R. (2010). Borrelia burgdorferi locus BB0795 encodes a BamA orthologue required for growth and efficient localization of outer membrane proteins. Mol. Microbiol. *75*, 692–709.

Letain, T.E., and Postle, K. (1997). TonB protein appears to transduce energy by shuttling between the cytoplasmic membrane and the outer membrane in Escherichia coli. Mol. Microbiol. *24*, 271–283.

Mahoney, T.F., Ricci, D.P., and Silhavy, T.J. (2016). Classifying β-Barrel Assembly Substrates by Manipulating Essential Bam Complex Members. J. Bacteriol. *198*, 1984–1992.

Majdalani, N., and Gottesman, S. (2005). THE RCS PHOSPHORELAY: A Complex Signal Transduction System. Annu. Rev. Microbiol. *59*, 379–405.

Majdalani, N., Hernandez, D., and Gottesman, S. (2002). Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. Mol. Microbiol. *46*, 813–826.

Majdalani, N., Heck, M., Stout, V., and Gottesman, S. (2005). Role of RcsF in Signaling to the Rcs Phosphorelay Pathway in Escherichia coli. J. Bacteriol. *187*, 6770–6778.

Malinverni, J.C., and Silhavy, T.J. (2009). An ABC transport system that maintains lipid asymmetry in the Gram-negative outer membrane. Proc. Natl. Acad. Sci. *106*, 8009–8014.

Manning, P.A., Beutin, L., and Achtman, M. (1980). Outer membrane of Escherichia coli: properties of the F sex factor traT protein which is involved in surface exclusion. J. Bacteriol. *142*, 285–294.

Mateus, A., Bobonis, J., Kurzawa, N., Stein, F., Helm, D., Hevler, J., Typas, A., and Savitski, M.M. (2018). Thermal proteome profiling in bacteria: probing protein state in vivo. Mol. Syst. Biol. *14*, e8242.

Mitchell, A.M., and Silhavy, T.J. (2019). Envelope stress responses: balancing damage repair and toxicity. Nat. Rev. Microbiol. *17*, 417–428.

Mitchell, A.M., Srikumar, T., and Silhavy, T.J. (2018). Cyclic Enterobacterial Common Antigen Maintains the Outer Membrane Permeability Barrier of Escherichia coli in a Manner Controlled by YhdP. MBio *9*, 1–16.

Monteiro, R., Chafsey, I., Leroy, S., Chambon, C., Hébraud, M., Livrelli, V., Pizza, M., Pezzicoli, A., and Desvaux, M. (2018). Differential biotin labelling of the cell envelope proteins in lipopolysaccharidic diderm bacteria: Exploring the proteosurfaceome of Escherichia coli using sulfo-NHS-SS-biotin and sulfo-NHS-PEG4-bismannose-SS-biotin. J. Proteomics *181*, 16–23.

Nakayama, T., and Zhang-Akiyama, Q.-M. (2017). pqiABC and yebST , Putative mce Operons of Escherichia coli, Encode Transport Pathways and Contribute to Membrane Integrity. J. Bacteriol. *199*, 1–13.

Nguyen, P.T., Lai, J.Y., Lee, A.T., Kaiser, J.T., and Rees, D.C. (2018). Noncanonical role for the binding protein in substrate uptake by the MetNI methionine ATP Binding Cassette (ABC) transporter. Proc. Natl. Acad. Sci. *115*, E10596–E10604.

Nichols, R.J., Sen, S., Choo, Y.J., Beltrao, P., Zietek, M., Chaba, R., Lee, S., Kazmierczak, K.M., Lee, K.J., Wong, A., et al. (2011). Phenotypic Landscape of a Bacterial Cell. Cell *144*, 143–156.

Noinaj, N., Guillier, M., Barnard, T.J., and Buchanan, S.K. (2010). TonB-dependent transporters: regulation, structure, and function. Annu. Rev. Microbiol. *64*, 43–60.

Noinaj, N., Gumbart, J.C., and Buchanan, S.K. (2017). The β-barrel assembly machinery in motion. Nat. Rev. Microbiol. *13*, 43–48.

Okuda, S., and Tokuda, H. (2011). Lipoprotein Sorting in Bacteria. Annu. Rev. Microbiol. *65*, 239–259.

Oliveros, J.C. (2015). Venny. An interactive tool for comparing lists with Venn's diagrams.

Orfanoudaki, G., and Economou, A. (2014). Proteome-wide subcellular topologies of E. coli polypeptides database (STEPdb). Mol. Cell. Proteomics *13*, 3674–3687.

Pathania, R., Zlitni, S., Barker, C., Das, R., Gerritsma, D.A., Lebert, J., Awuah, E., Melacini, G., Capretta, F.A., and Brown, E.D. (2009). Chemical genomics in Escherichia coli identifies an inhibitor of bacterial lipoprotein targeting. Nat. Chem. Biol. *5*, 849–856.

Perkins, D.N., Pappin, D.J.C., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis *20*, 3551–3567.

Pfeiffer, V., Sittka, A., Tomer, R., Tedin, K., Brinkmann, V., and Vogel, J. (2007). A small non-coding RNA of the invasion gene island (SPI-1) represses outer membrane protein synthesis from the Salmonella core genome. Mol. Microbiol. *66*, 1174–1191.

Pride, A.C., Herrera, C.M., Guan, Z., Giles, D.K., and Trent, M.S. (2013). The Outer Surface Lipoprotein VolA Mediates Utilization of Exogenous Lipids by Vibrio cholerae. MBio *4*.

R Core Team (2019). R: A Language and Environment for Statistical Computing.

Reichel, M., Liao, Y., Rettel, M., Ragan, C., Evers, M., Alleaume, A.-M., Horos, R., Hentze, M.W., Preiss, T., and Millar, A.A. (2016). In Planta Determination of the mRNA-Binding Proteome of Arabidopsis Etiolated Seedlings. Plant Cell *28*, 2435–2452.

Ricci, D.P., Hagan, C.L., Kahne, D., and Silhavy, T.J. (2012). Activation of the Escherichia coli -barrel assembly machine (Bam) is required for essential components to interact properly with substrate. Proc. Natl. Acad. Sci. *109*, 3487–3491.

Rigel, N.W., Schwalm, J., Ricci, D.P., and Silhavy, T.J. (2012). BamE Modulates the Escherichia coli Beta-Barrel Assembly Machine Component BamA. J. Bacteriol. *194*, 1002–1008.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Robinson, L.S., Ashman, E.M., Hultgren, S.J., and Chapman, M.R. (2006). Secretion of curli fibre subunits is mediated by the outer membrane-localized CsgG protein. Mol. Microbiol. *59*, 870–881.

Roesli, C., Neri, D., and Rybak, J.-N. (2006). In vivo protein biotinylation and sample preparation for the proteomic identification of organ- and disease-specific antigens accessible from the vasculature. Nat. Protoc. *1*, 192–199.

Rose, A.S., Bradley, A.R., Valasatava, Y., Duarte, J.M., Prlić, A., and Rose, P.W. (2018). NGL viewer: web-based molecular graphics for large complexes. Bioinformatics *34*, 3755–3758.

Rugg-Gunn, P.J., Cox, B.J., Lanner, F., Sharma, P., Ignatchenko, V., McDonald, A.C.H., Garner, J., Gramolini, A.O., Rossant, J., and Kislinger, T. (2012). Cell-Surface Proteomics Identifies Lineage-Specific Markers of Embryo-Derived Stem Cells. Dev. Cell *22*, 887–901.

Ruiz, N., Wu, T., Kahne, D., and Silhavy, T.J. (2006). Probing the Barrier Function of the Outer Membrane with Chemical Conditionality. ACS Chem. Biol. *1*, 385–395.

Savitski, M.M., Reinhard, F.B.M., Franken, H., Werner, T., Savitski, M.F., Eberhard, D.,

Martinez Molina, D., Jafari, R., Dovega, R.B., Klaeger, S., et al. (2014). Tracking cancer drugs in living cells by thermal profiling of the proteome. Science *346*, 1255784.

Schenk, M., Belisle, J.T., and Modlin, R.L. (2009). TLR2 Looks at Lipoproteins. Immunity *31*, 847–849.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Schulze, R.J., and Zückert, W.R. (2006). Borrelia burgdorferi lipoproteins are secreted to the outer surface by default. Mol. Microbiol. *59*, 1473–1484.

Sekar, K., Rusconi, R., Sauls, J.T., Fuhrer, T., Noor, E., Nguyen, J., Fernandez, V.I., Buffing, M.F., Berney, M., Jun, S., et al. (2018). Synthesis and degradation of FtsZ quantitatively predict the first cell division in starved bacteria. Mol. Syst. Biol. *14*, e8623.

Selkrig, J., Mosbahi, K., Webb, C.T., Belousoff, M.J., Perry, A.J., Wells, T.J., Morris, F., Leyton, D.L., Totsika, M., Phan, M.-D., et al. (2012). Discovery of an archetypal protein transport system in bacterial outer membranes. Nat. Struct. Mol. Biol. *19*, 506–510.

Semchenko, E.A., Day, C.J., and Seib, K.L. (2017). MetQ of Neisseria gonorrhoeae Is a Surface-Expressed Antigen That Elicits Bactericidal and Functional Blocking Antibodies. Infect. Immun. *85*, 1–17.

Sherman, D.J., Xie, R., Taylor, R.J., George, A.H., Okuda, S., Foster, P.J., Needleman, D.J., and Kahne, D. (2018). Lipopolysaccharide is transported to the cell surface by a membrane-to-membrane protein bridge. Science (80-.). *359*, 798–801.

Shimada, T., Makinoshima, H., Ogawa, Y., Miki, T., Maeda, M., and Ishihama, A. (2004). Classification and Strength Measurement of Stationary-Phase Promoters by Use of a Newly Developed Promoter Cloning Vector. J. Bacteriol. *186*, 7112–7122.

Sittka, A., Pfeiffer, V., Tedin, K., and Vogel, J. (2007). The RNA chaperone Hfq is essential for the virulence of Salmonella typhimurium. Mol. Microbiol. *63*, 193–217.

Sperandeo, P., Martorana, A.M., and Polissi, A. (2017). The lipopolysaccharide transport (Lpt) machinery: A nonconventional transporter for lipopolysaccharide assembly at the outer membrane of Gram-negative bacteria. J. Biol. Chem. *292*, 17981–17990.

Spratt, B.G. (1975). Distinct penicillin binding proteins involved in the division, elongation, and shape of Escherichia coli K12. Proc. Natl. Acad. Sci. *72*, 2999–3003.

Stevenson, K., McVey, A.F., Clark, I.B.N., Swain, P.S., and Pilizota, T. (2016). General calibration of microbial growth in microplate readers. Sci. Rep. *6*, 38828.

Storek, K.M., Auerbach, M.R., Shi, H., Garcia, N.K., Sun, D., Nickerson, N.N., Vij, R., Lin, Z., Chiang, N., Schneider, K., et al. (2018). Monoclonal antibody targeting the β-barrel assembly machine of Escherichia coli is bactericidal. Proc. Natl. Acad. Sci. *115*, 3692–3697.

Surendran, N., Barnard, D., Hellman, J., Michel, L.V., Bettinger, J., Pichichero, M.E., MacPherson, V., Shaw, J., and D'Arcy, B. (2015). Dual orientation of the outer membrane lipoprotein Pal in Escherichia coli. Microbiology *161*, 1251–1259.

Szewczyk, J., and Collet, J.-F. (2016). The Journey of Lipoproteins Through the Cell (Elsevier Ltd.).

Tao, K., Narita, S. -i., and Tokuda, H. (2012). Defective Lipoprotein Sorting Induces lolA Expression through the Rcs Stress Response Phosphorelay System. J. Bacteriol. *194*, 3643–3650.

Tata, M., and Konovalova, A. (2019). Improper Coordination of BamA and BamD Results in Bam Complex Jamming by a Lipoprotein Substrate. MBio *10*, 1–10.

The UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. *47*, D506–D515.

Tokuda, H., and Matsuyama, S. (2004). Sorting of lipoproteins to the outer membrane in E. coli. Biochim. Biophys. Acta *1694*, IN1-9.

Typas, A., Banzhaf, M., Gross, C.A., and Vollmer, W. (2012). From the regulation of peptidoglycan synthesis to bacterial growth and morphology. Nat. Rev. Microbiol. *10*, 123–136.

Valguarnera, E., Scott, N.E., Azimzadeh, P., and Feldman, M.F. (2018). Surface Exposure and Packing of Lipoproteins into Outer Membrane Vesicles Are Coupled Processes in Bacteroides. MSphere *3*, 1–14.

Vatanen, T., Kostic, A.D., D'Hennezel, E., Siljander, H., Franzosa, E.A., Yassour, M., Kolde, R., Vlamakis, H., Arthur, T.D., Hämäläinen, A.-M., et al. (2016). Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. Cell *165*, 842–853.

Wall, E., Majdalani, N., and Gottesman, S. (2018). The Complex Rcs Regulatory Cascade. Annu. Rev. Microbiol. *72*, 111–139.

Ward, J.E., and Lutkenhaus, J. (1985). Overproduction of FtsZ induces minicell formation in E. coli. Cell *42*, 941–949.

Webb, C.T., Selkrig, J., Perry, A.J., Noinaj, N., Buchanan, S.K., and Lithgow, T. (2012). Dynamic association of BAM complex modules includes surface exposure of the lipoprotein BamC. J. Mol. Biol. *422*, 545–555.

Werner, T., Sweetman, G., Savitski, M.F., Mathieson, T., Bantscheff, M., and Savitski, M.M. (2014). Ion Coalescence of Neutron Encoded TMT 10-Plex Reporter Ions. Anal. Chem. *86*, 3594–3601.

Wexler, A.G., Schofield, W.B., Degnan, P.H., Folta-Stogniew, E., Barry, N.A., and Goodman, A.L. (2018). Human gut Bacteroides capture vitamin B12 via cell surface-exposed lipoproteins. Elife *7*, 1–20.

Whitney, J.C., and Howell, P.L. (2013). Synthase-dependent exopolysaccharide secretion in Gram-negative bacteria. Trends Microbiol. *21*, 63–72.

Wilson, M.M., Anderson, D.E., and Bernstein, H.D. (2015). Analysis of the Outer Membrane Proteome and Secretome of Bacteroides fragilis Reveals a Multiplicity of Secretion Mechanisms. PLoS One *10*, e0117732.

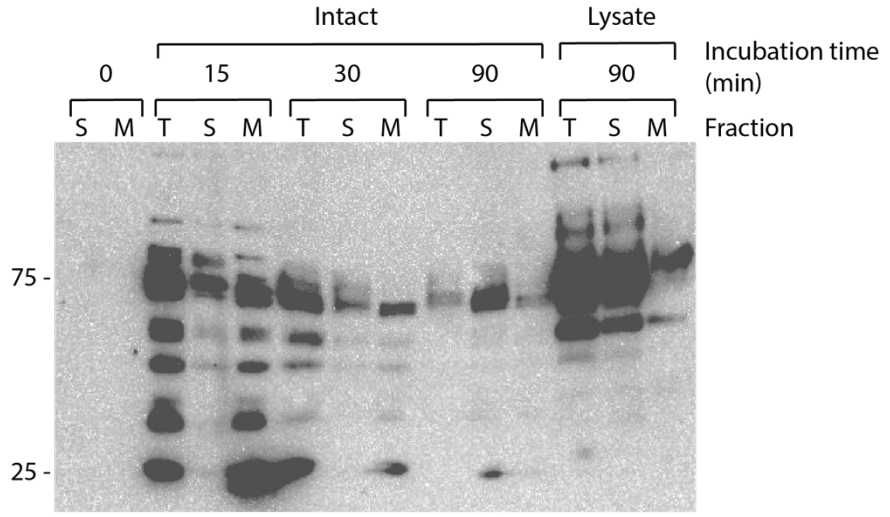
Yamaguchi, K., Yu, F., and Inouye, M. (1988). A single amino acid determinant of the membrane localization of lipoproteins in E. coli. Cell *53*, 423–432.

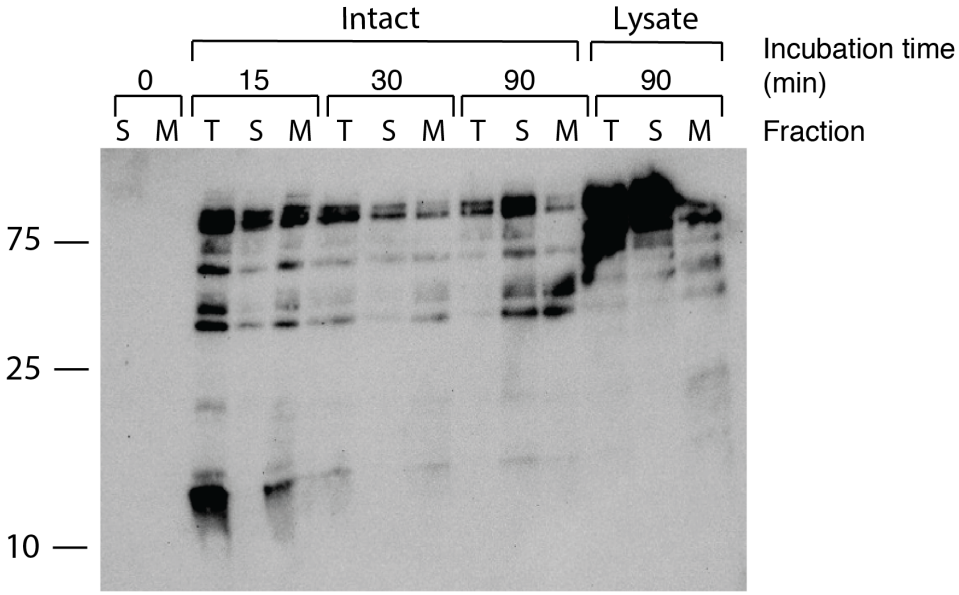
Yang, D.C., Peters, N.T., Parzych, K.R., Uehara, T., Markovski, M., and Bernhardt, T.G. (2011). An ATP-binding cassette transporter-like complex governs cell-wall hydrolysis at the bacterial cytokinetic ring. Proc. Natl. Acad. Sci. *108*, E1052–E1060.

Zückert, W.R. (2014). Secretion of Bacterial Lipoproteins: Through the Cytoplasmic Membrane, the Periplasm and Beyond. Biochim. Biophys. Acta - Mol. Cell Res. *1843*, 1509–1516.

Appendix

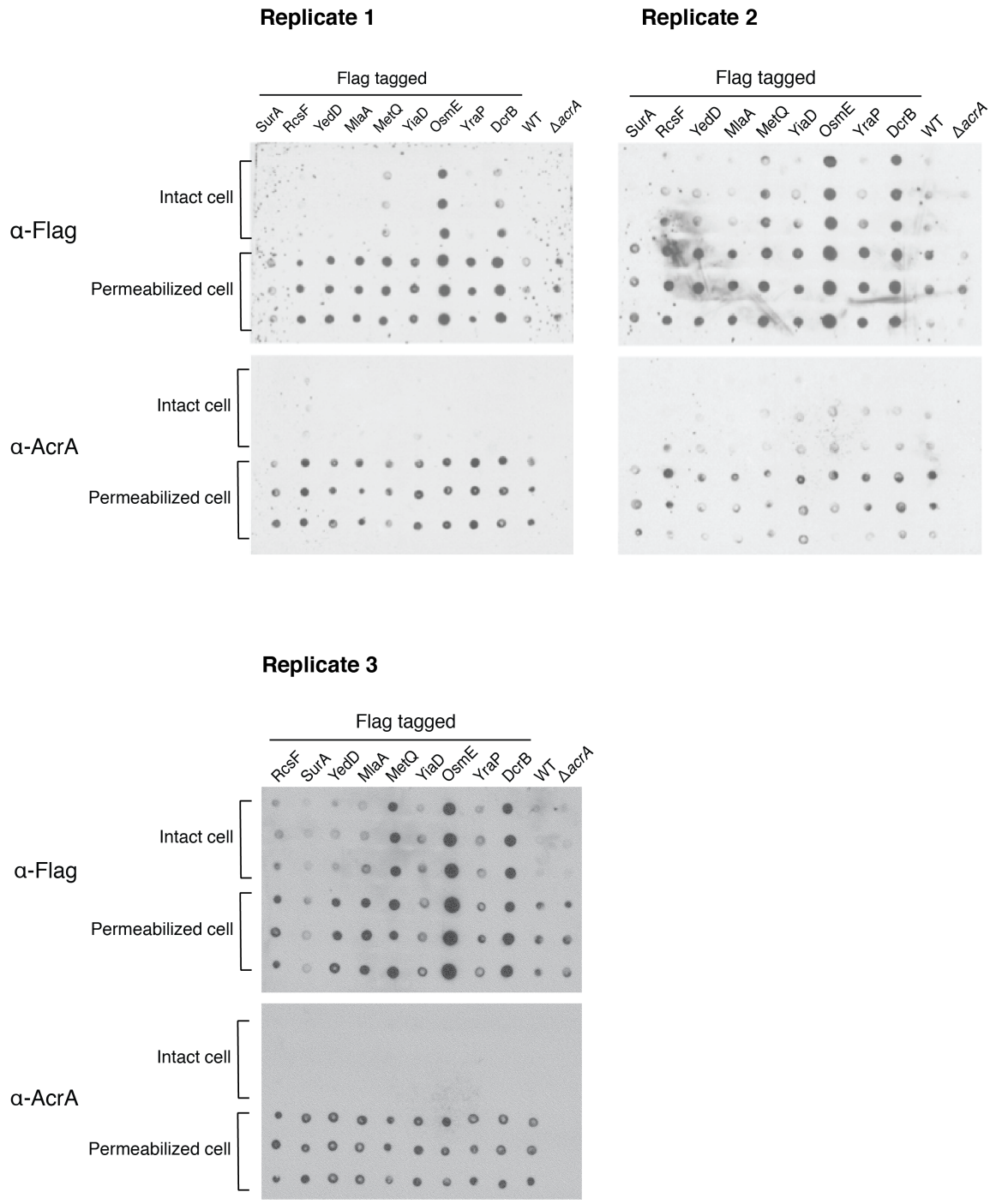
Appendix Figure 1: Original Western blots for biotinylation level quantification.
Western blots using Streptavidin-HRP (against biotin). Upper panel: 8% gel, lower panel: 15% gel. Corresponding to Chapter 3, Figure 3.1





Appendix Figure 2: Replicates of dot blot.

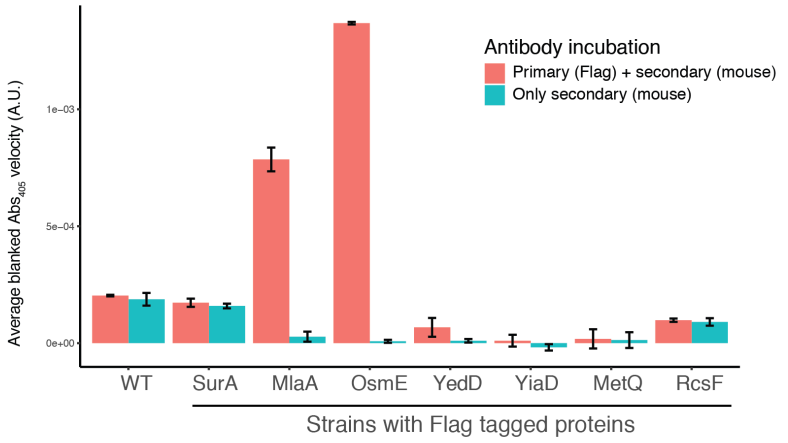
Three biological replicates of dot blot with three technical replicates each. Corresponding to Chapter 3, Figure 3.9



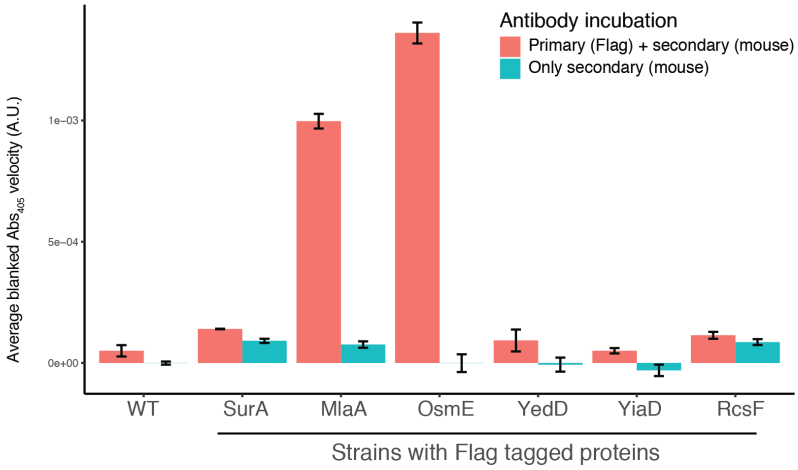
Appendix Figure 3: Replicates of whole-cell ELISA.

Two other biological replicates of ELISA for Flag tagged proteins. Corresponding to Chapter 3, Figure 3.10

Replicate 2

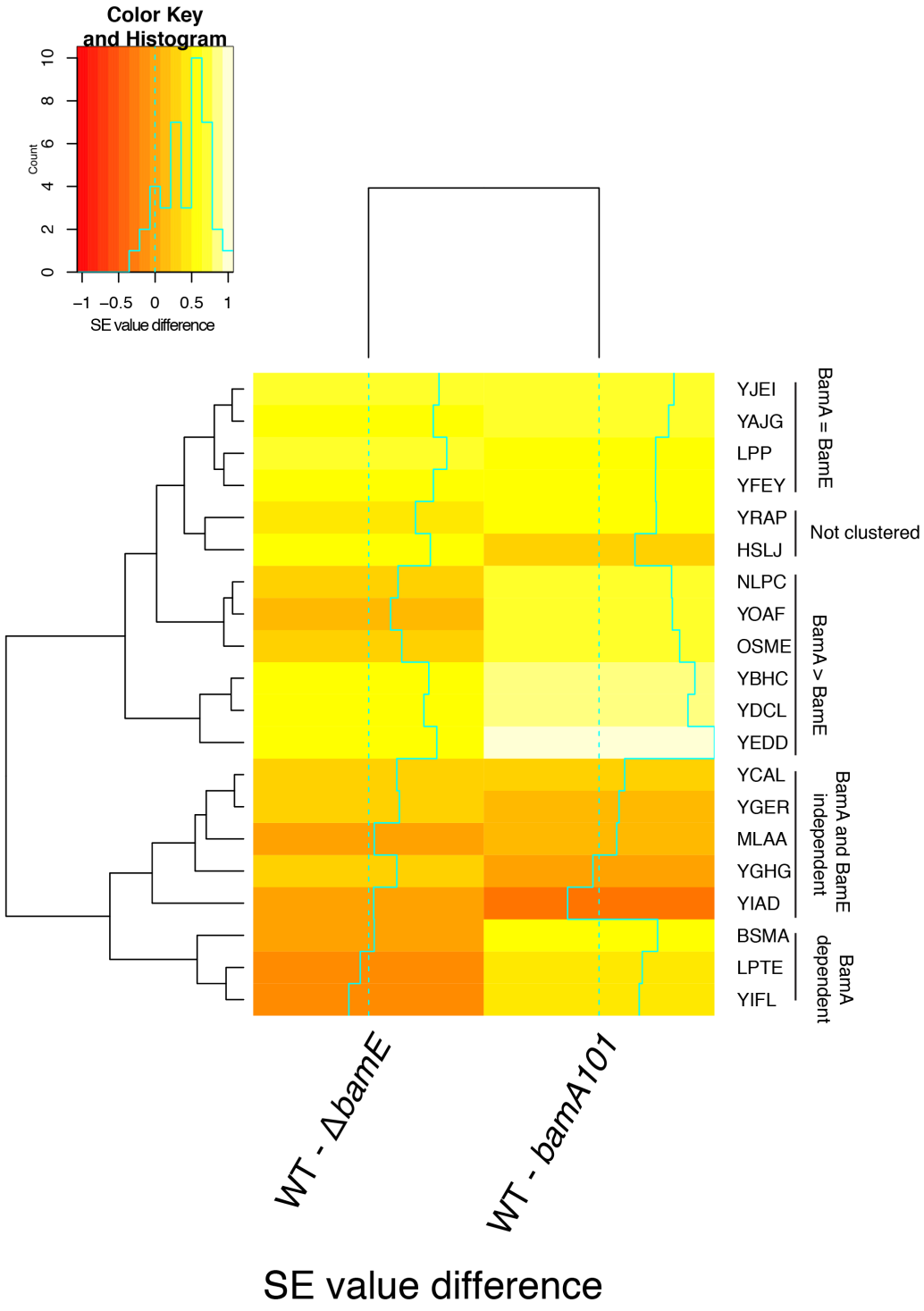


Replicate 3



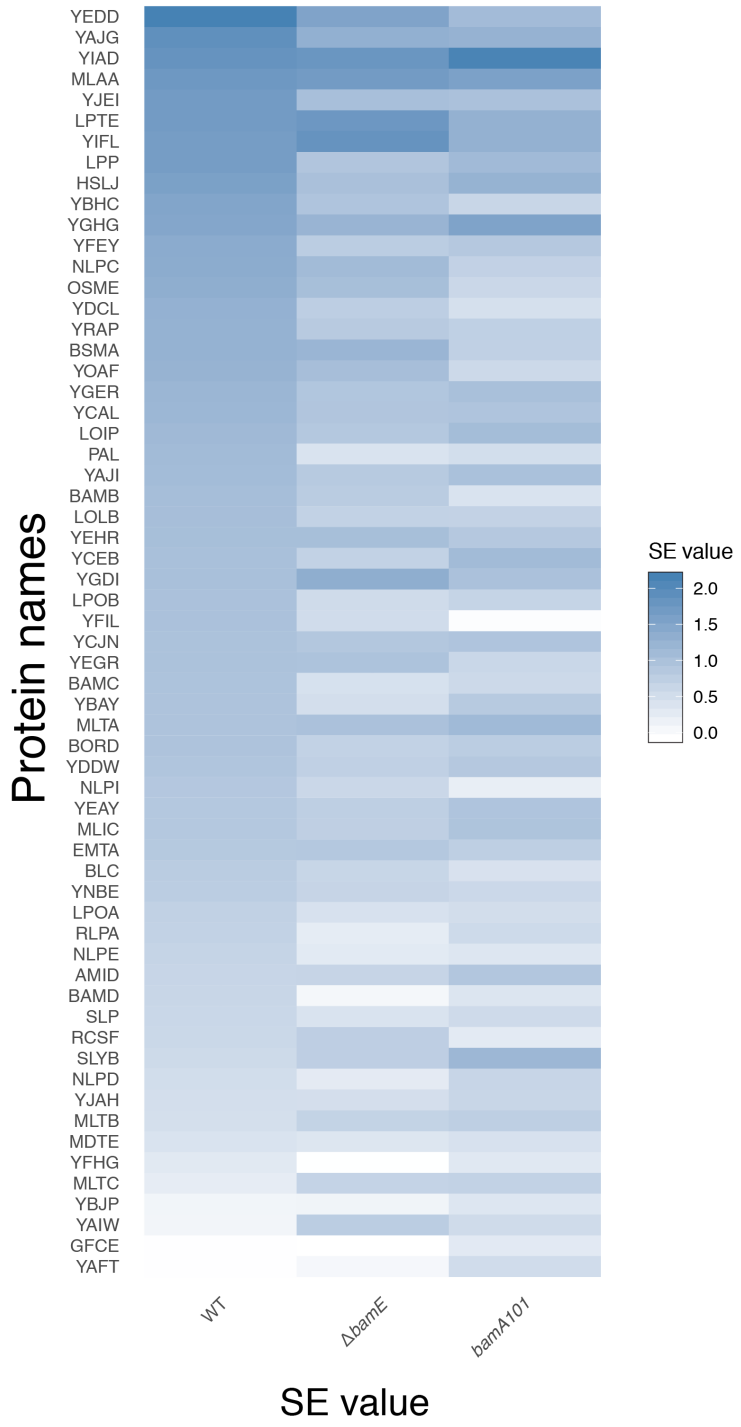
Appendix Figure 4: Hierarchical clustering of SE values.

Hierarchical clustering of SE value differences, 1:WT – *bamA101* and 2: WT – Δ*bamE*, Corresponding to Chapter 4, Figure 4.3



Appendix Figure 5: SE values of lipoproteins for three strains.

SE values of all identified lipoproteins (n = 61), for the three strains: WT, *bamA101,* Δ*bamE* shown as a heatmap (in the order of SE value in WT). Corresponding to Chapter 4, Figure 4.3



Appendix Figure 6: Replicate of FtsZ level quantification.

An independent biological replicate for FtsZ level quantification upon Rcs system activation. Western blot and quantification for FtsZ and RecA (loading control) and the ratio of FtsZ/RecA. Corresponding to Chapter 6, Figure 6.5