Evaluation of the Tat export pathway for the production of recombinant proteins in *Escherichia coli*

Gilles Malherbe

Thesis submitted to the University of Kent for the degree of Doctor of Philosophy in the Faculty of Microbiology

2019

ABSTRACT

Recently, the twin-arginine translocation (Tat) pathway has raised interest due to reports of its unique proofreading ability to export correctly folded proteins. However, only a few recombinant proteins fused to an N-terminal Tat signal peptide have been reported to be exported via Tat. The hypothesis for this thesis was to evaluate whether the range of proteins of interest that can be exported by the Tat pathway can be extended by C-terminally fusing them to a natural Tat substrate (NTS). Thus, the NTS would act as a soluble carrier to improve cytoplasmic solubility and facilitate Tat export. Moreover, the CyDisCo technology enabling cytoplasmic formation of disulphide bonds was investigated to reach this goal.

Initially, a robust fractionation method was developed to assess protein localisation with due diligence. A panel of NTS and reporter proteins were evaluated for the respective role of the soluble carrier and the protein of interest. None of the selected NTS and only PhoA and hGH of the tested reporter proteins were shown to export specifically via Tat. After the review of the fusion design to use the valid reporter proteins as the soluble carrier, three proteins of interest were expressed with and without the soluble carriers. This fusion strategy did not extend Tat acceptance. Overall, this project confirmed the difficulty to export a recombinant protein via the Tat pathway. Additionally, CyDisCo was demonstrated to improve expression of recombinant proteins but did not form native disulphide bonds.

However, this project revealed unexpected export of sfGFP, hGH and FABP4 including the recombinant proteins fused to hGH without a signal peptide. This unidentified translocation mechanism(s) needs further characterisation which can lead to the discovery of a new export pathway(s) and may highlight a promising alternative way to target proteins of interest to the periplasm.

TABLE OF CONTENT

Final word count: 54,197

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

DECLARATION

The work presented in this thesis is original work, conducted by myself under the supervision of Dr Emma Davé (UCB Celltech), Dr David P. Humphreys (UCB Celltech) and Professor Colin Robinson (University of Kent). I was based at the UCB Celltech company (Slough, United Kingdom) where all laboratory activities took place. No part of this thesis has been submitted in support of an application for any degree or other qualification of the University of Kent, or any other University or Institution of learning.

The research leading to these results has received funding from the People Programme (Marie Skłodowska-Curie Actions) of the European Union's Horizon 2020 Programme under REA grant agreement no. 642836. This material reflects only the authors' views and the Union is not liable for any use that may be made of the information contained therein.

Some of the results presented in this thesis have been published in the following journal:

MALHERBE, G., HUMPHREYS, D. P. & DAVE, E. 2019. A robust fractionation method for protein subcellular localisation studies in *Escherichia coli*. *Biotechniques*, 66, 171-178.

Gilles Malherbe

ABBREVIATIONS AND SYMBOLS

%	percent
°C	Celsius degree
λ_{em}	wavelength of emission
λ_{ex}	wavelength of excitation
ArfA	alternative ribosome rescue factor A
ATP	adenosine triphosphate
bp	base pair
cDNA	complementary deoxyribonucleic acid
CDR	complementary domain region
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleotide acid
dNTP	deoxynucleotide
DSB	disulphide bond
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
eGFP	enhanced green fluorescent protein
FABP4	fatty acid-binding protein 4
Fc	Fragment crystallizable
Ffh	Fifty-four homologue
FL	flexible linker
FRET	Förster resonance energy transfer
FW	framework
g	gravity
GFP	green fluorescent protein
h	hour
hcAb	Heavy-chain only antibody
hGH	human growth hormone
IPTG	isopropyl β-D-1 thiogalactopyranoside
kbp	kilobase pair
kDa	kilodalton
L	litre
LDLR	low-density lipoprotein receptor
LB	Luria-Bertani broth according to Lennox recipe

M	molar
MBP	maltose binding protein
MCS	multiple cloning site
mg	milligram
min	minute
ml	millilitre
mM	millimolar
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger ribonucleic acid
MscL	large mechanosensitive channel
NADPH	nicotinamide adenine dinucleotide phosphate
NEM	N-ethylmaleimide
ng	nanogram
nm	nanometre
NMR	nuclear magnetic resonance
NTS	natural Tat substrate
OD_{600}	optical density at 600 nm
ONPG	ortho-nitrophenyl-β-galactoside
PAGE	polyacrylamide gel electrophoresis
PaoA	aldehyde dehydrogenase iron-sulphur subunit
PBS	phosphate buffer saline
PCR	polymerase chain reaction
PDI	protein disulphide isomerase
PEG	polyethylene glycol
pH	potential of hydrogen
PhoA	alkaline phosphatase
PI	Post-induction
pNPP	para-nitrophenyl-phosphate
prl	specific genetic modification leading to a different protein localisation phenotype
RBS	ribosome binding site
RNA	ribonucleic acid
rpm	rotation per minute
scFv	single-chain variable fragment
SDS	sodium dodecyl-sulphate
sec	second

Sec	general secretory pathway
sfGFP	super-folder green fluorescent protein
SOC	super optimal broth
$_{SP}$	signal peptide
SRP	signal recognition particle
SUMO	small ubiquitin-like modifier
T4	bacteriophage T4
TAE	tris acetate EDTA
Tat	twin-arginine translocation
Tat_{SP}	Signal peptide leading for the twin-arginine translocation pathway
TF	trigger factor
T_M	melting temperature
TMAO	Trimethylamine-N-oxide
TorA	Trimethylamine-N-oxide reductase
TPM	transcript per million
Tris	trisaminomethane
Tris-Cl	trisaminomethane chloride
TY	LB medium with twice the amount of yeast extract
U	unit
µg	microgram
µl	microliter
µM	micromolar
US	United States
UV	ultraviolet
V	volt
VHH	antigen binding fragment of heavy chain only antibodies
WT	wild-type
w/v	weight per volume
X_{SP}	signal peptide of the protein X

ACKNOWLEDGEMENTS

I dedicate my PhD to my late grandparents who genuinely inspired me in life and continuously encouraged me during the course of this degree but will, unfortunately, not have the opportunity to see me bring it to completion.

Je dédicace ma thèse à mes défunts grands-parents qui n'auront malheureusement pas la possibilité de voir la fin de ce projet mais qui m'ont profondément inspiré et continuellement soutenu.

Firstly, I must express my profound gratitude to my research supervisor Dr Emma Davé whose support, guidance and encouragement have played a major role throughout this PhD project. Her patience, helpfulness and knowledge have genuinely made this experience more enjoyable. I am particularly grateful to Dr David Humphreys, at UCB, who granted me the opportunity to undertake my PhD, as well as Prof Colin Robinson from the University of Kent. They have been both helpful in terms of guidance, advice and feedback. I would also like to thank Prof Jan Maarten van Dijl and Dr Sierd Bron from the University of Groningen who created and coordinated the ProteinFactory consortia enabling me to have a rich networking experience during my PhD.

I would like to thank UCB which has been a great environment for professional and personal development. Special thankyou goes to Dr Neha Dhami for her help with real-time PCR experiments, Mark Ellis with bacterial genome mutagenesis and sequencing, Dr Remi Okoye with microscopy, Dr Linghong Huang with immunofluorescence, Rebecca Munro with raising antibodies and lastly Dr Hannah Edwards with RNA-Sequencing. I am also grateful to all my colleagues who have created such a friendly, yet professional environment.

Je voudrais remercier particulièrement ma famille pour leurs soutiens et encouragements tout au long de cette expérience. Ils ont su trouver les mots pour me motiver et me pousser pendant ces années. Je remercie très particulièrement mes parents pour m'avoir donné l'opportunité de relever ce défi.

I would like to thank Thomas Bergersen, Hans Zimmer and Lindsey Stirling whose work have truly inspired me and kept me going through the long days of the few last years.

I would also like to express my gratitude towards the British government and the Brexit committee for facilitating this advanced level training but not encouraging me, or the skills that I developed here, to stay in the country after graduation.

Last, but not least, I would like to give a huge thankyou to my friends at and beyond UCB for their help, support and advice, making this experience stimulating. A special thankyou goes to Guillaume Valette, Anna Mastela, Jérémie Monribot and Maria Tsoumpeli who have been there to support me psychologically throughout this PhD.

1. INTRODUCTION

1.1. General introduction

Proteins are required in academic research, the pharmaceutical, food and textile industries. They have multiple usage such as therapeutics, research proteins, cooking enzymes and proteases in laundry detergents. Proteins used as biopharmaceutical medicines encompass antibodies as well as non-antibody molecules and in 2016, represented US$ 228 billion in global sales (Moorkens et al., 2017). These proteins are produced as recombinant proteins for development and manufacture at laboratory and large scales for both academic and industrial applications. They can be produced in eukaryotic (mammalian cells, insect cells, yeast, plant) or prokaryotic (bacteria) systems (Gasser et al., 2013, Butler and Spearman, 2014, Stolt-Bergner et al., 2018, Sack et al., 2015, Chen, 2012). The most popular hosts are mammalian cells and bacteria. The latter were historically the first developed hosts for recombinant protein production thanks to their quick generation time, easy manipulation, low cost equipment and high productivity (Terpe, 2006). However, in many occurrences, bacteria are not able to produce active heterologous proteins as some tend to form instable polypeptides which aggregate and end up as inclusion bodies due to the lack of access of different machineries (Golgi apparatus, endoplasmic reticulum) or necessary chaperones (García-Fruitós, 2010). Many engineering works have emerged in the past decades to contour these limitations to render bacteria even more attractive for recombinant protein production (section 1.4.5). On the other hand, mammalian cells are the systems of choice for glycosylated therapeutic protein production. The main drawbacks of these hosts are the high cost brought by the slow cell growth, expensive media and specialized equipment needed for culture conditions.

Amongst bacterial systems, both Gram-positive and Gram-negative strains are used and therefore offer different advantages made possible by way of their cell wall organization. The cell wall of Gram-positive bacteria consists of a thin periplasmic space between a thick peptidoglycan layer and the cell membrane as represented in Figure 1-1 (Matias and Beveridge, 2005). In Gram negative bacteria, a periplasmic space containing a thin peptidoglycan layer is delimited by the inner and outer cell

membranes (Figure 1-1). In terms of recombinant protein production, this difference is translated by the ease of protein excretion in Gram-positive compared to Gram-negative bacteria. Indeed, species like *Bacillus subtilis* are often used for production of proteins intended for secretion into the extra-cellular medium (Song et al., 2015). Other bacterial expression systems have emerged like the Gram-positive *Lactococcus lactis* and the Gram-negative *Pseudomonas* (Chen, 2012). For instance, proteins production in *Pseudomonas fluorescens* is less dependent on oxygen concentration, does not accumulate acetate and gave a yield greater than 30 times higher compared to production in *E. coli* (Huang et al., 2007). This species can also be used for periplasmic export, but its pathogenicity refrains the development as a production platform especially for therapeutics (Retallack et al., 2007, Scales et al., 2014). Additionally, *Lactococcus lactis* is largely used for the production of membrane proteins due its single membrane cell envelope but the protein production titers have never been reported higher than 300 mg/L (Chen, 2012).

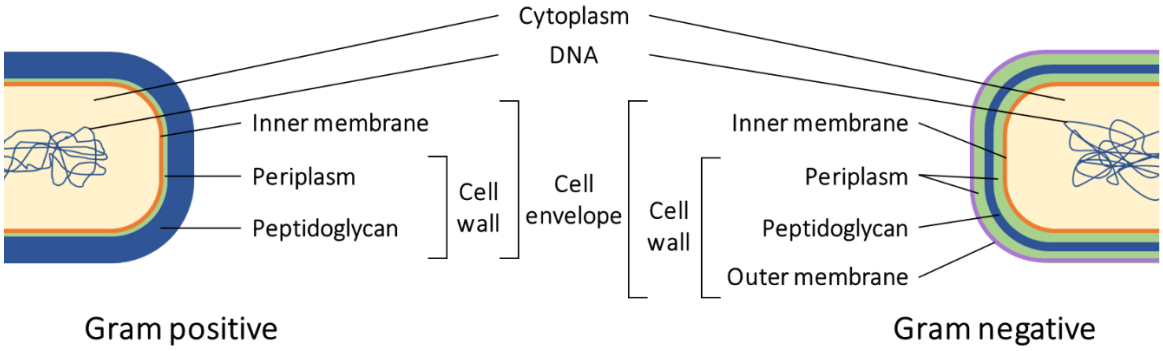


Figure 1-1: Diagram of bacterial envelopes

Somatostatin was, in 1977, the first human recombinant protein to be generated (Itakura et al., 1977) and in 1982, human insulin was the first commercial recombinant drug to receive approval from the US food and drug administration (Johnson, 1983). Both proteins were produced in *E. coli* and since then, this species has been used extensively for recombinant protein production platform. Indeed in 2009, a report indicated that about 30 % of protein-based pharmaceuticals approved by

competent authorities were produced in *E. coli* (Ferrer-Miralles et al., 2009). The various therapeutics produced by this bacterium include interferons, insulins and monoclonal antibodies (Jozala et al., 2016, Baeshen et al., 2015, Sanchez-Garcia et al., 2016). *E. coli* offers several additional advantages including fast growth kinetics, high cell density and the availability of non-pathogenic species (Rosano and Ceccarelli, 2014).

1.2. The *Escherichia coli* K12 strain as a production host

As mentioned in section 1.1, *E. coli* is used as the first-choice microorganism for recombinant protein production but many strains from this species are pathogenic to humans (Croxen et al., 2013). However, the descendants of a non-pathogenic isolate, *E. coli* K12 (Tatum and Lederberg, 1947) have given rise to strains such as MG1655 and W3110 which are routinely used for recombinant protein expression (Daegelen et al., 2009). K12 derivatives do not possess the F plasmid necessary for conjugation and are therefore designated F^- and are defective for lambda lysogeny (λ). Many examples of successful protein expression in *E. coli* K12 strains can be found for expression of a research protein at laboratory scale such as interleukin 3 (Hercus et al., 2013), amylases (Zafar et al., 2016), myosines (Pylypenko et al., 2016). Also, such strains are used to produce therapeutics at industrial scale (Huang et al., 2012) including the enzyme Glucarpidase (Voraxaze®) used to treat delayed methotrexate clearance in patients with renal failure (Rattu et al., 2013), the pegylated Fab' fragment Certolizumab pegol (Cimzia®) for the treatment of rheumatoid arthritis in combination with methotrexate (Goel and Stephens, 2010), the human growth hormone (hGH) produced in the periplasm (de Oliveira et al., 1999) and the human stem cell factor in fusion with erythropoietin (Su et al., 2006). For the present study, the *E. coli* K12 W3110 strain was selected has the host since it is used routinely in-house for periplasmic expression of antibody fragments (section 3.6.).

Furthermore, a plethora of parameters need to be considered in order to optimize the recombinant protein production process in the selected host strain. Examples of such parameters are listed in Table 1-1. Moreover, the cytoplasm contains more host cell proteins, including proteases, than the

periplasm or the extra-cellular medium. Hence, the localisation of the protein of interest within the cell impacts the purification process and product quality. Periplasmic expression was previously associated with lower production yields than cytoplasmic expression but recent studies have demonstrated than protein can be expressed to up to 1.1 g/L or even 2.4 g/L in the periplasm of *E. coli* (Matos et al., 2012, Ellis et al., 2017).

Table 1-1: Parameters to consider for recombinant protein expression optimization

Parameters to optimize		**Examples**	**Reference**
Expression vector	Selection marker	antibiotic resistance, plasmid addiction system	Rosano and Ceccarelli (2014), Goh and Good (2008)
	Copy number	Low (SC101), moderate (p15A), high (pUC) copy numbers	Friehs (2004)
	Induction system	IPTG inducible (*tac*), arabinose inducible (pRHaBAD) promoters	de Boer et al. (1983), Szeliova et al. (2016)
Expression cassette	Transcription and translation	RBS and promoter strengths	Mutalik et al. (2013), Li and Zhang (2014)
	Codon usage		Plotkin and Kudla (2011)
	Gene organization	Single, mutlicistronic	Tan (2001)
	Detection/purific-ation/solubility tag	His, FLAG, V5, HA, SUMO, MBP, GST	Kosobokova et al. (2016)
	Cleavage site	TEV, thrombin cleavage sites	Kosobokova et al. (2016)
	Signal peptide	OmpA, MBP, PelB, TorA, HlyC signal peptides	Freudl (2018)
	Chaperone(s) co-expression	GroEL, GroES, Trigger factor	Gräslund et al. (2008)
Culture condition	Mode of culture	Batch, fed-batch, continuous fermentation	Costa et al. (2014)
	Container	Deep well plate, shake flask, bioreactor	Overton (2014)
	Medium	Rich or minimum, defined or undefined	Arya et al. (2015)
	Temperature	Optimal, suboptimal	Costa et al. (2014)
	Agitation		Overton (2014)
	Induction level		Overton (2014)
	Harvest time		Arya et al. (2015)

1.3. Secretion systems in *Escherichia coli*

E. coli possesses several ways to transport proteins from the cytoplasm into the periplasm (export) or to the extra-cellular medium (secretion). Gram-negative bacteria possess eight known pathways to transport proteins across the biological membranes (Green and Mecsas, 2016). They are listed as the general secretory (Sec), twin-arginine translocation (Tat), TypeI, TypeII, TypeIII, TypeIV, TypeV and TypeVI pathways. The Sec and Tat pathways have been exploited for recombinant protein production whereas the remaining systems are involved in bacterial virulence (Green and Mecsas, 2016).

1.3.1. The Sec pathway

The Sec pathway is an essential pathway used by around 96 % of *E. coli* exportome of which 60 % are inner membrane proteins and the remaining 40 % are outer membrane, periplasmic and secreted proteins (Yu et al., 2010). It allows substrate proteins to translocate into or across the plasma membrane as unfolded polypeptide chains using two mechanisms: the SecB and the signal recognition particle (SRP) pathways (Pugsley, 1993). *De novo* translated polypeptides are exported by Sec in three stages: firstly, they are targeted to the Sec machinery via a N-terminal signal peptide and sorted to the SecB or SRP pathways. Secondly, they are translocated across or into the inner membrane and finally, the signal peptide is cleaved by membrane anchored peptidase prior to release of the preprotein.

1.3.1.1. Sec-specific signal peptides

Sec substrates are targeted for Sec translocation via specific signal sequences borne at their N-termini (Green and Mecsas, 2016). SRP targets highly hydrophobic signal peptides to the SecYEG channel in a co-translational way whereas SecB recognizes a less hydrophobic signal peptide. The signal peptides are typically 18-27 residues long and can be divided into three regions as described in Figure 1-2 (Gouridis et al., 2009). The n-region is located at the amino terminal and is characterised by a net positive charge. One or more basic residues are responsible for the positive charge(s) of the region allowing insertion of the signal peptide into the SecYEG translocon. The net charge varies from

one substrate to the other and can influence protein synthesis and translocation rates (Caspers et al., 2010). Moreover, the net charge should be maintained to a positive value, otherwise the translocation rate can suffer significant reduction (Inouye et al., 1982). For instance, a cutinase from the fungus *Fusarium solan pisi* was expressed in *Bacillus subtilis* fused to the signal peptide of AmyE (Caspers et al., 2010). The authors found that Phe2Asp, Phe2Glu or Lys4Leu mutation, each lowering the net charge from +3 to +2, led to a three-fold increase in terms of secretion and activity. Screening the net charge of signal peptides for a selected recombinant protein can lead to improved secretion yields but is protein specific (Puziss et al., 1992, Nesmeyanova et al., 1997, Ismail et al., 2011).

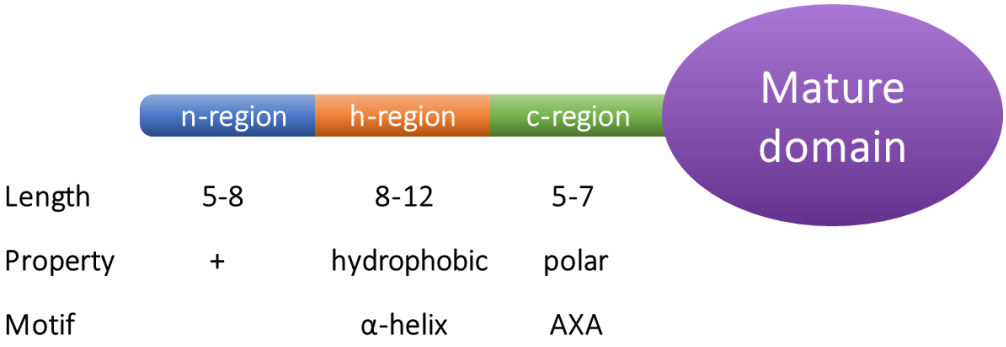


Figure 1-2: General features of a Sec-specific signal peptide

The Sec signal peptide is composed of a positively charged amino-terminal region (n-region), followed by a hydrophobic region (h-region) which forms an α-helix upon binding to SecA, and is terminated by a carboxyl-terminal region (c-region) containing a signal peptidase cleavage site (AXA). The lengths are indicated in amino acid residues. + means positively charges.

The following region is the h-region of 8 to 12 amino acid residues characterised by its hydrophobicity and its helical propensity (Figure 1-2). Decreasing the hydrophobicity, by deletion or mutation of amino acid residues, considerably decreases protein export resulting in cytoplasmic accumulation in a misfolded state (Izard and Kendall, 1994, Suominen et al., 1995). Although, the possible deflection from the SRP to the SecB pathways was not investigated. The low hydrophobicity can exceed a limit, resulting in a significant decrease in translocation of native Sec substrates. For instance, a minimum of five leucine residues on the signal peptide of alkaline phosphatase (PhoA) is required for effective

export (Rusch et al., 1994). On the other hand, mutating the h-region to increase the hydrophobicity by the addition of leucine(s) has proven to increase secretion efficiency (Chen et al., 1996) but again, the re-routing of nascent polypeptide from SecB to SRP was not assessed. Moreover, introducing a helix breaker reduces the affinity of SRP for the highly hydrophobic signal peptides. Similarly, an extension to this helix by mutating glycine for cysteine or leucine residues shifted the affinity of the PhoE signal peptide towards SRP (Adams et al., 2002).

The C-terminal region of a Sec signal peptide is named the c-region and contains a cleavage site to allow release of the mature domain into the periplasm or into the membrane for integral substrates (Figure 1-2). This region should be at least five residues long to ensure export and cleavage efficiency (Suominen et al., 1995). During export, the signal peptide becomes embedded into the SecYEG translocon and upon complete translocation of the nascent polypeptide into the periplasm, it is cleaved by the signal peptidase I, LepB which is located on the outer leaf of the inner membrane (Date, 1983). This enzyme recognizes a 3 amino acid motif where the -1 and -3 positions before the cleavage site favour residues with small neutral chains (von Heijne, 1984). Since the common residue at these positions is alanine, the motif is often referred to as AXA (Karamyshev et al., 1998). Protein export levels can be optimized by mutating the c-region since it is strongly influenced by signal peptide cleavage efficiency (Low et al., 2013). Indeed, a relatively large residue (i.e. phenylalanine, tyrosine, leucine and histidine) at position -2, an alanine at position -6 and hydrophobic residues (i.e. phenylalanine) at position -7 to -10 have been shown for more effective signal peptide cleavage. LepB is an essential enzyme in *E. coli*, hence no study determining the faith of Sec substrates after translocation was conducted (Mora et al., 2015). To conclude, no single known signal peptide will export all recombinant proteins into the periplasm via Sec. Therefore, a screen and optimization of signal peptides is necessary to obtain optimum yields of a specific recombinant protein, as depicted in this review (Freudl, 2018).

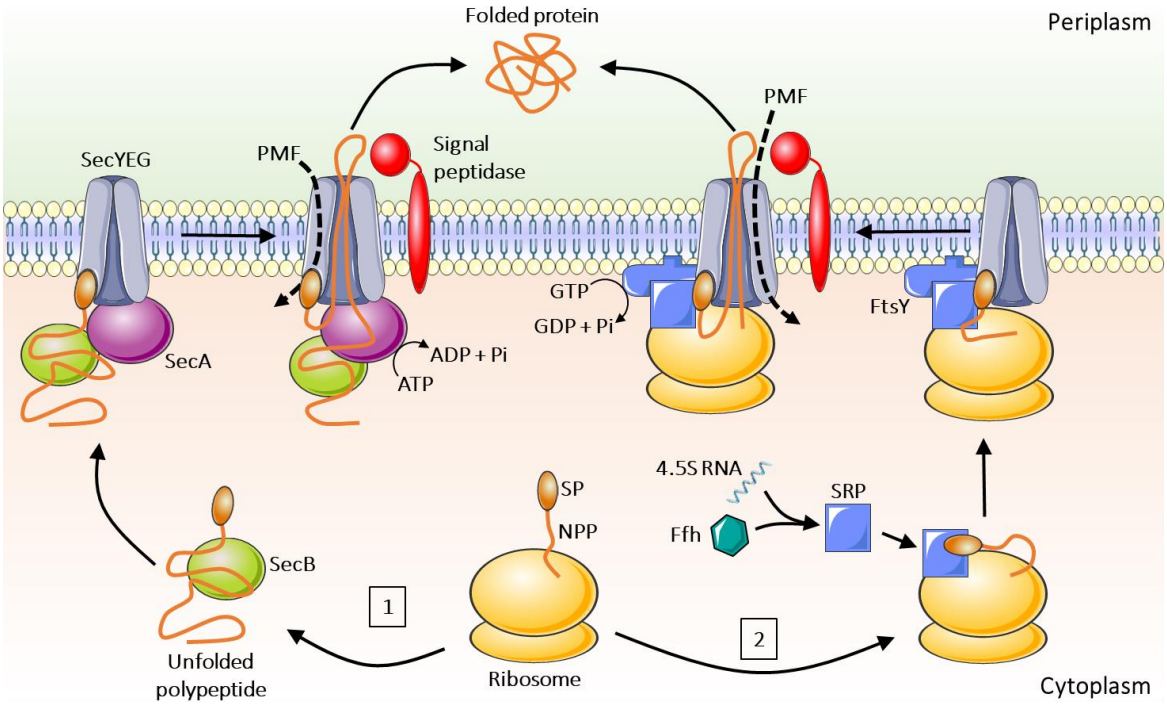


Figure 1-3: Bacterial Sec pathway

Polypeptides exported to the periplasm by the Sec pathway can be routed either via the post-translational SecB (1) or the co-translational SRP (2) systems. 1 - After binding the signal peptide (SP), SecB prevents folding of the native polypeptide (NPP) and recruits SecA which targets the complex to the SecYEG translocase. Using ATP hydrolysis and the proton motive force (PMF), the polypeptide is transported across the cytoplasmic membrane, where the signal peptidase cleaves the SP before the release and folding of the mature protein. 2 - The SRP complex is made of the Ffh protein and the small 4.5S RNA which binds to the hydrophobic domains of the SP to prevent aggregation. This ribo-complex is targeted to the SecYEG translocon via the docking protein FtsY. The polypeptide is transported during translation using GTP hydrolysis and the PMF. The mature protein can be either integrated into the membrane or released into the periplasm. Figure adapted from Green and Mecsas (2016) and Freudl (2018).

1.3.1.2. The SecB pathway

SecB substrates, targeted to the periplasm or to the extra-cellular medium, are transported through the inner membrane via the SecB pathway in a post-translational manner (Tsirigotaki et al., 2017). SecB is a non-essential protein which, when knocked-out, leaves the cell with a strong cold-sensitive phenotype below 23°C, as well as sensitivity to many compounds including ethanol, copper and triclosan (Sala et al., 2014). The SecB protein is a cytoplasmic homotetrameric chaperone which binds to either completely or partially synthesized nascent polypeptides via the interaction with a specific signal peptide (Green and Mecsas, 2016). This interaction prevents the polypeptide from folding and the complex then binds to SecA; an essential cytoplasmic translocation motor protein

(Figure 1-3). The latter directs this complex, composed of the substrate and SecB, to the SecYEG membrane channel. Translocation is initiated by SecA which pulls the polypeptide into the SecYEG channel using ATP hydrolysis as energy source. In the late stage of the polypeptide translocation, SecA detaches from the signal peptide and is released into the cytoplasm while the proton motive force pulls the polypeptide completely through the channel (Tsirigotaki et al., 2017). The movement of protons and preproteins is enabled by the accessory factor SecDF, a periplasmic complex embedded in the inner membrane (Tsukazaki et al., 2011). The signal peptide is then cleaved by the membrane-anchored signal peptidase II (LspA) for lipoproteins or by the signal peptidase I (LepB) for other substrates which allows the release of the polypeptide into the periplasm (Yu et al., 1984). Once in the periplasm, the mature protein folds and can be exported further to its destination via either the Type II or the Type V secretion systems (section 1.3.3).

1.3.1.3. The SRP pathway

The SRP pathway is utilized by polypeptides whose unfolded status cannot be maintained by SecB or possess hydrophobic patches (corresponding to future transmembrane domains) which cannot be processed by the SecB pathway. Such polypeptides are exported into the periplasm or inserted into the membrane via the SRP pathway in a co-translational manner (Green and Mecsas, 2016). The SRP is a ribonucleoprotein complex composed of the fifty-four homologue (Ffh) protein and a small 4.5S RNA (Tsirigotaki et al., 2017). SRP binds to ribosomes but is only kinetically stabilised by nascent polypeptide through the recognition of a specific signal peptide or transmembrane domains after which, it temporarily halts translation at the ribosome (Tsirigotaki et al., 2017). The SRP-polypeptide complex, also known as the ribo-complex, then binds to the, inner membrane-located, docking protein FtsY which transfers the complex to the SecYEG channel (Figure 1-3). This event obtains the necessary energy from the hydrolysis of GTP by the GTPase domains of Ffh and FtsY which triggers the release of SRP and reinitiates the translation. The polypeptide is synthetized directly into the SecYEG channel which uses the proton motive force to pull the nascent chain through. For membrane protein insertion, YidC drives the nascent transmembrane helixes integration into the

inner membrane while the chaperone PpiD is considered to run a quality control on transmembrane domains exiting the SecYEG channel (Sachelaru et al., 2014). Finally, the signal peptide is cleaved by LepB during the later stages of the membrane protein insertion or periplasmic release (Auclair et al., 2012).

1.3.1.4. Recombinant protein export via Sec

Since its discovery, the Sec pathway has been used in expression systems to export recombinant proteins into the periplasm of *E. coli*. By fusing a Sec signal peptide to the N-terminus of the protein of interest, heterologous proteins can be translocated either by the SecB or the SRP pathway depending on the signal peptide selected. For example, active human epidermal growth factor, parathyroid hormone and the T cell receptor glycoprotein were successfully produced in the periplasm via SecB export using the native signal peptide of the *E. coli* outer membrane protein, OmpA ($OmpA_{SP}$) (Sutherlan and Wong, 1993). The ice-structuring protein A and the chitonase enzyme from *Bacillus subtilis* were successfully expressed in the periplasm of *E. coli* using $OmpA_{SP}$ to 16 and 18.5 mg/L in shake flask respectively (Tong et al., 2000, Pechsrichuang et al., 2016). Sec offered a viable alternative for these proteins for which cytoplasmic expression led to lower yields (Pechsrichuang et al., 2013, Rivas et al., 2000). Therapeutic proteins such as Fab' fragments can also be expressed in the periplasm via Sec. Also, using an $OmpA_{SP}$, a Fab' fragment was expressed to 2.4 g/L in batch-fed fermentation (Ellis et al., 2017). It is important to note here, that this high yield was obtained using an engineered strain lacking the protease Tsp and co-expressing the chaperone DsbC (section 1.3.1.4. for more details).

1.3.2. The twin arginine translocation pathway

The twin arginine translocation (Tat) pathway is an alternative to Sec for proteins to cross the plasma membrane in bacteria. The major difference is that Sec exports unfolded polypeptide chains into the periplasm whereas Tat transports only folded proteins (Patel et al., 2014). Sec is used for translocation of about 98 % of *E. coli* exportome (Orfanoudaki and Economou, 2014). Of the remaining 2 %, only about 30 substrates are known for the Tat pathway (Tullman-Ercek et al., 2007).

The proteins exported by the Tat translocon are involved in respiratory metabolism, virulence, cell division, phosphate, copper and iron metabolism and cell wall formation (Palmer and Berks, 2012). The utility of the Tat machinery instead of Sec can be physiologically rationalised as a prerequisite for: (i) the insertion of complex co-factors, (ii) the prevention of periplasmic metal ions competing for insertion into the active site, (iii) the transport of hetero-oligomeric complexes which require cytoplasmic assembly, (iv) the instability of some unfolded proteins in the periplasm and (v) the requirement of cytoplasmic protein for folding and maturation (Palmer and Berks, 2012).

1.3.2.1. The role of Tat through its substrates

The total amount of substrates exported via Tat is estimated to be around 30 proteins for enteric bacteria but the number fluctuates between bacterial species (Palmer and Berks, 2012). The reason behind the absence of an absolute value is that many substrates are putative, determined from bioinformatic predictions and have not been experimentally confirmed. For instance, Tullman-Ercek et al. (2007) listed 27 potential Tat substrates by testing Tat export of the predicted Tat signal peptides when fused to the maltose binding protein (Appendix 1).

TorA is a trimethylamine N-oxide reductase from the molybdoenzyme family and is encoded by the *torCAD* operon (Mejean et al., 1994). The connection between the double arginine motif within its signal peptide and its periplasmic localisation was made and TorA was one of the first Tat substrates to be discovered (Weiner et al., 1998, Santini et al., 1998). Indeed, the protein is involved in anaerobic respiration which explains its periplasmic localisation. TorA requires a molybdenum cofactor for activity, therefore justifying the need for Tat export. Moreover, TorA requires the chaperone TorD for cytoplasmic maturation, cofactor binding and addressing the correctly folded proteins to the Tat system (Pommier et al., 1998, Ilbert et al., 2003, Ilbert et al., 2004, Oresnik et al., 2001, Jack et al., 2004).

DmsA also belongs to the molybdoenzyme family and is a subunit of the dimethyl sulfoxide (DMSO) reductase which reduces S- and N-oxides (Yokoyama and Leimkuhler, 2015). The protein binds to a molybdenum cofactor which is inserted with the help of the cytoplasmic chaperone DmsD (Chan et

al., 2008, Ray et al., 2003). The DMSO reductase is a complex containing the DmsA, DmsB and DmsC subunits. The DmsC is an integral transmembrane protein whereas DmsA and DmsB are soluble periplasmic proteins. The correctly folded DmsAB complex is co-transported via the Tat pathway thanks to the specific signal peptide of DmsA.

NapA, FdnG and FdoG are subunits of other molybdoenzymes exported via Tat to the periplasm. They are first matured in the cytoplasm by their respective NapD, FdhD and FdhD chaperones and associates in complex with their respective NapB, FdnH and FdoH partners prior to export (Yokoyama and Leimkuhler, 2015). NapAB is a nitrate reductase while FdnGH and FdoGH complexes associates with their respective FdnI and FdoI subunits in the periplasm to form the active FdnGHI and FdoGHI formate dehydrogenase complexes (Jormakka et al., 2002, Yokoyama and Leimkuhler, 2015).

In a different tone, FhuD and SufI are examples of monomeric Tat substrates which do not contain a complex cofactor (Ize et al., 2004, Samaluru et al., 2007). They bind, however, to an iron and copper metal ion respectively. FhuD is involved in the Fe(III) import system and SufI is involved in cell division under stress conditions (Koster, 1991, Samaluru et al., 2007).

1.3.2.2. Tat-specific signal peptides

In the same way as Sec, Tat substrates bear a specific N-terminal signal peptide to drive export (Patel et al., 2014). Some examples are detailed in the Appendix 2. They share the tripartite structure with respect to the n-, h- and c-region including the AXA motif allowing release of the mature domain upon translocation (Figure 1-4). However, several differences between these signal peptides allow selective targeting of the protein to its rightful pathway. Tat signal peptides are in average 14 residues longer with an average length of 38 compared to 24 amino acids for Sec signal peptides (Cristobal et al., 1999). But the main features which prevent Tat substrates from being mistargeted to Sec, are the lower hydrophobicity of the h-region, a conserved motif and the basic residues of the c-region (Berks et al., 2000, Palmer and Berks, 2012). Considerably increasing the hydrophobicity of the h-region was found to lead an artificial Tat substrate away from the Tat translocon towards Sec export and up to complete Tat-independence (Cristobal et al., 1999). The conserved S/T-R-R-X-F-L-K

motif is located at the junction of the n- and the c-region and contains a characteristic twin-arginine giving the translocase its name (Berks, 1996). Mutating both arginine residues to lysine on the natural Tat signal peptide of SufI completely abolished Tat export (Yahr and Wickner, 2001). But the mutation of one of its arginine to lysine resulted in reduced Tat export efficiency (Stanley et al., 2000). Also, mutation of the first arginine to glutamate was reported to abolish the export of β-lactamase bearing the Tat signal peptide of *Desulfovibrio vuljpis* hydrogenase while mutation to a range of other residues (valine, isoleucine, lysine and glutamine) only reduced Tat secretion yield (Niviere et al., 1992). The last Sec-repelling feature is the presence of basic residues in the c-region. Indeed, mutating these residues to the polar asparagine or glutamine, allowed Sec export (Cristobal et al., 1999, Berks et al., 2000, DeLisa et al., 2002). Despite the Sec-avoidance features, 16 of the 27 signal peptides identified by Tullman-Ercek et al. (2007) including DmsA, FhuD and SufI have been found to be promiscuous when expressed in fusion with MBP.

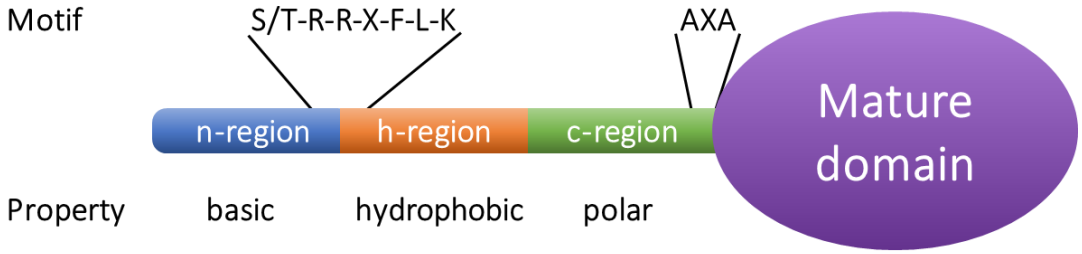


Figure 1-4: General features of a Tat-specific signal peptide

The Tat signal peptide is composed of the basic amino-terminal region (n-region), followed by the hydrophobic region (h-region) and terminated by the carboxyl-terminal region (c-region) containing a signal peptidase cleavage site (AXA). The consensus motif S/T-R-R-X-F-L-K is located at the junction of the n- and the h-regions.

1.3.2.3. Components of Tat

Five proteins are reported to be involved in the Tat pathway in *E. coli*: the four integral membrane proteins TatA, TatB, TatC and TatE and the cytoplasmic protein TatD (Figure 1-5A). TatA is a small 89 amino acid long (9.7 kDa) protein consisting of a short periplasmic N-terminus followed by a transmembrane domain, a cytoplasmic amphipathic helix and an unstructured cytoplasmic C-

terminus (Lange et al., 2007). Lange et al. (2007) also showed that the amphipathic helix is aligned parallel to the membrane. TatE, with its 67 residues (7.0 kDa), is a paralogue protein to TatA with 42.7 % identity and 57.3 % similarity (Sargent et al., 1998). This protein is 22 residues shorter than TatA but still presents a similar membrane conformation (Figure 1-5A). Its function overlaps with TatA with some substrates seemingly to require one, the other or both (Sargent et al., 1998). Single mutant strains lacking either TatA or TatE are still able to translocate Tat substrates but to a lower rate whereas Tat export is inactivated in a double mutant.

TatB is a 171 amino acid long protein (18.4 kDa) sharing a 23.8 % similarity with TatA and a very similar predicted secondary structure (Sargent et al., 1999). TatB possesses two short cytoplasmic α-helixes on top of the amphipathic and the transmembrane helixes (Figure 1-5A) (Zhang et al., 2014b). Despite the resemblance, these proteins have a functionally distinct role. Moreover, the abundance of TatA is roughly 20 times higher than TatB in the membrane (Sargent et al., 2001, Jack et al., 2001). A TatB-null mutant strain was reported to mislocalise the Tat-dependant trimethylamine N-oxide (TMAO) reductase indicating that TatB is essential for a functional Tat pathway (Chanal et al., 1998). TatC is a 258 amino acids (28.9 kDa) protein characterised by six transmembrane domains with both N- and C-termini located in the cytoplasm (Punginelli et al., 2007).

The Tat proteins TatA, TatB, TatC and TatD are expressed from the *tatABCD* operon and TatE encoded as a single gene (Figure 1-5B) (Sargent et al., 1998, Weiner et al., 1998). The gene coding for TatE locates at about 14 min on the genome while the *tatABCD* operon locates at about 78 min and is thought to be the result of *tatA* gene duplication (Yen et al., 2002). TatD was identified as a cytoplasmic protein with magnesium-dependant DNase activity. This protein was found to be non-essential for the export of Tat-dependant proteins as a TatD knock-out strain was not impaired for Tat translocation (Wexler et al., 2000). Matos et al. (2009) discovered that, even though TatD is not essential for Tat substrates export, it is required for the quality control mechanism (discussed in section 1.3.2.5). Indeed, Matos et al. (2009) expressed the NrfC, NapG and FhuD Tat substrates which were rendered unsuitable for Tat export by mutating one of their iron-sulphur site. The team found

these mutants to be stable in the cytoplasm of the TatD-null strain whilst rapidly degraded in the wild-type (WT) strain. The latest research on this protein revealed that TatD possesses a 3'-5' exonuclease activity responsible for chromosomal DNA degradation during apoptosis and repair of single stranded DNA damage (Chen et al., 2014). The correlation between the DNase activity and the role in quality control mechanism of the Tat pathway is still unclear.

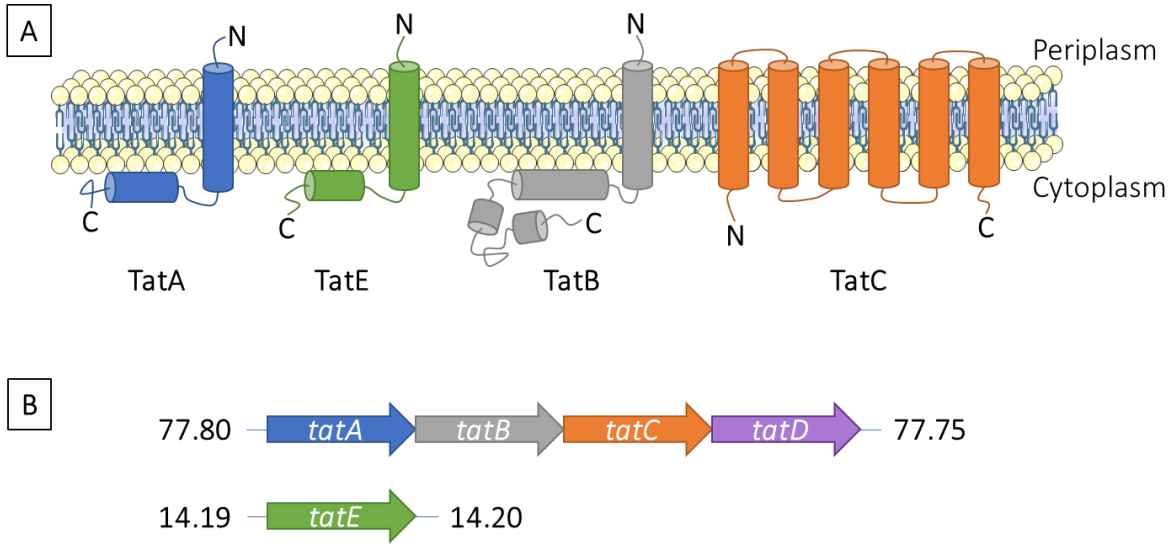


Figure 1-5: Organisation of the *E. coli* Tat system

A. Representation of the integral inner membrane Tat proteins where N and C represent the amino- and carboxy-termini of each protein. The cylinders represent α-helixes. **B.** Arrangement of the *tat* genes in the genome of *E. coli* K12 W3110 strain: the *tatABCD* operon and the isolated *tatE* gene. The minutes are indicated on each side of the loci. Adapted from Patel et al. (2014).

1.3.2.4. Tat export mechanism

The general mechanism by which Tat exports folded proteins across the inner membrane is not fully understood. However, the protein structures of some of the Tat components can be used to provide insights. The structure of TatA and TatB proteins from *E. coli* has been solved by NMR in detergent solution (Rodriguez et al., 2013, Zhang et al., 2014a, Zhang et al., 2014b) but no structure of the other *E. coli* Tat proteins as isolated components have been published. However, the structure of TatC from *Aquifex aeolicus*, a hyperthermophilic Gram-negative bacterium, has been solved by

crystallography and found to share 52% similarity with its *E. coli* homologue (Rollauer et al., 2012, Ramasamy et al., 2013). Some structures of complexes have also been published such as TatBC of *E. coli* with and without substrate binding by single-particle electron microscopy (Tarry et al., 2009).

Using the solved structures, many researchers have attempted to elucidate this mechanism and collectively their data has led to the proposal of a mechanism for the recruitment of the Tat complex by folded Tat substrates which is generally accepted. However, the actual mechanism for the formation of a channel and subsequent translocation is not fully understood. The presence of folded Tat substrates with presentation of the N-terminal Tat-specific signal peptide in the cytoplasm triggers the translocation event by reversible recruitment of the translocon components (Rose et al., 2013). The first step of Tat translocation is the association of TatB and TatC proteins as dimers in the membrane which act as receptor for the substrates (Alami et al., 2003, McDevitt et al., 2006, Frobel et al., 2019). This association is critical to prevent premature release of the substrate. TatC is an insertase which alone can recognize Tat substrates through their signal peptide but can prematurely trigger their cleavage in the absence of TatB (Frobel et al., 2012). TatB is therefore required for the correct initiation of translocation by preventing early cleavage of the Tat signal peptide. Together they form a hydrophobic cavity inside the lipid bilayer capable of docking the Tat signal peptides (Blummel et al., 2017). The twin-arginine motif is first recognized by the loop located between the transmembrane helixes 2 and 3 of TatC (Alami et al., 2003, Gerard and Cline, 2006). The Tat signal peptide is then inserted deeply into a cavity formed by the TatBC complex (Figure 1-6) (Blummel et al., 2015, Lausberg et al., 2012, Huang et al., 2017). A recent study claims that the early mature part of the substrate also plays a role in TatBC binding by having a tighter interaction with the first residues of the mature domain when their hydrophobicity is increased (Ulfig and Freudl, 2018).

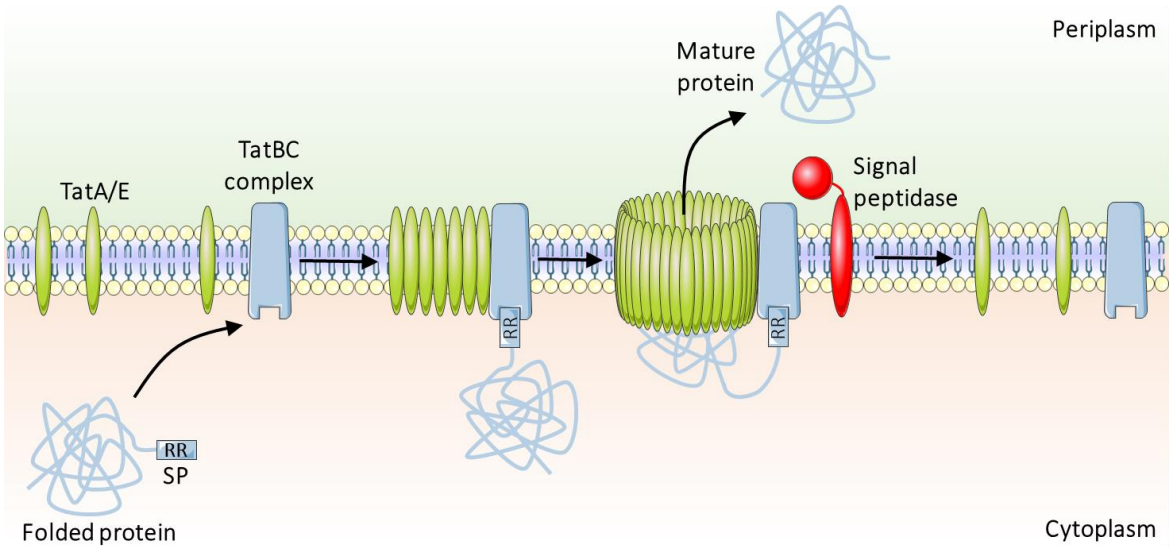


Figure 1-6: The TatA/E-mediated channel model of the Tat pathway of *E. coli*

Tat-specific substrates contain a N-terminal signal peptide (SP) with a conserved twin-arginine (RR). They fold in the cytoplasm prior to targeting to the membrane TatBC complex. TatA/E proteins are then recruited to the TatBC/substrate complex and form a transient pore to accommodate the size of the substrate. Once the substrate is translocated across the cytoplasmic membrane, the SP is cleaved by the signal peptidase and the mature protein is released in the periplasm while the pore dissociates. Adapted from Lee et al. (2006) and Palmer and Berks (2012)

The second step is the recruitment and oligomerization of TatA proteins upon binding of the substrate to the TatBC complexes (Alcock et al., 2013). This recruitment is dependent on the proton motive force (Patel et al., 2014). TatE was also demonstrated to be recruited along with TatA and clustered at the Tat complex proving it plays a part of Tat translocation (Eimer et al., 2015). From this point, two models are still being discussed to explain the translocation of the substrate: the TatA-mediated channel (Figure 1-6) and the transient membrane disruption models (Berks, 2015, Cline, 2015). The channel hypothesis describes oligomerization of TatA proteins at the TatBC/substrate complex to form a size-variable pore adapted for the bound substrate (Lange et al., 2007, Patel et al., 2014). In a different way, the latter model suggests that upon oligomerization of TatA proteins, the membrane is thinned and destabilized thanks to the N-terminus of these proteins and a transient pore into the lipid bilayer is opened, allowing transit of the substrate (Hou et al., 2018). Altogether, the Tat translocon is most likely to use both mechanisms to adapt to the various substrates up to 142 kDa which is the upper molecular weight limit identified (Berks et al., 2000).

1.3.2.5. The proofreading mechanism

All the molybdoenzymes cited in section 1.3.2.1 require their respective cytoplasmic chaperone to be exported by the Tat pathway. Without their chaperones, these Tat substrates cannot reach a native conformation which is sensed by the translocon thus blocking transport. The ability to sense the folded state of Tat-targeted proteins is referred to as the 'proofreading' or 'quality control' mechanism (DeLisa et al., 2003, Sutherland et al., 2018). This unique ability has increased interest to exploit the Tat pathway (Waraho-Zhmayev et al., 2013). Rocco et al. (2012) identified suppressors mutations on the TatABC proteins which allowed a misfolded model protein to be exported to the periplasm. Such a protein was designed for genetic selection as a tripartite fusion containing a test protein inserted between the signal peptide of the natural Tat substrate TorA and β-lactamase. The test protein itself was designed to be a monomeric, three-helix-bundle protein forming molten globules. Moreover, reduced PhoA, another misfolded substrate, fused to the signal peptide of TorA was found to reach the Tat translocase but was not exported (Panahandeh et al., 2008, Richter and Bruser, 2005). The latter associated with the TatBC complex, but this did not trigger the recruitment of the TatA proteins. The degradation of substrates that cannot reach a native conformation is triggered by the proofreading system upon interaction with the Tat translocase (Matos et al., 2008). Overall, it was concluded that the Tat translocase "senses" the targeted substrate by the so called "quality control" system and rejects the incorrectly folded proteins leading to their degradation. However, the limit of the folding sensitivity is still unclear.

Recent studies have shed light on the conformational tolerance of the Tat proofreading mechanism. The natural Tat substrate cuprous oxidase (CueO) is exported without its copper cofactor which is acquired in the periplasm (Djoko et al., 2010, Roberts et al., 2002). This apo form of the protein is exported by Tat in native conditions (Stolle et al., 2016). The apo form is therefore sufficiently folded to pass the quality control system but still presents a higher conformational flexibility than the holo-form. The tolerance of the Tat proofreading mechanism on the substrate flexibility was further demonstrated by Jones et al. (2016) with a non-natural Tat substrate. Indeed, they showed that a

Tat targeted signal-chain variable fragment (scFv) had about 10-fold enhanced export when the scFv was more tightly folded with its two intra-molecular DSB than without them. Moreover, they discovered that this scFv was not Tat-translocated when fused to a 26 amino acids unstructured tail. The tolerance to a certain conformational flexibility was further demonstrated on another recombinant Tat substrate. Using a heme-binding protein known as BT6 maquette, Sutherland et al. (2018) observed the correlation between enhancement of Tat export and increase of substrate's rigidity. Indeed, the changes in the folding extent of three variants of BT6 binding to either two, one or no heme(s) was demonstrated by circular dichroism and NMR spectroscopy. The three variants were then individually fused to the signal peptide of TorA and expressed in *E. coli*. The rising abundance of periplasmic BT6 clearly correlated with the increasing heme-binding capacity indicating a link between rigidity status of the substrate and the translocation event. However, it is also claimed that the Tat pathway offers a degree of flexibility with respect to surface hydrophobicity and net charge. Indeed, mutating surface hydrophobic residues from hydrophilic residues to leucines, or the net charge from uncharged residues to lysines and/or glutamines of the scFv were shown not to affect Tat export (Jones et al., 2016).

The proofreading mechanism is still not clear on how it rejects certain proteins. Misfolded proteins still interact with the Tat translocon as demonstrated by the co-purification of a mutant PhoA with the TatBC complex (Richter and Bruser, 2005). Thus, Tat may not possess an inbuilt proofreading mechanism but is rather linked to an efficient degradation system (Frain et al., 2019). This idea is supported by the direct interaction of the Tat machinery with the protease WprA which is essential for Tat export in *Bacillus subtilis* (Krishnappa et al., 2012, Monteferrante et al., 2013).

1.3.2.6. The Tat systems beyond *E. coli*

The Tat pathway is an evolutionary conserved mechanism found in 77 % of known bacteria (Simone et al., 2013). Not only is it present in Gram-negative bacteria, it is also present in other prokaryotes like Gram-positive and Archaea bacteria which both possess a system often considered "minimalistic" named TatAC. For instance, *Bacillus subtilis* has two TatAC isoforms, TatAdCd and

TatAyCy, with specific substrates (Simone et al., 2013). *E. coli* TatA (89 bp) has 24%, 42% and 35% similarity with TatB (171 bp) from *E. coli*, TatAd (70 bp) and TatAy (57 bp) from *Bacillus subtilis* respectively (Jongbloed et al., 2000). The *tatA*-like gene *tatB* is the result of a divergent evolution of the *tatA* gene found in the Proteobacteria Actinobacteria while *tatE* is the result of a second duplication event found only in enterobacteria (Wu et al., 2000). Hence, TatA encoded by one or more copy of a *tatA*-like gene fulfils the function of both TatA and TatB of a TatABC system like *E. coli* (Palmer and Berks, 2012). This was further demonstrated by the successful complementation of a Tat-deficient *E. coli* strain by the TatAdCd system (Barnett et al., 2008, Albiniak et al., 2013). While most bacteria have a low abundance of Tat substrates, the predicted Tat substrate within the secretome of Archae bacteria such as *Natrialba magadii* can reach up to 55% (Storf et al., 2010).

Beyond the prokaryotic organisms, a homologous system has been discovered in plant cells. A Tat pathway was identified in the thylakoid membranes of plant chloroplasts (Cline et al., 1992, Henry et al., 1994, Hulford et al., 1994). It was first named ΔpH-dependant pathway and was distinguished from the Sec-equivalent pathway since the latter requires both energy from ATP and the thylakoidal ΔpH whereas the plant Tat pathway only requires the ΔpH (Mori and Cline, 2001). This system uses three membrane proteins: thylakoid assembly 4 (Tha4), high chlorophyll 106 (Hcf106) and chloroplast TatC (cpTatC) which are the bacterial orthologues of TatA, TatB and TatC respectively (Natale et al., 2008). This system is essential for photosynthesis because it is responsible of the assembly of photosystem II and cytochrome b_6f by inserting the Rieske iron-sulphur complex into the thylakoid membrane (Molik et al., 2001). Recent findings suggest that plant mitochondria, which possess their own copy of TatB- and TatC-like proteins, use a functional Tat pathway to insert the Rieske iron-sulphur complex protein into the membrane (Carrie et al., 2016, Kolli et al., 2018).

The Tat system is absent from the animal and fungus kingdoms with only two exceptions. The first one is the choanoflagellate *Monosiga brevicollis* (Burger et al., 2003). The second one is the homoscleromorph sponges, Oscarellidae (Wang and Lavrov, 2007). In both organisms, only a TatC-like protein has been identified and share a respective 39.3 % and 39.7 % similarity with *E. coli* TatC.

Genes coding for these proteins are located in the mitochondrial DNA and the sponge TatC-like protein has been shown to be active in *Oscarella carmela* for membrane protein insertion (Pett and Lavrov, 2013). The authors theorized that, since the sponge Tat system does not have a TatA-like protein which is necessary to form the pore even in minimalist systems, the mitochondrial Tat system is only able to insert membrane proteins rather than protein translocation.

1.3.3. The Type I to VI secretion pathways

The Type I secretion system is a one-step process which translocates unfolded substrates across the cell envelope using energy from catalysing ATP (Thomas et al., 2014). Three essential components are required for this process: one ABC transporter protein in the inner membrane, one membrane fusion protein bridging the inner membrane to the outer membrane and one outer membrane porin-like protein. Type 1 substrates include proteins that often contribute to bacterial virulence including proteases, lipases, adhesins and heme-binding proteins. These substrates bare a C-terminal signal sequence specific for Type I secretion which is not cleaved following export. As this system allows DSB formation during transport and offers the advantage of protein secretion into the extracellular environment, thus simplifying downstream processing, it has been exploited for overexpression of recombinant proteins. Indeed, β-galactosidase, prochymosin, chloramphenicol acetyltransferase and green fluorescent protein (GFP) were shown to be expressed when fused to the C-terminal domain of the Type I natural substrate HlyC, although absolute expression yields have not been published (Kenny et al., 1991, Linton et al., 2012). However, to avoid saturation of the ABC transporter, overexpression of the translocase components is necessary (Fernandez and de Lorenzo, 2001).

The Type II secretion pathway exports proteins from the periplasm into the extracellular media and is therefore considered a two-step process (Cianciotto and White, 2017). The first step exports the protein into the periplasm using Sec or Tat systems. These substrates contain a N-terminal signal peptide for Sec or Tat export (sections 1.3.1.1 and 1.3.2.1) which is cleaved before release of the mature protein into the periplasm. The second step consists of the translocation of the folded

substrate from the periplasm to the extra-cellular medium through a complex channel containing at least the 12 "core" proteins named T2S C, D, E, F, G, H, I, J, K, L, M, and O and uses ATP as an energy source (Cianciotto and White, 2017). The Type II system is restricted to *Proteobacteria* and mostly contributes to the pathogenic virulence by exporting proteases for example.

The Type III secretion system secretes proteins across the inner and outer membranes and even across a target cell membrane using a syringe-like channel upon contact with the target cell (Buttner, 2012). The complex secretes unfolded effectors essential for the virulence of many pathogens in an ATP-dependent manner. The secretion signals are specific peptides embedded within the N-terminus of the substrates and are not cleaved through the secretion process.

The Type IV secretion pathway is also triggered by contact with the target cell and transport substrates from the cytoplasm to the extra-cellular medium or to the target cell via a pilus-based structure (Cascales and Christie, 2003). The recipient cell can be a bacterium from a different or the same species or even a eukaryotic cell. Substrates include single proteins, interacting proteins and protein-DNA complexes which serve a variety of roles such as DNA conjugation, DNA uptake and release, biofilm formation and effector translocation.

The Type V secretion system is unique as the substrates do not require a translocation apparatus or the assistance of a channel but rather transport themselves (van Ulsen et al., 2014). Three classes of substrates have been identified: auto-transporter, two-partner secretion and chaperone-usher secretion (Green and Mecsas, 2016). In each class, the substrate is first translocated across the inner membrane via the Sec pathway using a N-terminal signal peptide which is cleaved off on entry into the periplasm. Then, the β-barrel domain of the substrate protein inserts itself into the outer membrane and forms a channel allowing the remainder of the protein or another protein to be transported. Most of the substrates are involved in virulence.

The Type VI secretion pathway is the most recently discovered and therefore, not yet fully characterised (Russell et al., 2014). It is however known to have an envelope-spanning complex which has a tubular structure and found to be related to the tail of contractile phages evolutionarily,

structurally and functionally (Zoued et al., 2014). This system secretes antagonistic and non-antagonistic effector proteins from one bacterium to another or to a eukaryotic target cell. Substrates are linked to interspecies bacterial competition, virulence, bacterial communication and interactions in the environment.

1.3.4. Unknown secretion pathway(s)

Some recombinant proteins have shown distinct secretion capabilities which are not yet explained. For instance, the L-asparaginase, an enzyme used as a therapeutic to treat leukaemia, has been produced in *E. coli* (Khushoo et al., 2004). The protein was expressed in fusion with the signal peptide of PelB, a Sec-exported periplasmic protein, and found in high amounts in the extra-cellular medium of high-cell density cultures. The reason for the translocation of the recombinant protein across the outer membrane is still unclear. Rinas and Hoffmann (2004) published that this phenomenon was not specific to the L-asparaginase but was indeed observed when expressing other recombinant proteins into the periplasm in high-cell density cultures. They also observed that host periplasmic proteins such as DppA and MalE were also found into the cell-free media and in one case, including, the cytoplasmic host chaperone DnaK. The localisation of native periplasmic proteins into the extra-cellular media was not linked to the overexpression of a recombinant protein since they proved that non-recombinant strains also show the same “export” pattern in high-cell density cultures. This “export” was hypothesized to be the result of the loss of cell envelope integrity (Han et al., 2003, Ni and Chen, 2009). Periplasmic leakage has been proposed to be due to a number of reasons including cell division prior to formation of the individual outer membranes, accumulation of recombinant proteins in the periplasm causing a build-up of osmotic pressure, perturbation of the membrane induced by recombinant proteins to cellular lysis as a result of periplasmic secretion (Mergulhao et al., 2005).

The above examples depict an unknown mechanism by which a recombinant protein targeted to the periplasm crosses the outer membrane. A recent study describes the secretion of a heterologous protein from the cytoplasm into the extra-cellular medium without the help of a N-terminal signal

sequence or dedicated secretion pathway. The superfolder green fluorescent protein (sfGFP) was indeed discovered in the cytoplasm, periplasm, outer membrane and extra-cellular compartments following expression in *E. coli* cultures (Zhang et al., 2017). No putative signal peptide within the N-terminal primary amino acid sequence of sfGFP was identified by prediction software. Deletion of the first 10 and 20 residues of the sfGFP sequence did not prevent extra-cellular export, providing further support that an N-terminal signal peptide was not involved in this type of autosecretion. The authors speculated that sfGFP uses its characteristic β-barrel structure to cross both membranes and this was further reinforced by their discovery that the related red fluorescent protein, mCherry which also forms a β-barrel, is also autosecreted in a similar fashion.

Morra et al. (2018) published even more recently, a detailed analysis on unknown secretion events. Using enhanced GFP (eGFP) as the overexpressed recombinant protein, they discovered that secretion was not due to cell lysis, does not involve outer membrane vesicle budding and was not related to cultivation conditions, but rather to high expression levels. The authors further studied the stresses related to recombinant protein overexpression and observed that overexpression induced greater hypo-osmotic conditions within the cytoplasm resulting in increased membrane permeability. They also discovered that the large mechanosensitive channel (MscL) had a direct role in this secretion. Finally, translational stress also causes 'autosecretion' of cytoplasmic proteins and this appears to be mediated by the alternative ribosome rescue factor A (ArfA) (Morra et al., 2018). The translational stress was induced by chloramphenicol, an antibiotic known to cause translation arrest (Marks et al., 2016). This stress caused several proteins to 'autosecrete' including the elongation factor Tu 1 and was found to be reproducible even in untransformed WT strain. This discovery matched the results from Kapoor et al. (2017) who identified other antibiotics such as aminoglycosides, tetracyclines and streptogramins that directly inhibit ribosome translation, which can also induce autosecretion to the extra-cellular medium. The conclusions from this study are limited since the authors used a GFP protein which could still be crossing the cell envelope as

indicated in the previous paragraph. Furthermore, implications of the role that MscL and ArfA play with respect to autosecretion of specific proteins also need to be considered.

1.4. Engineering to improve *E. coli* as an expression host

Proteins expressed in high organisms have access to different environments, post-translational machineries as well as different chaperones. Therefore, when such proteins are produced in simpler organism such as *E. coli*, they may require access to similar conditions to acquire the same biological activity. These conditions can be created by refolding or the use of engineered strains for instance. The post-translational modifications, codon usage and inclusion body formation are the main drawbacks when using *E. coli* as a host (Rosano and Ceccarelli, 2014, Baeshen et al., 2015). Heterologous proteins may require post-translational modifications for their solubility and activity such as glycosylation, phosphorylation, acetylation and disulphide bonds (DSB) (Chen, 2012). Although *E. coli* do not possess the same apparatus as eukaryotes for post-translational modifications, it still has some features like the disulphide bond formation which naturally takes place in the only oxidising environment of the cell: the periplasm. Moreover, the success rate of producing soluble proteins in *E. coli* decreases considerably when the molecular weight is above 60 kDa (Gräslund et al., 2008). However, *E. coli*, being a bacterium, has as number of advantages in relation to being an expression host (section 1.1) and therefore systems have been developed to bypass these bottlenecks.

1.4.1. Disulphide bond formation

Many recombinant proteins require the formation of correct DSB to reach their biologically active three-dimensional conformation. DSB require an oxidising environment to be formed such as the periplasm and cannot form into the cytoplasm of *E. coli* since it is naturally a reducing environment. Several strategies have been developed to bypass this limitation.

1.4.1.1. The naturally oxidising periplasmic space

E. coli possesses an oxidising environment, the periplasm, which hosts DSB formation and isomerization catalysts (Mergulhao et al., 2005, de Marco, 2009). This naturally oxidising environment hosts the Dsb family proteins which enable DSB formation and are conserved between Gram-negative bacteria (Collet and Bardwell, 2002, Messens and Collet, 2006). The soluble periplasmic DsbA oxidizes cysteine pairs of substrate proteins to form DSBs and is maintained in an oxidizing form by the integral inner membrane DsbB. However, the chance of forming a native DSB decreases as the number of cysteine pairs in a protein increases and thus these DSBs are often incorrectly paired, trapping the protein in a non-native form. The isomerases DsbC use their quality control mechanism to shuffle DSBs into their native conformation. This protein is also known to have chaperone activities by which DsbC assists GAPDH reactivation (Chen et al., 1999, Denoncin et al., 2010). DsbG acts as a chaperone to reduce the amounts of protein aggregates in cells with a damaged envelope (Depuydt et al., 2009, Kouidmi et al., 2018). The isomerases are kept in a reducing state by the integral inner membrane DsbD.

1.4.1.2. Disruption of the reducing pathways

The cytoplasm is maintained as a reducing environment with a negative redox potential by the thioredoxin and the glutaredoxin systems which also alleviates damaging oxidising stress (Rosano and Ceccarelli, 2014). Thioredoxin 1 (encoded by *trxA*) reduces substrate proteins and is recycled by the thioredoxin reductase (encoded by *trxB*) (Aberg et al., 1989, Stewart et al., 1998). On the other hand, glutathione uses the reducing potential of NADPH via the glutathione oxidoreductase (encoded by *gor*) to reduce the three glutaredoxins (encoded by *grxA*, *grxB*, *grxC*) (Zheng et al., 1998, Stewart et al., 1998). The latter three enzymes have DSB reduction activities and therefore participate in maintaining the negative redox potential. Therefore, efforts have been made to disrupt these pathways to obtain an oxidising cytoplasm enabling DSB formation (Prinz et al., 1997). By mutating both the thioredoxin and the glutaredoxin pathways, substantially more DSBs were formed than in mutant missing only one reducing pathway (Prinz et al., 1997). However, such a double

mutant was unable to grow aerobically and required the addition of a disulphide reductant to grow at a normal rate. Suppressor mutations in the *ahpC* gene has been shown to solve this problem by conferring protein an enhanced activity towards the reduction of glutathionylated glutaredoxins to the AhpC (Ritz et al., 2001, Yamamoto et al., 2008). The commercially available strain Origami (Novagen) is a double knock-out *trxB* and *gor* mutant with the suppressor mutations on *ahpC* allowing cytoplasmic DSB formation without impairing aerobic growth (Bessette et al., 1999). This strain has been optimized further onto the commercial strain SHuffle (New England Biolabs) which also expresses a chromosomal copy of the periplasmic *dsbC* without its signal peptide (Lobstein et al., 2012).

1.4.1.3. The CyDisCo technology

An alternative from disrupting these pathways is to co-express cytoplasmic DSBs catalysts. Indeed, Hatahet et al. (2010) expressed cytoplasmic mutants of the periplasmic enzymes PhoA and phytase from *E. coli* which require two and four DSBs respectively for activity. The team demonstrated that these enzymes can be produced in an active form in the cytoplasm of a WT *E. coli* strain by co-expressing the sulfhydryl oxidase Erv1p. This essential DSB catalyst was isolated from the inner membrane space of *Saccharomyces cerevisiae* (Lee et al., 2000).

Other DSB-modifying proteins were later co-expressed with Erv1p to improve DSB conformation (Nguyen et al., 2011). A fragment of tissue plasminogen activator, bovine pancreatic trypsin inhibitor and resistin were all produced with their respective nine, three and six DSBs. When expressing these proteins, similar or higher yields were obtained when the catalyst for DSB formation Erv1p was co-expressed with a disulphide isomerase DsbC compared to expression of Erv1p alone (Nguyen et al., 2011). Disulphide isomerases enable rearrangement of incorrect DSB. This result supported a growing body of evidence highlighting that the production of eukaryotic proteins with multiple DSBs in the cytoplasm of *E. coli* is possible.

A new system named CyDisCo for cytoplasmic formation of disulphide bonds in *E. coli*, was hence developed and co-expresses Erv1p and PDI along with the protein of interest to allow formation of

cytoplasmic DSB (Matos et al., 2014). PDI is the human protein disulphide isomerase which has been shown to complement a DsbC-dependent phenotype (Stafford and Lund, 2000). This technology was used in the cytoplasm with intact reducing pathways to produce antibody fragments scFv and Fab presenting specific binding to their ligand proving the correct status of the DSBs (Gaciarz et al., 2016, Gaciarz and Ruddock, 2017). CyDisCo was also tested in glucose fed-batch fermentation where hen avidin, single chain IgA_1 antibody fragment, hGH and interleukin 6 reached around 71 mg/ml, 139 mg/ml, 1 g/L and 1 g/L respectively (Gaciarz et al., 2017). Although the correct number of DSBs were proven on the purified proteins using Ellman's assay, no activity assays were performed to confirm the active state of these proteins. This technology offers an alternative to the gene disruption schemes discussed in section 1.4.1.2., where engineered strains are known to have reduced fitness. Although, it has been reported that expression appears with the latter system to be variable and the activity of expressed recombinant proteins still need to be demonstrated.

1.4.2. Glycosylation in *E. coli*

Glycosylation is the modification of polypeptide chains or lipids through the attachment of carbohydrate(s) and are found in eukaryotes as well as archaea and eubacteria (Varki, 2017). The nomenclature to identify a glycosylation is based on their linkage type. The most widely distributed carbohydrate-peptide bonds are the *N*- and the *O*-glycosidic bonds (Spiro, 2002). They can be linked to an asparagine or an arginine residue or to a serine, a threonine, a hydroxylysine, a hydroxyproline or a tyrosine residue respectively. In eukaryotes, glycosylation is the predominant post-translational modification and is difficult to replicate since *E. coli* does not naturally own a glycosylation pathway. Some native glycoproteins can still be expressed without their glycans in an active form in this organism such as interleukin-2 (Roifman et al., 1985) while others are unstable and aggregate like EPO (Crommelin et al., 2008). When the glycans are mandatory, a couple of solutions can be tested including mutagenesis of the residues which are the sites of the modifications (Narhi et al., 2001) or by transferring a glycosylation pathway to *E. coli* (Wacker et al., 2002, Nothaft and Szymanski, 2013, Valderrama-Rincon et al., 2012). Indeed, Wacker et al. (2002) successfully transferred the

glycosylation system from the bacterium *Campylobacter jejuni* to *E. coli* allowing N-glycosylation of recombinant proteins. Although bacterial N-glycans are different from their eukaryotic counterparts, efforts have been made to develop mutant strains allowing eukaryotic glycosylation (Valderrama-Rincon et al., 2012, Nothaft and Szymanski, 2013).

1.4.3. Codon usage

In all living organisms, synonymous codons, with the same encoded amino acid, are not used in equal frequencies and this phenomenon is termed codon bias (Zhou et al., 2016). The rarity of a codon can impact the expression of a transgene by up to 1000 fold (Gustafsson et al., 2004). Moreover, codon usage differs significantly in species across all taxa but seldom between closely related species (Sharp et al., 1995, Plotkin and Kudla, 2011). Indeed, some codons from heterologous genes are considered rare in *E. coli* and can therefore, led to translational errors (Kane, 1995).

Two main options are available to bypass this issue: modifying the rare codons in the target gene to reflect the codon usage of *E. coli* either by directed mutagenesis or by whole gene synthesis or increasing the availability of limiting tRNAs using a plasmid or a mutant strain carrying the genes coding for the underrepresented tRNAs (Baeshen et al., 2015, Rosano and Ceccarelli, 2014). However, a recent study demonstrated that keeping rare codons at the N-terminal can increase expression yields by ≈ 14 instead of using common codons (Goodman et al., 2013).

1.4.4. Stabilization of the recombinant product

Heterologous proteins can aggregate when their solubility is impaired resulting in the formation of inclusion bodies. While solutions are available for the resolubilization of inclusion bodies into soluble and active proteins, they remain long, tedious and expensive processes (Singh et al., 2015, He and Ohnishi, 2017). Heterologous protein aggregates are mostly due to solubility, cofactors, folding mechanisms and osmolarity which differ from the native organism (Rosano and Ceccarelli, 2014). To counter these issues, one can enable DSB formation, use an engineered strain permitting bacterial glycosylation, improve the codon usage, aid the protein stability by knocking-out proteases, co-

expressing chaperones or use of genetic fusions or slow down the production rate (Rosano and Ceccarelli, 2014, Baeshen et al., 2015).

1.4.4.1. Use of protease-deficient strains

One way to increase the yield of the recombinant protein is by preventing its degradation. For instance, the Lon and OmpT proteases were knocked-out in the popular strain BL21 (Daegelen et al., 2009). Indeed, the cytoplasmic Lon protease has been shown to degrade abnormal proteins and many heterologous proteins (Gottesman, 1996). On the other hand, OmpT is a periplasmic protein located in the outer membrane and degrades proteins of interest in cell extracts after lysis (Gottesman, 1996, Rosano and Ceccarelli, 2014). Another example of protease deficient strain was published by Ellis et al. (2017) where the deletion of the Tsp, protease III and DegP proteases were assessed on the expression of a Fab' antibody fragment. The Tsp mutant strain gave the optimum yields but required a suppressor mutation on the *spr* gene to maintain the cell viability. The deficiency of the two cell envelope DegP and Tsp proteases were previously shown to be beneficial for the growth rate and the production yield (Park et al., 1999, Chen et al., 2004).

1.4.4.2. Co-expressing chaperones

Chaperone proteins improve protein solubility by interacting with nascent polypeptide chains during synthesis and prevent aggregation or promote refolding and solubilisation of misfolded proteins (Carrio and Villaverde, 2003). Co-expressing chaperones have been shown to enhance the solubility of the protein of interest and hence, the yields (Rosano and Ceccarelli, 2014). The most popular chaperones are GroEL, GroES, DnaK, DnaJ, GrpE and trigger factor which are used individually or in sets (Baeshen et al., 2015). For instance, 65 % of a scFv co-expressed with GroEL and GroES were found soluble in the cytoplasm against 27 % when expressed without these chaperones (Maeng et al., 2011). Another example was the co-expression of Tat-targeted organophosphorus hydrolase with the GroEL/GroES combination, allowing 5.5 fold higher activity of the enzyme than without these chaperones (Kang et al., 2012). However, co-expressing a chaperone does not guarantee improved

solubility as it may lower the overall protein productivity by causing an extra-metabolic burden to the cells (Huang et al., 2012).

1.4.4.3. Fusion of a soluble partner

Improving the solubility of a recombinant protein by fusion with a soluble partner is a well-established concept (Khow and Suntrarachun, 2012). A range of soluble proteins are available from the literature such as MBP, GST and SUMO and offer a successful alternative to solubilizing inclusion bodies (Peciak et al., 2014, Sorensen and Mortensen, 2005). For instance, MBP (43 kDa) was shown to be an excellent solubility enhancer when fused to the N-terminal of the protein of interest (Raran-Kurussi et al., 2015). Fusion partners are also used to improve folding, production yield, to facilitate purification or to prevent degradation of the target protein (Costa et al., 2014). Also, the small antigen Fh8 (8 kDa), isolated from the parasite *Fasciola hepatica*, has a combined feature of enhancing the target protein solubility and its purification (Costa et al., 2013a, Costa et al., 2013b).

1.4.4.4. Improving the process

The options mentioned earlier in this section require genetic manipulations but optimisation of the expression process can also influence the protein yields (Rosano and Ceccarelli, 2014). Indeed, slowing the rate of protein production has been demonstrated to allow more time for protein folding and reduce the cellular protein concentration which both improves the protein solubility. The most common way to achieve this objective is by lowering the incubation temperature during protein expression which was shown to also minimize the formation of inclusion bodies (Fahnert et al., 2004). However, this approach implicates to work with reduced growth and synthesis rates leading to lower protein yields. Several other parameters can be optimised such as medium composition, plasmid copy number, promoter strength, type of tag or inducer concentration (Baeshen et al., 2015).

1.4.5. Other engineered strains

Over the past decades, more strains have been developed to bypass the bottlenecks of expressing recombinant proteins, especially of eukaryotic origin, in *E. coli*. For instance, a strain carrying

mutation on the *hns* gene allowed greater protein production upon addition of indole (or derivative) which stops cell growth but maintain metabolic activity (Chen et al., 2015, Thomson et al., 2018). Moreover, a *E. coli* K12 strain was developed to counter the low abundance of Tat components in WT cells in order to achieve higher yields when exporting a recombinant protein via the Tat pathway. Indeed, Thomson et al. (2018) engineered a W3110 strain capable of over-expressing the TatABC proteins upon induction and obtained a 5-fold higher periplasmic yield of Tat-targeted (hGH) and scFv compared to the WT strain.

1.5. Aims and hypothesis

The cytoplasm of *E. coli* is a reducing environment hostile for DSB formation, whereas the periplasmic space is an oxidising environment where DSBs can naturally be formed by way of the Dsb machinery (section 1.3.1.). It also offers the advantages of containing chaperones that can enhance the correct folding of recombinant proteins and contains fewer proteases than the cytoplasm enabling a higher protein stability (Mergulhao et al., 2005). Overall, the periplasmic space is less crowded, facilitating downstream processing (Balasundaram et al., 2009). For these reasons, producing recombinant proteins into the periplasm offers an attractive alternative solution to cytoplasmic expression. To this aim, proteins of interest have been exported via the Sec mechanism. The more recently discovered Tat pathway offers the advantage of translocating folded proteins whereas Sec can only handle unfolded polypeptides. Moreover, the proofreading ability of Tat would be a first screening step for selecting active proteins easing further the downstream processing. The aim of this study is to evaluate the Tat pathway for the production of the recombinant proteins. Such production has already been successfully performed in *E. coli* but only for a limited number of proteins (section 3.1.). The approach chosen was to fuse a soluble partner to the proteins of interest to broaden the range of recombinant proteins accepted by Tat. The initial hypothesis was to use a natural Tat substrate (NTS) as the soluble partner to both improve the solubility of the partner and target the complex to the Tat pathway.

First, a robust fractionation method was developed to thoroughly analyse the localisation of proteins of interest in the cellular sub-compartments (Chapter 4). Then, a selection of NTS was evaluated for their use as soluble partner in the fusion strategy (Chapter 5). In parallel, several reporter proteins were assessed to play the role of the protein of interest in the fusion design (Chapter 6). Following the failure to identify a suitable soluble partner amongst the *E. coli*'s NTS, the strategy was adapted to use the validated reporter proteins as soluble partner and fuse them to protein of interests (Chapter 7). Finally, a feature near the *tat* operon suggested by the recent development of an engineered Tat strain was investigated (Chapter 8).

2. MATERIALS AND METHODS

2.1. Strains

The strains used during the course of this project are listed in Table 2-1. Chemically competent cells were prepared using the following protocol for all strains except TOP10 (Life Technologies) and SHuffle (New England BioLabs) which were bought as ready-to-use aliquots.

Chemically competent cells were prepared using the TSS method described previously (Chung et al., 1989). Briefly, 10 ml of LB was inoculated with a single colony from a fresh overnight LB agar plate and the culture was incubated for 16 h at 37°C with shaking at 220 rpm. The next day, 100 ml of 2xTY media was inoculated with 1 ml of the preculture and incubated at 37°C with shaking at 220 rpm. Once the culture reached an OD_{600} = 0.5, the cells were cooled down on ice for 20 min. The cells were harvested by centrifugation for 15 min at 1838 g at 4°C and the resulting cell pellet was resuspended in 10 ml of TSS buffer (85 % LB, 10 % PEG 8000, 5 % DMSO, 50 mM $MgCl_2$). Finally, the cell suspension was aliquoted in 200 µl volumes and snap frozen in dry ice before storing at -80°C.

Table 2-1: Details of strains used in this project

Strain	Phenotype	Genotype	Reference
TOP10		F-, *mcrA,* Δ(*mrr-hsd*RMS-*mcr*BC), Φ80*lac*ZΔM15, Δ*lac*X74, *rec*A1, *ara*D139, Δ(*araleu*)7697, *gal*U, *gal*K, *rps*L, (StrR), *end*A1, *nup*G	One Shot™ TOP10 Chemically Competent *E. coli*, Life Technologies
W3110	Wild-type	F-, lambda-, IN(*rrnD-rrnE*)1, *rph*-1	ATCC 27325
MC4100	Wild-type	F-, [*araD*139]B/r, Del(*argF-lac*)169, lambda-, e14-, *flhD*5301, Δ(*fruK-yeiR*)725(*fruA*25), *relA*1, *rpsL*150(*strR*), *rbsR*22, Del(*fimB-fimE*)632(::IS1), *deoC*1	CGSC 6152
MC4100	Tat-null	F-, [*araD*139]B/r, Del(*argF-lac*)169, lambda-, e14-, *flhD*5301, Δ(*fruK-yeiR*)725(*fruA*25), *relA*1, *rpsL*150(*strR*), *rbsR*22, Del(*fimB-fimE*)632(::IS1), *deoC*1, Δ*tatABCDE*	(Wexler et al., 2000)
SHuffle		F-, *lac*, *pro*, *lacI^q* / *Δ(ara-leu)7697, araD139, fhuA2, lacZ::T7, gene1, Δ(phoA)PvuII, phoR, ahpC*, galE, (or U), galK, λatt*::pNEB3-r1-*cDsbC*(Spec^R, *lacI^q*), *ΔtrxB, rpsL150*(Str^R), *Δgor, Δ(malF)3*	SHuffle® T7 Competent *E. coli*, New England BioLabs

2.2. Molecular biology

2.2.1. Transformation of chemically competent bacteria by heat shock

Bacteria were transformed using the heat shock method described (Sambrook and Russel, 2001). Chemically competent cells were mixed with 100 ng of plasmid DNA and incubated for 10 min on ice. After a heat shock of 45 sec at 42°C, the cells were incubated for 5 min on ice. Next, 350 µl of recovery SOC medium (Life Technologies) was added to the tube which was then incubated for 1 h at 37°C with shaking at 220 rpm. A volume of 100 µl of the suspension was plated on LB agar plates supplemented with the appropriate antibiotic and incubated for 16 h at 37°C.

2.2.2. Plasmid amplification

Plasmid transformants were grown in 50 ml of LB with appropriate antibiotic for 16 h at 37°C with shaking at 220 rpm. Plasmids were purified from the cells using Plasmid Plus Midi kits (QIAGEN) according to the manufacturer's instructions. The DNA was recovered in 100 µl of the supplied EB buffer (10 mM Tris-Cl, pH 8.5). DNA concentrations were measured using a NanoDrop™ 2000 (Life Technologies) and samples were stored at -20°C.

2.2.3. Amplification by PCR

All primers were chemically synthesised by Sigma-Aldrich. Polymerisation chain reactions (PCR) were performed by mixing 4 µl of 5X Herculase II reaction buffer (Agilent), 0.2 µl of dNTPs (25 mM mixture of dATP, dGTP, dTTP, dCTP) (Life Technologies), 1.25 µl of each primer (10 mM), 100 ng of DNA template, 1 % DMSO (final), 0.5 µl of Herculase II (Agilent) and adjusted to 50 µl with UltraPure™ water (Ambion). The reactions were carried out in a thermocycler T3000 (Biometra®). An example of amplification reaction is given below:

Initial denaturation	95°C	2 min	
Denaturation	95°C	20 sec	35 cycles
Annealing	Primer T_M minus 5°C	20 sec	
Extension	72°C	30 sec per kbp	
Final extension	72°C	5 min	

2.2.4. PCR purification

Amplified DNA fragments from PCR were purified using the QIAquick PCR purification kit (QIAGEN) according to the manufacturer's instructions. The DNA was recovered in 40 µl of supplied EB buffer. DNA concentrations were measured using a NanoDrop™ 2000 (Life Technologies) and samples were stored at -20°C.

2.2.5. DNA digestion with restriction enzymes

Typically, 1 µg of plasmid DNA was digested with 20 units of restriction enzyme (New England Biolabs®), 1X CutSmart® buffer (New England Biolabs®) and adjusted to 30 µl with UltraPure™ water (Ambion). The reaction was incubated for 1 h at 37°C. The DNA was analysed by agarose gel electrophoresis (section 2.2.6).

2.2.6. Agarose gel electrophoresis

DNA samples were separated by electrophoresis on agarose gels. Typically, 1 % agarose (w/v) (Life Technologies) was added to 100 ml of TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.2) and dissolved by heating in a microwave with occasional swirling. The solution was then cooled to ≈ 45-60^{o}C and ethidium bromide (Sigma-Aldrich) was added to a final concentration of 0.5 µg/ml. The solution was poured into a sealed gel tray (Bio-Rad) containing a comb and once set, the seals and comb were removed. The gel was submerged in TAE buffer in the electrophoresis tank. The DNA sample was mixed with Gel Loading Dye, Purple (6X) (New England Biolabs®) before loading on the gel alongside 100 bp and 1 kpb DNA Ladders (Bio-Rad). Electrophoresis was then carried out at 100 V for 60 min. Gels were visualised and digitalized by the imaging system E-Gel Imager (Life Technologies). DNA sizes were estimated by comparison to the 100 bp and 1 kbp DNA Ladders (Bio-Rad).

2.2.7. Agarose gel extraction

To extract DNA fragments from agarose gels, the appropriate bands were first visualised on a UV transilluminator (Spectroline) at wavelength 365 nm and the relevant bands were excised from the

gel using a scalpel. The DNA was extracted from the gel slices using QIAquick Gel Extraction kits (QIAGEN) according to the manufacturer's instructions. The DNA was recovered in 40 µl of the supplied EB buffer. DNA concentrations were measured using a NanoDrop™ 2000 (Life Technologies) and samples were stored at -20°C.

2.2.8. Ligation

Typically, the restriction-digested backbone vector was mixed with the restriction-digested insert to a molar ratio of 1:3 into Quick Ligase Reaction Buffer (New England Biolabs®). After addition of 1 µl of Quick Ligase (New England Biolabs®), the solution was incubated at RT for 5 min. A 2 µl volume of the mix was then used to transform chemically competent TOP10 cells (section 2.2.1).

2.2.9. Directed mutagenesis

The Q5® Site-Directed Mutagenesis kit (New England Biolabs®) was used to introduce nucleotide substitutions, deletions or insertions into plasmid DNA according to the manufacturer's instructions. Primers were designed according to the manufacturer's recommendations and chemically synthesised by Sigma-Aldrich. Typically, 25 ng of plasmid DNA was mixed with 0.5 µM of each primer and 12.5 µl of Q5® Hot Start High-Fidelity 2X Master Mix. The final volume was adjusted to 25 µl with UltraPure™ water (Ambion). The reaction was carried out in a thermocycler T3000 (Biometra) as follows: 98°C for 30 sec, 30 cycles of 98°C for 20 sec and 72°C for 20 sec per kbp, 72°C for 2 min. Then, 1 µl of the reaction was mixed with 5 µl of KLD reaction buffer and 1 µl of KLD enzyme mix. The final volume was adjusted to 10 µl with UltraPure™ water (Ambion). The solution was incubated 5 min at RT. Chemically competent TOP10 cells were transformed with 5 µl of the reaction mix (section 2.2.1).

2.2.10. Type IIS cloning strategy

2.2.10.1. Design of backbone vector

The pTTO expression vector pDPH340 (Ellis et al., 2017) was used to clone all the gene inserts expressed in this study unless otherwise stated. This plasmid contains the low copy number origin of replication p15A, a tetracycline resistance gene and a *lacI* gene to repress the *tac* promoter

controlling the expression of the gene(s) of interest. A stuffer fragment containing the *lacZα* gene, flanked by the TypeIIS restriction site BsaI, to enable in-frame gene cloning using Golden Gate assembly (Engler et al., 2008), was designed *in silico,* in the format of [BsaI-lacZ *α*-Stop codon-BsaI]. This cassette was chemically synthesised (ATUM, Inc) and inserted into the MCS, under the control of an upstream *tac* promoter (Malherbe et al., 2019).

2.2.10.2. Design of inserts

Typically, the gene of interest, flanked by *BsaI* restriction sites compatible with the backbone vector was designed *in silico* using *E. coli* W3110 codon usage (Neidhardt et al., 1996) with a 5' RBS and proprietary spacing before the ATG start site. The inserts were also designed to contain a second RBS at the 3' end with a Type IIS site SapI to enable in-frame insertion of additional downstream gene(s) for polycistronic expression, in the format of [BsaI-RBS+spacer^{1}-Gene^{1}-stop codon-RBS+spacer^{2}-SapI-SapI-BsaI]. All gene inserts were chemically synthesised (ATUM, Inc, section 2.2.10.1).

For polycistronic expression, the downstream gene(s) of interest, flanked by *SapI* sites compatible with the upstream gene and vector sites, was designed *in silico* as [SapI-gene^{2}-stop codon-SapI] and chemically synthesised (ATUM, Inc). In the case of cloning multiple downstream genes, an RBS with the spacer was introduced upstream of the ATG start site of each additional gene, terminating with a 3' stop codon, as [SapI-gene^{2}-stop codon-RBS+spacer^{3}-gene^{3}-Stop codon-SapI].

2.2.10.3. Type IIS cloning reaction

The TypeIIS cloning reaction was adapted from Golden Gate Assembly (Engler et al., 2008) and optimised to give >99 % efficiency in-house (E. Dave, personal communication). Briefly, the chemically synthesised gene insert (ATUM, Inc) was received as a lyophilised pellet and re-dissolved in 50 µl of UltraPure™ water (Ambion). The first step of assembly consisted of mixing 2 µl of the gene insert (about 20 ng) with 20 ng of the TypeIIS-enabled backbone vector, 3 µl of 10X CutSmart® buffer (New England Biolabs®), 0.5 mM ATP (Life Technologies), 10 units of *BsaI*-HF (New England Biolabs®) and 200 units of T4 DNA ligase (New England Biolabs®). The solution was adjusted to a final volume

of 30 µl with UltraPure™ water (Ambion) and the reaction was thermocycled (thermocycler T3000 (Biometra®) for 30 cycles of 1 min at 37°C and 1 min at 16°C followed by 20 min at 85°C. The second step involved adding 0.5 µl of 10X CutSmart® buffer (New England Biolabs®), 10 units of *BsaI*-HF (New England Biolabs®) and 4 µl of UltraPure™ water (Ambion) to the reaction mix to linearize any DNA containing the TypeIIS restriction site including empty vector. The reaction was incubated for 20 min at 37°C then 20 min at 85°C in a thermocycler T3000 (Biometra®). Chemically competent TOP10 cells were transformed with 2 µl of the final reaction (section 2.2.1).

To clone additional downstream genes for polycistronic expression, the newly constructed plasmid as given above was used as the accepting vector. Subsequently, the above reactions were performed using *SapI* (New England Biolabs®) enzyme instead of *BsaI*-HF.

2.2.11. Plasmid DNA sequencing

DNA sequencing was carried out using Sanger sequencing (Sambrook and Russel, 2001) either with UCB Celltech's internal service or externally with Macrogen services using the provided oligos. Oligos to span the genes(s) of interest and *LacI* in the forward and reverse directions were designed *in silico* and chemically synthesized by Sigma-Aldrich. For internal sequencing, 0.5 µg of DNA sample was prepared using 1 µl of primers at 3.2 µM and the BigDye® Direct Sanger Sequencing Kit (Life Technologies) according the manufacturer's instructions. The solution was thermocycled in (thermocycler T3000 (Biometra®) for 30 sec at 96°C and 35 cycles of 10 sec at 96°C, 5 sec at 50°C and 4 min at 60°C. The completed reactions were ethanol precipitated with 1/3 volume of NaOAc and 3 volumes of 95 % ethanol and incubated for 30 min at RT. After a 45 min centrifugation at 4186 g, the supernatant was discarded, and the DNA pellets were washed with 70 % ethanol and air dried. DNA was then dissolved in 15 µl of formamide (Life Technologies) and denatured at 95°C for 2 min. Samples was sequenced by capillary electrophoresis on an ABI 3130xL sequencer (Life Technologies). Sequencing data was obtained by ABI Sequence Analysis 5.2 software (Life Technologies) and processed by Sequencher 5.4.5 software (Gene Codes). The sequence-validated plasmids used in this study are listed in Table 2-2.

Table 2-2: List of plasmids used in this study in order of appearance.

Name	Vector	Insert	Reference
pDPH340	pTTO		(Ellis et al., 2017)
pGM2	pTTO	*lacZ*α	This study, (Malherbe et al., 2019)
pGM36	pTTO	phoA-His	This study
pGM44	pTTO	phoA-His + CyDisCo	This study
pGM37	pTTO	$ompA_{SP}$-phoA-His	This study
pGM45	pTTO	$ompA_{SP}$-phoA-His + CyDisCo	This study
pGM39	pTTO	$ycbK_{SP}$-phoA-His	This study
pGM47	pTTO	$ycbK_{SP}$-phoA-His + CyDisCo	This study
pGM100	pTTO	$torA_{SP}$-phoA-His + CyDisCo	This study
pGM74	pTTO	Empty vector	This study
pGM88	pTTO	$paoA_{SP}$-phoA-His + CyDisCo	This study
pGM13	pTTO	$amiA_{SP}$-amiA-His	This study
pGM14	pTTO	$hyaA_{SP}$-hyaA-His	This study
pGM15	pTTO	$paoA_{SP}$-paoA-His	This study
pGM16	pTTO	$torA_{SP}$-torA-His	This study
pGM17	pTTO	$ycbK_{SP}$-ycbK-His	This study
pGM21	pTTO	paoA-His	This study
pGM26	pTTO	$ompA_{SP}$-paoA-His	This study
pGM101	pTTO	$torA_{SP}$-paoA-His	This study
pGM102	pD874	paoB-V5 + paoC-V5	This study
pGM103	pD874	paoB-V5 + paoC-V5 + paoD-V5	This study
pGM113	pTTO	paoASP-paoA-His + paoB-v5 + paoC-v5 + paoD-HA	This study
pGM29	pTTO	lacZ-His	This study
pGM30	pTTO	$amiA_{SP}$-lacZ-His	This study
pGM31	pTTO	$hyaA_{SP}$-lacZ-His	This study
pGM32	pTTO	$paoA_{SP}$-lacZ-His	This study
pGM33	pTTO	$torA_{SP}$-lacZ-His	This study
pGM34	pTTO	$ycbK_{SP}$-lacZ-His	This study
pGM35	pTTO	$ompA_{SP}$-lacZ-His	This study
pGM73	pTTO	CyDisCo	This study
pGM87	pTTO	$paoA_{SP}$-phoA-His	This study
pGM99	pTTO	$torA_{SP}$-phoA-His	This study
pGM118	pTTO	sfGFP-His	This study
pGM119	pTTO	$ompA_{SP}$-sfGFP-His	This study
pGM120	pTTO	$paoA_{SP}$-sfGFP-His	This study
pGM121	pTTO	$torA_{SP}$-sfGFP-His	This study

Name	Vector	Insert	Reference
pGM198	pTTO	hGH-His	This study
pGM199	pTTO	hGH-His + CyDisCo	This study
pGM196	pTTO	$ompA_{SP}$-hGH-His	This study
pGM197	pTTO	$ompA_{SP}$-hGH-His + CyDisCo	This study
pGM267	pTTO	$torA_{SP}$-hGH-His	This study
pGM268	pTTO	$torA_{SP}$-hGH-His + CyDisCo	This study
pGM159	pTTO	hGH-FLAG-phoA-His	This study
pGM160	pTTO	hGH-FLAG-phoA-His + CyDisCo	This study
pGM146	pTTO	$ompA_{SP}$-hGH-FLAG-phoA-His	This study
pGM147	pTTO	$ompA_{SP}$-hGH-FLAG-phoA-His + CyDisCo	This study
pGM122	pTTO	$torA_{SP}$-hGH-FLAG-phoA-His	This study
pGM123	pTTO	$torA_{SP}$-hGH-FLAG-phoA-His + CyDisCo	This study
pGM265	pTTO	phoA-FLAG-hGH-His	This study
pGM266	pTTO	phoA-FLAG-hGH-His + CyDisCo	This study
pGM252	pTTO	$ompA_{SP}$-phoA-FLAG-hGH-His	This study
pGM253	pTTO	$ompA_{SP}$-phoA-FLAG-hGH-His + CyDisCo	This study
pGM200	pTTO	$torA_{SP}$-phoA-FLAG-hGH-His	This study
pGM201	pTTO	$torA_{SP}$-phoA-FLAG-hGH-His + CyDisCo	This study
pGM223	pTTO	aP2-His	This study
pGM269	pTTO	aP2-His + CyDisCo	This study
pGM234	pTTO	$ompA_{SP}$-aP2-His	This study
pGM270	pTTO	$ompA_{SP}$-aP2-His + CyDisCo	This study
pGM141	pTTO	$torA_{SP}$-aP2-His	This study
pGM271	pTTO	$torA_{SP}$-aP2-His + CyDisCo	This study
pGM167	pTTO	hGH-FLAG-aP2-His	This study
pGM272	pTTO	hGH-FLAG-aP2-His + CyDisCo	This study
pGM154	pTTO	$ompA_{SP}$-hGH-FLAG-aP2-His	This study
pGM273	pTTO	$ompA_{SP}$-hGH-FLAG-aP2-His + CyDisCo	This study
pGM130	pTTO	$torA_{SP}$-hGH-FLAG-aP2-His	This study
pGM274	pTTO	$torA_{SP}$-hGH-FLAG-aP2-His + CyDisCo	This study
pGM245	pTTO	phoA-FLAG-aP2-His	This study
pGM275	pTTO	phoA-FLAG-aP2-His + CyDisCo	This study
pGM258	pTTO	$ompA_{SP}$-phoA-FLAG-aP2-His	This study
pGM276	pTTO	$ompA_{SP}$-phoA-FLAG-aP2-His + CyDisCo	This study
pGM208	pTTO	$torA_{SP}$-phoA-FLAG-aP2-His	This study
pGM277	pTTO	$torA_{SP}$-phoA-FLAG-aP2-His + CyDisCo	This study
pGM226	pTTO	VHH1-His	This study
pGM227	pTTO	VHH1-His + CyDisCo	This study
pGM237	pTTO	$ompA_{SP}$-VHH1-His	This study

Name	Vector	Insert	Reference
pGM238	pTTO	$ompA_{SP}$-VHH1-His + CyDisCo	This study
pGM178	pTTO	$torA_{SP}$-VHH1-His	This study
pGM179	pTTO	$torA_{SP}$-VHH1-His + CyDisCo	This study
pGM190	pTTO	hGH-FLAG-VHH1-His	This study
pGM191	pTTO	hGH-FLAG-VHH1-His + CyDisCo	This study
pGM184	pTTO	$ompA_{SP}$-hGH-FLAG-VHH1-His	This study
pGM185	pTTO	$ompA_{SP}$-hGH-FLAG-VHH1-His + CyDisCo	This study
pGM172	pTTO	$torA_{SP}$-hGH-FLAG-VHH1-His	This study
pGM173	pTTO	$torA_{SP}$-hGH-FLAG-VHH1-His + CyDisCo	This study
pGM248	pTTO	phoA-FLAG-VHH1-His	This study
pGM249	pTTO	phoA-FLAG-VHH1-His + CyDisCo	This study
pGM261	pTTO	$ompA_{SP}$-phoA-FLAG-VHH1-His	This study
pGM262	pTTO	$ompA_{SP}$-phoA-FLAG-VHH1-His + CyDisCo	This study
pGM213	pTTO	$torA_{SP}$-phoA-FLAG-VHH1-His	This study
pGM214	pTTO	$torA_{SP}$-phoA-FLAG-VHH1-His + CyDisCo	This study
pGM228	pTTO	VHH2-His	This study
pGM229	pTTO	VHH2-His + CyDisCo	This study
pGM239	pTTO	$ompA_{SP}$-VHH2-His	This study
pGM240	pTTO	$ompA_{SP}$-VHH2-His + CyDisCo	This study
pGM180	pTTO	$torA_{SP}$-VHH2-His	This study
pGM181	pTTO	$torA_{SP}$-VHH2-His + CyDisCo	This study
pGM192	pTTO	hGH-FLAG-VHH2-His	This study
pGM193	pTTO	hGH-FLAG-VHH2-His + CyDisCo	This study
pGM186	pTTO	$ompA_{SP}$-hGH-FLAG-VHH2-His	This study
pGM187	pTTO	$ompA_{SP}$-hGH-FLAG-VHH2-His + CyDisCo	This study
pGM174	pTTO	$torA_{SP}$-hGH-FLAG-VHH2-His	This study
pGM175	pTTO	$torA_{SP}$-hGH-FLAG-VHH2-His + CyDisCo	This study
pGM250	pTTO	phoA-FLAG-VHH2-His	This study
pGM251	pTTO	phoA-FLAG-VHH2-His + CyDisCo	This study
pGM263	pTTO	$ompA_{SP}$-phoA-FLAG-VHH2-His	This study
pGM264	pTTO	$ompA_{SP}$-phoA-FLAG-VHH2-His + CyDisCo	This study
pGM215	pTTO	$torA_{SP}$-phoA-FLAG-VHH2-His	This study
pGM216	pTTO	$torA_{SP}$-phoA-FLAG-VHH2-His + CyDisCo	This study

2.2.12. Genomic DNA extraction

The extraction of genomic DNA from bacterial strains of interest were achieved using the QIAamp DNA Mini kit (QIAGEN) according to the manufacturer's instructions. Briefly, bacteria were grown

for 16 h at 37°C on LB agar. A single colony was used to inoculate 10 ml of LB medium in a 50 ml tube which was then incubated for 16 h at 37°C with shaking at 220 rpm. The DNA was then extracted using the kit mentioned above and eluted in 100 µl of supplied Buffer AE. DNA concentrations were measured using a Qubit 3.0 (Life Technologies) and samples were stored at -20°C.

2.2.13. Genomic DNA sequencing

Genomic sequencing was performed externally by GenomeScan using an Illumina HiSeq4000. The 150 bp paired-end reads obtained were mapped to the reference genome (NCBI, AP009048.1) using the Geneious 10.1.3 software.

2.2.14. Total RNA extraction

A preculture was set with single colony in 5 ml of LB and incubated overnight at 37°C with shaking at 220 rpm. A volume of 5 ml of fresh LB was inoculated with 50 µl of the preculture and incubated at 37°C with shaking at 220 rpm until mid-log phase (OD_{600} = 0.8). Then, total RNA was extracted from 0.5 ml of culture using the RNAprotect Bacteria Reagent (QIAGEN) and the RNeasy Protect Bacteria mini kit (QIAGEN) according to the manufacturer's instructions. The RNA was eluted in 40 µl of RNase-free water in a fresh 1.5 ml RNase-free tube. RNA concentrations were measured using a NanoDrop™ 2000 (Life Technologies) and samples were stored at -20°C.

2.2.15. Reverse transcription

The mRNA from the total RNA extracted was reverse-transcribed into cDNA using the SuperScript III kit (Life Technologies). In a RNase-free microcentrifuge tube, 200 ng of random primers (Life Technologies), 1 µg of total RNA, 1 µl of dNTP mix (at 10 mM each) and RNase-free water up to 13 µl were mixed. After a 5 min incubation at 65°C, the samples were cooled on ice for at least 1 min and the contents were collected by a brief centrifugation. Then, 4 µl of 5X First Strand Buffer (Life Technologies), 1 µl of DTT at 0.1 M, 1 µl of RNaseOUT Recombinant RNase Inhibitor (Life Technologies) and 1 µl of Superscript III Reverse-Transcriptase were added to the mixture. The solutions were mixed by gently pipetting up and down and incubated for 60 min at 55°C. After inactivating the enzyme at 70°C for 15 min, the cDNA was used immediately or stored at -20°C.

2.3. Protein expression

2.3.1. Culture media

Unless indicated, all liquid cultures were grown in Lennox medium (LB) and Lennox agar (LA) was used for growth on Petri dishes (Sambrook and Russel, 2001). The 2xTY medium was composed of LB medium containing twice as much yeast extract (Sambrook and Russel, 2001). The minimal medium M9 was composed of M9 salts (Gibco) supplemented with $MgSO_4$, $CaCl_2$ and glucose to working concentrations of 2 mM, 0.1 mM and 20 mM respectively. For protein expression, all media were supplemented with $MgSO_4$ (0.5 mM), $ZnSO_4$ (0.5 mM) and appropriate antibiotic: tetracycline (10 µg/ml) (FisherScientific) or kanamycin (15 µg/ml) (Sigma-Aldrich). When PaoA was expressed, the cofactors $FeCl_3$ (0.5 mM) and Na_2MoO_4 (1 mM) were also added to the culture media.

2.3.2. Expression culture

The culture conditions used are listed in Table 2-3. A preculture was set up by inoculating 15 ml of selective media with a single colony from a fresh overnight selective LA plate and incubated for 16 h at 37°C with shaking at 220 rpm. The preculture was used to inoculate 25 ml of selective fresh media at a starting OD_{600} = 0.05 in a 125 ml shake flask. The culture was incubated at the appropriate temperature as indicated in Table 2-3, with shaking at 220 rpm until the OD_{600} reached 0.5. Protein expression was induced by adding 100 µM IPTG (Sigma Aldrich) and the culture was incubated at the relevant temperature for a defined period, as indicated in Table 2-3. The feed, used in condition 7, was a 25X solution containing 1 M MOPS pH 7.2, 40 % glycerol, 0.5 % MgSO4 and 0.42 % $MgCl_2$ (C. Doyle, personal communication).

2.4. Fractionation methods

All solutions were freshly supplemented with 100 µM of the capping agent N-ethylmaleimide (NEM) to trap the thiol-disulphide status during sample processing due to air oxidization (Matos et al., 2014, Hua et al., 2005). Moreover, all buffers were supplemented with one cOmplete™ Protease Inhibitor

Cocktail (Roche) tablet per 50 ml to prevent protein degradation by proteases. From harvest onwards, the samples and solutions were always kept on ice and all centrifugation steps were carried out at 4°C. All liquid removal was carried away by careful pipetting without disturbing the cell pellet.

At harvest, a volume of $\frac{8}{\mathrm{OD}_{600}}$ of the culture was transferred to a 50 ml tube and centrifuged at 4816 g for 15 min. Where relevant, the supernatant was recovered in a 2 ml tube and stored at -20°C as the media fraction. The cell pellet was washed with 850 ml of PBS, transferred into a fresh 2 ml tube and centrifuged at 20,817 g for 3 min.

Table 2-3: Culture conditions

Condition	Pre-culture medium	Culture medium	Growth temperature	OD_{600}	Feed	Temperature PI	Harvest PI
1	LB	LB	30°C	0.5	-	30°C	2 h
2	LB	LB	30°C	0.5	-	30°C	4 h
3	LB	LB	37°C	0.5	-	37°C	2 h
4	LB	LB	30°C	0.5	-	18°C	16 h
5	2x TY	2x TY	30°C	0.5	-	30°C	2 h
6	2x TY	2x TY	37°C	3.5	Yes	18°C	16 h
7	M9	M9	30°C	2	-	18°C	16 h

2.4.1. Periplasmic extraction

Following periplasmic extraction, the pellet was stored at -20°C prior to further fractionation.

2.4.1.1. Cold-osmotic shock

The cold osmotic shock method was adapted from Manoil and Beckwith (1986). The washed cell pellet was resuspended in 900 µl of spheroplast buffer (0.1 M Tris pH 8.0, 500 mM sucrose, 0.5 mM EDTA pH8.0), incubated for 5 min, centrifuged 3 min at 20,817 g and the supernatant was discarded. The pellet was then resuspended in 400 µl of distilled water and incubated 15 sec before adding 20 µl of $MgSO_4$ at 20 mM. After a 3 min centrifugation at 20,817 g, the supernatant was transferred to a 1.5 ml tube and stored at -20°C as the periplasmic fraction.

2.4.1.2. Cold-osmotic shock with magnesium

The washed cell pellet was resuspended in 900 µl spheroplast buffer (0.1 M Tris pH 8.0, 500 mM sucrose, 0.5 mM EDTA pH8.0), incubated for 5 min, centrifuged 3 min at 20,817 g and the supernatant was discarded. The pellet was then resuspended in 400 µl of distilled water containing 1 mM of $MgCl_2$ and incubated 1 min. After a 3 min centrifugation at 20,817 g, the supernatant was transferred to a 1.5 ml tube and stored at -20°C as the periplasmic fraction.

2.4.1.3. EDTA/lysozyme/cold-osmotic shock

The EDTA/lysozyme/cold osmotic shock method was adapted from Jones et al. (2016). The washed cell pellet was resuspended in 500 µl Buffer 1 (100 mM Tris-Acetate pH 8.2, 500 mM sucrose, 5 mM EDTA pH 8.0). After addition of 500 µl of water then 40 µl of lysozyme at 2 mg/ml, the suspension was incubated for 5 min after which 20 µl of $MgSO_4$ at 1 M was added. The sample was centrifuged at 20,817 g for 2 min and the supernatant was transferred to a 1.5 ml tube and stored at -20°C as the periplasmic fraction.

2.4.2. Cell disruption

2.4.2.1. Freeze/thaw cycling

The method using freeze/thaw cycling was adapted from Manoil and Beckwith (1986). After periplasmic extraction, the cell pellet was resuspended in 900 µl of spheroplast buffer followed by the addition of 120 µl of fresh lysozyme at 2 mg/ml and 900 µl of water. The suspension was incubated for 5 min then centrifuged 30 min at 20,817 g. After discarding the supernatant, the pellet was resuspended in 1.2 ml of Tris pH 8.0 at 10 mM. For cell lysis, the suspension was rapidly freeze/thaw cycled four times in liquid nitrogen for 2 min and in a 37°C water bath for 3 min. Then 10 U of benzonase was added to the solution with 20 µl of $MgSO_4$ at 1 M and incubated for 10 min. After centrifuging for 20 min at 20,817 g, the supernatant was transferred to a 1.5 ml tube and stored at -20°C as the cytoplasmic fraction. The resulting pellet was resuspended with 1 ml of Tris pH 8.0 and the tube was stored at -20°C as the insoluble fraction.

2.4.2.2. Ultrasonication

The method for fractionation by ultrasonication was adapted from Jones et al. (2016). After periplasmic extraction, the cell pellet was resuspended in 1 ml of Buffer 2 (50 mM Tris-acetate pH 8.2, 250 mM sucrose, 10 mM $MgSO_4$, U/ml benzonase) and centrifuged for 5 min at 20,817 g. The supernatant was discarded, and the pellet was then resuspended in 750 µl of Buffer 3 (50 mM Tris-acetate pH 8.2, 2.5 mM EDTA pH 8.0). The suspension was sonicated on ice with 5 cycles of 10 sec at amplitude 8 µm and 10 sec of incubation (Soniprep 150, MSE) before ultracentrifugation for 30 min at 186,000 g (Beckman). The supernatant was transferred to a 1.5 ml tube and stored at -20°C as the cytoplasmic fraction. The pellet was resuspended in 1 ml of Buffer 3 and stored at -20°C as the insoluble fraction.

2.5. Protein analysis

2.5.1. Western-Blotting

Protein samples were separated by SDS-PAGE as follows: a Novex™ 4-20 % or 16 % Tris-Glycine Mini Gels, WedgeWell™ format, 15-well (Life Technologies) was placed into a gel tank (Life Technologies) and the tank was filled with 1X Tris-glycine SDS running buffer (Life Technologies). Loading materials were prepared by mixing 10 µl of samples with NuPAGE™ Sample Reducing Agent (10X) (Life Technologies) and NuPAGE™ LDS Sample Buffer (4X) (Life Technologies) prior to heating for 5 min at 98°C in a thermocycler T3000 (Biometra®). The samples, PageRuler Plus Prestained Protein Ladder (Life Technologies) and MagicMark™ (Life Technologies) molecular weight markers were loaded onto the gel and separated by electrophoresis at 225 V for 35 min. The proteins were dry transferred from the gel onto a PVDF membrane (Life Technologies) using an iBlot™ 2 (Life Technologies), according to the manufacturer's instructions. After blocking the membrane in blocking solution (PBS, 0.1 % Tween20, 5 % milk powder), the membrane was immunologically detected with a primary antibody (Table 2-4) diluted in blocking solution, washed twice with PBST (PBS, 0.1 % Tween20) and revealed with appropriate secondary antibody (Table 2-4). The membrane was washed twice in PBST, twice in PBS and twice in distilled water for 15 min each. Finally, protein bands were detected with Pierce™

ECL Plus Western Blotting Substrate (Life Technologies) according to the manufacturer's instructions and imaged with the ImageQuant LAS 4000 (GE Healthcare).

Table 2-4: List of antibodies used in this study

Antibody	Target	Clonality	Species	Conjugation	Dilution	Supplier	Reference
Primary	His tag	Polyclonal	Rabbit	HRP	1/2000	Universal biologicals	A190-114P
	FLAG tag	Monoclonal	Mouse	-	1/1000	Sigma	F1804
	cMyc tag	Monoclonal	Mouse	-	1/1000	Abcam	Ab32
	LacI	Monoclonal	Mouse	-	1/2000	Abcam	AB33832
	G6PD	Polyclonal	Rabbit	-	1/2000	Assaypro	33115-05111
	MBP	Polyclonal	Rabbit	-	1/5000	Abcam	AB21144
	DsbA	Monoclonal	Mouse	-	1/1000	Abcam	AB106061
	V5 tag	Monoclonal	Mouse	-	1/2000	Thermo	MA5-15253
	HA tag	Polyclonal	Rabbit	-	1/4000	Abcam	ab9110
Secondary	Rabbit IgG HC+LC	Polyclonal	Goat	HRP	1/2000	Universal biologicals	A120-101P
	Mouse IgG, Fcγ	Polyclonal	Goat	HRP	1/2500	Jackson ImmunoResearch	115-036-071

2.5.2. Protein activity assays

All samples and standards were measured in triplicate.

2.5.2.1. β-galactosidase activity assay

β-galactosidase activity was measured using ONPG substrate (Sigma-Aldrich) as recommended by the manufacturer. A titration was prepared in a 96 well plate using concentrations ranging from 1.25 to 1280 U/L of purified β-galactosidase standard (Sigma-Aldrich). The samples were diluted in PBS in the same plate before ONPG substrate (Sigma-Aldrich) was added, then the solutions were homogenized by a brief and vigorous shaking. The plate was incubated for 30 min at 37°C and the reaction was stopped by the addition of 50 µl Na_2CO_3 (3 M) to each well followed by brief and vigorous shaking. The amount of product formed was measured at A_{405} on a Synergy 2 plate reader (BioTek). A standard curve was fitted from which the activity in each fraction sample was calculated.

2.5.2.2. Alkaline phosphatase activity assay

The activity of PhoA was measured using pNPP substrate (Sigma-Aldrich) as recommended by the manufacturer. A titration was prepared in a 96 well plate using purified PhoA (Sigma-Aldrich) at concentrations ranging from 1 to 1024 U/L. In the same plate, 10 µl of the fractionated samples were diluted in 90 µl PBS before 100 µl of pNPP substrate was added followed by a brief and vigorous shaking. The plate was incubated for 30 min at 37°C and the reaction was stopped by the addition of 50 µl of 3 M NaOH to each well followed by another brief and vigorous shaking. The activity was measured at A405 on a Synergy 2 plate reader (BioTek). A standard curve was fitted from which the activity in each fraction sample was calculated.

2.5.2.3. Green fluorescent protein activity assay

The activity of sfGFP was estimated by reading the fluorescence on the different sample fractions. In a 96 well plate, 50 µl of sample were loaded per well and the relative fluorescence activity was read by a Synergy 2 plate reader (BioTek) with λ_{ex} = 485 nm and λ_{em} = 528 nm.

2.6. Microscopic observations

Bright field observations of bacteria were performed on live cells at the time of harvest. A 5 µl volume of culture were mounted between a microscopic slide and a cover slip. The cells were imaged using an EVOS™ FL imaging microscope (Invitrogen) with an oil-immersion 100 times objective.

The fluorescence imaging was made using the same protocol, but a green fluorescence filter was set (λ_{ex} = 470 nm/λ_{em} = 525 nm).

3. DESIGN STRATEGY TO IMPROVE TAT EXPORT AND ITS ASSESSMENT

3.1. The limited current use of the Tat pathway

Export of recombinant proteins through the Tat pathway have been shown successfully over the past decade. Indeed, GFP is an interesting example since it was used as a model protein in many studies. Sec export of GFP is possible by fusion to the signal peptide of MBP but the recombinant protein was not active in the periplasm (Feilmeier et al., 2000). However, GFP was successfully translocated via Tat and found to be active in the periplasm (Thomas et al., 2001). The highest export yields of GFP through the Tat pathway were obtained in fermenters when co-expressing the TatABC components to 1.1 g/L (Matos et al., 2012). These process optimisations have shown that the Tat system can be exploited to produce an active recombinant protein to high titres. Fisher et al. (2008) confirmed this conclusion by testing the export between Sec and Tat of several recombinant proteins. Comparable and improved secretion yields were obtained for MBP, PhoA, GFP and a scFv when transported through Tat instead of Sec. The PhoA, phytase AppA, hGH, interferon α2b, a scFv and a VH antibody fragments recombinant proteins were also reported to export via Tat in *E. coli* (Matos et al., 2014, Alanen et al., 2015). Overall, Tat can be used to export recombinant proteins in the periplasm of *E. coli*, but, since its discovery, less than a dozen of proteins of interest have been reported. Furthermore, the active state of these proteins, to confirm correct folding following export, has only been demonstrated for two of these proteins.

The reason of the low success presented so far may be due to the insolubility of the proteins of interest in the cytoplasm resulting in the rejection by the proofreading system. The use of a soluble partner to improve protein solubility and hence Tat export, forms the hypothesis of this thesis. Indeed, numerous publications justify how the fusion of a soluble partner improved the expression of proteins of interest both in the cytoplasm and exported via Sec (section 1.4.4.3.) (Nygren et al., 1994, Esposito and Chatterjee, 2006, Cheng et al., 2017).

3.2. Design of screening strategy

Based on the relative success of using a fusion partner strategy for Sec export as discussed in section 3.1., a similar approach will be used here where a NTS will be used as a 'carrier' to export a recombinant protein into the periplasm via Tat. A diagrammatic representation of the fusion protein is given in Figure 3-1.

Figure 3-1: Protein fusion strategy to evaluate Tat export

The fusion protein consists of a natural Tat substrate baring its N-terminal signal peptide (SP) fused to a His-tagged protein of interest via a linker consisting of a FLAG tag (FLAG), a TEV cleavage site (TEV) and a flexible linker (FL).

A N-terminal NTS bearing a signal peptide is fused to a C-terminally polyhistidine (His)-tagged protein of interest via a 34 amino acids long linker region consisting of a FLAG tag, TEV cleavage site and a flexible linker. The His (6xHis) tags will be used first to detect the presence of the respective individual NTS and protein of interest, and secondly to aid purification if needed.

Tat export of C-terminally His-tagged proteins has been previously reported, suggesting that the flexibility of this unfolded tail does not prevent Tat export (Matos et al., 2014, Alanen et al., 2015, Matos et al., 2012). The FLAG tag (DYKDHDGDYKDHDIDYKDDDDK) was also used for detection with the final aim to detect the NTS after separation of the two domains. This particularly long version of the FLAG tag has been shown to increase the detection sensitivity while still being hydrophilic (Terpe, 2003).

A TEV cleavage site was positioned between the two proteins to allow the possibility to remove the carrier protein, if necessary, following purification. The TEV cleavage motif (ENLYFQ↓G/S) is recognized very specifically by a protease from the Tobacco Etch Virus which cleaves at the position denoted by the arrow with high specificity, (Cesaratto et al., 2016).

The flexible linker consisting of five amino acid residues of GSSGG and the N-terminal elements of the linker were used to allow enough space and flexibility between the two folded protein structures. Indeed, the GSSGG linker is known to provide flexibility and maintain solubility through the polarity of serine residues (Chen et al., 2013).

3.3. Selection of natural Tat substrates as the carrier partner

Selecting an NTS as the soluble carrier protein for the fusion construct has dual interest: increasing the solubility of the protein of interest hence reducing the risk of forming inclusion bodies and improving Tat export. To date, there are 28 known *E. coli* NTS, of which 19 have been somewhat characterized experimentally (such as TorA and DmsA), whereas little is known of the others, apart from the signal peptide sequence and its specificity, the molecular weight of NTS and information gleaned from analysis of its genetic organization *in silico* (section 1.3.2.4., Appendix 1) (Tullman-Ercek et al., 2007, Mejean et al., 1994, Sambasivarao et al., 1990). The largest Tat-exported complex described was the FdnGH subcomplex of the formate dehydrogenase-N which is an NTS of 142 kDa (Sargent et al., 2002). Although there is no report defining the maximum protein size limit of the Tat translocation capability, the translocon is thought to have an export size limit (Berks et al., 2000). Therefore, a small NTS would be preferable as a soluble partner, as this would allow a much larger protein of interest to be used. Of the 28 NTS, 12 NTS are within the smallest size range (< 41 kDa), therefore NTS larger than this were eliminated from this panel (Appendix 1).

Another criterium for selection was grounded on whether the NTS with their natural signal peptides were obligates for Tat-export. Signal peptide-MBP reporter fusions in a Tat knock-out strain have shown that some Tat signal peptides (Tat_{SP}) led to promiscuous export into the periplasm via Sec (Tullman-Ercek et al., 2007). Subsequently NTS that failed to meet this specificity according to available published data, were eliminated. Of the 5 NTS remaining, 2 appear to have partners and are therefore not monomeric and the oligomeric state of 2 others are unknown. The potential of expressing additional partners is replete with issues such as complicated vector design and larger

plasmid size, increased carrier protein size, subsequent overexpression, folding and association of the carrier complex, and increased metabolic drain. Ideally the NTS candidate should be monomeric, yet of the remaining panel only AmiA (31.4 kDa) is reported to be monomeric (Table1); the oligomeric state of YcbK (20.4 kDa) although small, is not known and TorA (94.5 kDa), although monomeric, is larger in size. TorA has been used in many published studies and proved to be Tat-specific (Santini et al., 1998, Weiner et al., 1998) and its signal peptide ($TorA_{SP}$) is the most widely used in Tat export studies (section 3.1.); it may prove to be a useful positive control. To increase the candidate shortlist to five, a heterodimer HyaA (40.7 kDa) and a heterotrimer PaoA (24.3 kDa) subunits were selected to investigate the potential for Tat export of these proteins (Table 3-1). Initially, these proteins will be expressed without their respective partners to gauge whether they alone can be used as a fusion partner.

Table 3-1: Selected NTS candidates

The panel of NTS was selected based on the criteria of small size, monomeric state and specific to Tat export. The table indicates the short protein name, molecular weight (kDa), the number of cysteine (#Cys), the number of disulphide bond (#DSB), the cofactor(s), the oligomeric state in which the said protein is transported via the Tat pathway and the required chaperone. N/D non-determined. N/A non- applicable.

Name	Size (kDa)	#Cys	#DSB	Cofactor	Chaperone	Oligomeric state during Tat export	complex size (kDa)	Reference
AmiA	31.4	0	0	Zn	GroEL	Monomer	N/A	(Ize et al., 2003)
YcbK	20.4	0	0	N/D	N/D	N/D	N/D	
TorA	94.5	10	0	Mo	TorD	Monomer	N/A	(Pommier et al., 1998)
HyaA	40.7	14	0	$[2Fe\text{-}2S]_3$	HyaE	Putative heterodimer with HyaB	102	(Dubini and Sargent, 2003)
PaoA	24.3	9	0	$[2Fe\text{-}2S]_2$	PaoD	Heterotrimer with PaoB and PaoC	135	(Neumann et al., 2009)

AmiA is a periplasmic amidase and functions as a cell wall hydrolase (Ize et al., 2003). This protein is non-essential, although the *amiA* gene deletion results in a morphology defect where cells form

growing chains and highlights incomplete septum cleavage during cell division (Heidrich et al., 2001). The *amiA* gene is the first gene of the *amiA-hemF* operon (Troup et al., 1994). The HemF protein is the oxygen-dependent coproporphyrinogen-III oxidase, an enzyme involved in heme biosynthesis in aerobic metabolism. The *ypeA* gene is located upstream from *amiA* and is transcribed in the opposite direction meaning that this operon does not contain more upstream genes.

YcbK is an uncharacterised protein where the signal peptide has been experimentally tested as an MBP fusion for Tat export (Tullman-Ercek et al., 2007). YcbK is encoded as a lone gene and is flanked by the downstream gene *gloC* encoding a cytoplasmic hydroxyacylglutathione hydrolase, presumably involved in methyglyoxal detoxification whereas the upstream gene *ycbB*, encodes for a transpeptidase involved in cell wall biosynthesis (Magnet et al., 2008).

TorA is the trimethylamine N-oxide (TMAO) reductase involved in the electron chain transport in anaerobia in *E. coli* by reducing the TMAO as the terminal electron acceptor (Mejean et al., 1994, Jourlin et al., 1996). This periplasmic enzyme receives electron from TorC, a pentahemic *c*-type cytochrome (Gon et al., 2001) and requires the cytoplasmic chaperone TorD for correct maturation (Pommier et al., 1998). The TorCAD proteins characterised as the active members of the trimethylamine N-oxide reductase system are encoded by the *torCAD* operon (Mejean et al., 1994). The *torT* and *torS* genes are located upstream from this operon and their respective proteins are responsible for the transcriptional regulation of *torCAD* (Jourlin et al., 1996).

HyaA (40.7 kDa) is the small subunit of the hydrogenase 1 complex which also contains the large subunit HyaB (66 kDa) (Forzi and Sawers, 2007). HyaA is an inner membrane bound protein facing the periplasm (Sawers and Boxer, 1986, Forzi and Sawers, 2007). It therefore contains a TM domain, suggesting that expression may lead to incorporation into the inner membrane on the periplasmic side. Although no direct evidence has been published, the 102 kDa hydrogenase 1 is believed to be Tat exported and uses the Tat_{SP} of HyaA to hitchhike the leaderless HyaB (Rodrigue et al., 1999). The *hyaA* gene belongs to the *hyaABCDEF* operon (Menon et al., 1990). The HyaCDEF proteins encoded by the *hyaCDEF* genes are essential chaperones for HyaAB synthesis and activity (Menon et al., 1991).

PaoA is an iron-sulphur binding subunit that forms a heterotrimer with PaoB (34 kDa) and PaoC (78 kDa) (Lee et al., 2014). The 135 kDa complex functions as a molybdenum-dependent aldehyde oxidoreductase and presumably plays a role in detoxification to prevent cell damage (Neumann et al., 2009). The subunits PaoB and PaoC are both devoid of export signals and are translocated into the periplasm in a piggyback fashion using the Tat_{SP} of PaoA (Lee et al., 2014). Both PaoB and PaoC partners are thought to be essential for Tat export. A cofactor protein PaoD (35 kDa) is also thought to be involved, improving complex stability and a requisite for its activity. The complex is genetically encoded by the *pao*ABCD operon (previously known as *yagTSRQ*) (Neumann et al., 2009).

3.4. Selection of proteins of interest

To validate the strategy, the selected NTS will initially be fused to reporter protein(s) as the protein of interest, as shown in Figure 3-1. An ideal reporter protein would be monomeric to avoid folding hindrance, selected with and without DSB(s) for choice flexibility and to test DSB formation in the cytoplasm, of small size so that the fusion protein does not breach the size threshold for Tat export, and have a rapid and measurable activity assay to quantify whether the reporter is folded correctly. The reporter proteins will be initially tested as N-terminal fusions to signal peptides to confirm their export via the Tat pathway prior to being used according to the fusion strategy. A valid reporter will be a protein that is exclusively exported by Tat and presents activity in the periplasm, thus confirming its correct folding (Chapter 6).

3.5. Evaluation of the success of this strategy

The success of the strategy was estimated by the detection of the cellular localisation of the protein of interest and reporter activity assays. The expressing cells were fractionated to separate the cellular compartments and the isolated fractions were separated by reducing SDS-PAGE. Gels were run in reducing conditions as many of the proteins of interest and reporter proteins possess DSBs. Analysis were confirmed by Western-blot and immunological detection of over-expressed proteins

via their respective tags using specific antibody reagents. A positive result was expected to be the localisation of the protein of interest in the periplasm when fused to a Tat_{SP}. In order to confirm the Tat specificity of the export, the protein of interest was expected to only localise in the periplasm of WT cells and be absent in the periplasm of Tat-null, a strain devoid of an effective Tat pathway (section 3.6.).

The NTSs were first be expressed with their own natural Tat_{SP} to select the valid candidates. Subsequently, the signal peptides of the latter were planned to be fused to the selected reporter proteins to test their export properties via the Tat pathway. However, to mitigate risk that this may not be successful, the signal peptide of TorA was also considered. Indeed, this signal peptide has been widely and successfully used in the literature for Tat export of recombinant proteins (Thomas et al., 2001, Matos et al., 2012, Matos et al., 2014, Alanen et al., 2015, Fisher et al., 2008). This leader sequence was proven to allow non-promiscuous export when fused to recombinant proteins (Tullman-Ercek et al., 2007, Berks et al., 2000).

To evaluate whether there is redundancy between the Sec and Tat pathways in relation to export of these proteins, a Sec-specific signal peptide was also used to test for Sec export efficiency. The latter pathway has the advantages of being a more efficient export pathway with a large capacity for export than the Tat pathway and therefore may be preferable. Amongst the discovered Sec substrates, the signal peptides of PhoA, pectate lyase B (PelB) or outer membrane protein A (OmpA) are of the most commonly used for the secretion of recombinant proteins (Low et al., 2013). In-house, the signal peptide of OmpA is preferred for Sec export based on data from periplasmic expression of antibody fragments (Ellis et al., 2017). Therefore, $OmpA_{SP}$ was selected to direct the proteins of interest into the periplasm via Sec. The potential to cytoplasmically express these proteins (without a signal peptide) will also be evaluated to understand whether these proteins have the natural ability to fold into a native and stable conformation in the cytoplasm, this being a prerequisite for Tat export.

3.6. Selection of host strain selection

The *E. coli* K12 W3110 (Table 2-1) strain was selected as the WT host, given that it is used routinely in-house for periplasmic expression of recombinant proteins. To assess the specificity of Sec and Tat export, engineered strains where the Sec or Tat pathway is deleted would be ideal. However, the Sec pathway is essential for bacterial growth and the temperature-permissive *sec* mutant strains offer a technically unreliable and conditional alternative (Ito et al., 1989). On the other hand, the Tat pathway is non-essential and a knock-out strain (Tat-null) has been successfully developed (Wexler et al., 2000). This mutant (Table 2-1) was constructed from *E. coli* K12 MC4100 by deletion of the *tatABCD* operon and *tatE* gene (Figure 1-5). Genetically, MC4100 and W3110 appear to have some differences, with MC4100 bearing ≈ 2.7 % gene deletions (119 genes) compared to MG1655, a W3110-related strain (Peters et al., 2003). From the 8 genetic differences noted between the genomes of MG1655 and W3110, only the mutation in *rpoS*, introducing a stop codon, change the gene number in W3110 (Hayashi et al., 2006). Without *rpoS*, this brings a total expected of 118 genes missing in MC4100 compared to W3110. Since W3110 is genetically different from the MC4100 Tat-null mutant, the MC4100 WT strain (Table 2-1) was also utilized for expression studies.

As no annotated genomes are available for MC4100 and to confirm authenticity of our strain stocks, DNA sequencing of W3110 WT, MC4100 WT and MC4100 Tat-null was conducted as detailed in section 2.2.13. The sequencing data was aligned to the theoretical W3110 reference genome (AP009048.1, NCBI). The differences highlighted are presented in Appendix 3 and first show no variances between the reference and our in-house W3110 WT strain, thereby confirming authenticity. Secondly, the sequenced MC4100 WT strain actually lacks 123 genes from seven loci compared to W3110 WT, including the *lac* and *paoABCD* operons which codes for the Tat substrate PaoA. This indicates that this genome is five genes shorter compared to published data by Peters et al. (2003). Two of the missing genes *yeiW* and *ykgN* (Hayashi et al., 2006) identified in this work were referenced after the publication from Peters et al. (2003). The three remaining discrepancies identified in this work are the *fnr*, *uspE* and *ynaJ* genes, all from the same genetic locus (Appendix 3, locus G). This suggests that these genes were either deleted by spontaneous mutation in our strain or the deletions were not detected by Peters and co-workers (2003). The *fnr* gene is worth

mentioning since it encodes for a global transcription factor essential for anaerobic respiratory processes (Spiro and Guest, 1990). Hence, the MC4100 WT strain was only used in aerobic conditions. Lastly, the MC4100 Tat-null strain is demonstrably identical to MC4100 WT, apart from the engineered deletion of the *tatABCDE* genes.

The *E. coli* Tat-null strain was confirmed to have a morphological phenotype as previously reported by Caldelari et al. (2008) (Figure 3-2). Cells cultured in LB were visualised with a 1000-fold magnification under oil immersion and shown to form long chains. This phenotype is due to the AmiA and AmiC Tat-substrates unable to translocate to the periplasm due to the absence of the Tat pathway (Bernhardt and De Boer, 2003, Ize et al., 2003). These proteins are cell-wall hydrolases involved in septum cleavage during cell division.

This Tat-null strain has also been reported to have other biological phenotypes including hypersensitivity to exogenous drugs and detergents, sensitive to lysis by lysozyme without EDTA, sensitive to copper, lower virulence, periplasmic leakiness and resistance to infection by P1 (Caldelari et al., 2008). It is thought that the lack of a functional Tat pathway leads to the mis-localisation of its substrates, thereby resulting in these phenotypes.

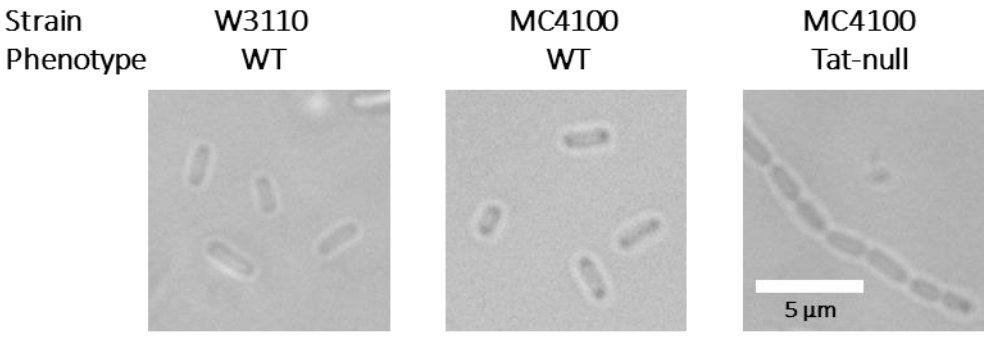


Figure 3-2: Morphological differences between *E. coli* WT and Tat mutant strains.

The images show bacteria from the strains W3110 WT, MC4100 WT and MC4100 Tat-null from left to right taken using an EVO™ FL imaging microscope. The pictures were taken with a 1000-fold magnification and under oil immersion. Scale bar is 5 µm.

4. DEVELOPMENT OF A ROBUST FRACTIONATION METHOD

4.1. Available methods to assess protein localisation

In bacteria, proteins localise at specific sites to form an elaborate subcellular architecture (Rudner and Losick, 2010). Gram negative bacteria such as *E. coli* have four protein locations: the outer membrane (OM), the periplasm, the inner membrane (IM) or the cytoplasm (Figure 1-1) (Costerton et al., 1974). Fractionation is an experimental process used to isolate proteins from their respective sub-cellular compartments. Moreover, the activity of the protein of interest can also be tested in the different isolated fractions. The risk with this technique is to have cross-fraction contamination which would result in artefactual data. Another method is to fuse GFP to the protein of interest, allowing direct observation by microscopy (Arigoni et al., 1995). This approach is limited because fusions may influence the localisation of the protein of interest through misfolding, aggregation or the GFP could be proteolytically clipped and therefore can no longer identify the localisation of the protein of interest. Moreover, the activity of the latter cannot be evaluated in its respective fractions using this approach. Hence, the fractionation method was considered for this project with a primary focus on fraction purity.

Although published fractionation protocols are not wholly different from each other, there appears to be no standardized method to extract all compartments with due-diligence against cross-contamination and inconsistency (Balasundaram et al., 2009). Sub-cellular fractionation begins with extracting the periplasm through the outer membrane. Then the resulting cells are lysed to release the cytoplasm which is separated from the insoluble fraction which includes membranes and potentially aggregated proteins by centrifugation. Further preparation to isolate the inner membrane from the outer membrane can also be performed (Thein et al., 2010) but this is out of the scope of this work. Periplasmic extraction is the most critical step of the process since disruption of the inner membrane can result in contamination from cytoplasmic proteins. Common methods to extract the periplasm use either an EDTA-lysozyme strategy (Thein et al., 2010, Pierce et al., 1997) or cold osmotic shock (Kang et al., 2012, Manoil and Beckwith, 1986), or a combination of both

(Alanen et al., 2015) (Figure 4-1). In the former method, EDTA is used to destabilise the outer membrane (Clifton et al., 2015), allowing the lysozyme to enter the periplasm and hydrolyse the peptidoglycan cell wall. This process subsequently releases the periplasm and leaves the cells as spheroplasts (Pierce et al., 1997). The second method generates an osmotic shock induced by successively adding a hypertonic solution to weaken the outer membrane, followed by a hypotonic solution to partially disrupt the outer membrane without compromising the integrity of the inner membrane (Kang et al., 2012).

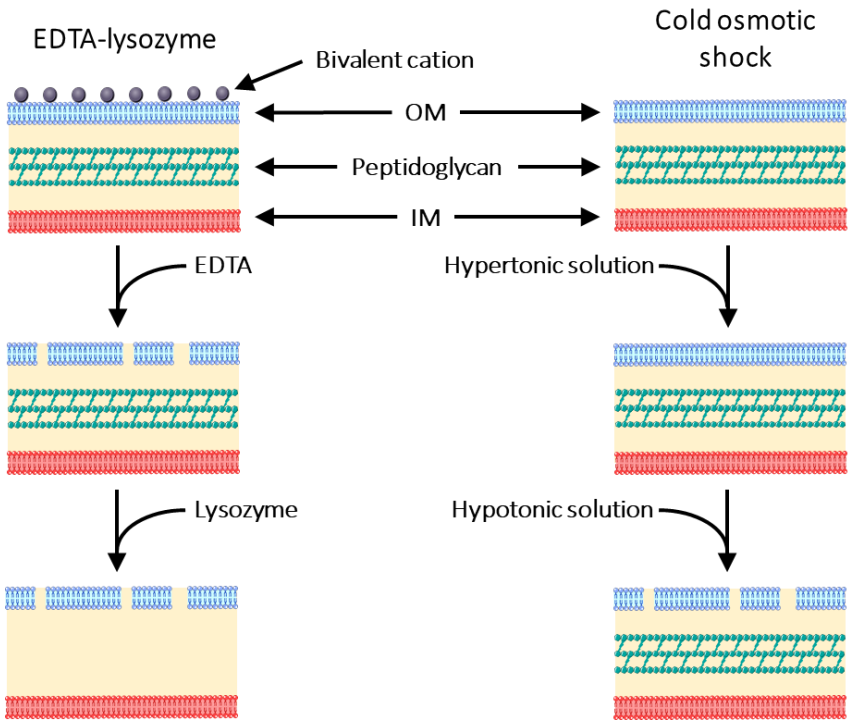


Figure 4-1: Diagram of the action mechanisms of the main periplasmic extraction procedures

On the left, the EDTA-lysozyme strategy aims to destabilise the outer membrane (OM) by chelating the bivalent cations such as Mg^{2+} and Ca^{2+}. The latter are necessary for the stability of the outer membrane and act as cofactors for enzymes involved in cell wall building. Once the outer membrane is fragilized, lysozyme can penetrate inside the periplasm and degrade the peptidoglycan. The periplasm is hence released while the inner membrane (IM) remains intact. On the right, the cold osmotic shock partially disrupts the outer membrane by the successive cell resuspension in hypertonic and hypotonic solutions.

The second and third steps in the fractionation process result in the recovery of the cytoplasmic and insoluble fractions simultaneously. This primarily involves cell lysis to disrupt the inner membrane and subsequent centrifugation to separate the cytoplasmic and insoluble fractions. The risks are poor

soluble protein recovery due to incomplete lysis or the use of harsh conditions resulting in loss of soluble protein through aggregation which in turn contaminates the insoluble fraction. The most common cell-lysis techniques used in small scale are freeze/thaw cycling and ultrasonication (Balasundaram et al., 2009, Thein et al., 2010). Freeze/thaw cycling causes the formation of ice crystals ultimately leading to cell disruption. However, loss of enzymatic activity have been reported (Lambert et al., 1983) as well as low yields (Stanbury and Whitaker, 1984). Ultrasonication uses the cavitation principle to disrupt the membranes by ultrasonic vibration creating local disruption of pressure in the form of cavitation bubbles which by collapsing release mechanical energy disintegrating the membranes (Harrison, 1991). The risks from this technique are the heat produced by the absorption of the acoustic energy which might denature proteins and the ionisation which was reported to inactivate enzymes (Lilly and Dunnill, 1969). Moreover, freeze/thaw cycling offers the advantage of treating several samples simultaneously whereas ultrasonication can treat only one sample at a time. Following cell lysis, the samples are centrifuged to isolate the insoluble debris. The pellet contains the membrane fragments and, when present in the cell, protein inclusion bodies.

To the best of our knowledge, no single study clearly describes in detail, a robust method to obtain uncontaminated soluble bacterial fractions. For instance, (Tao, 2008) and Thein et al. (2010) compared periplasmic fractions obtained by EDTA/lysozyme and cold osmotic shock and demonstrated cross-contamination in each case. Many publications show no due-diligence to fraction purity (Pechsrichuang et al., 2016, Alanen et al., 2015, Pierce et al., 1997, Kang et al., 2012, Matos et al., 2014). Some do utilize controls in the form of sub-compartment specific host proteins as purity markers, but these are incomplete. For instance, Zhang et al. (2017) merely controlled the fraction purity in only one of their experiments by Western-blot using an anti-β-lactamase antibody. β-lactamase is an *E. coli* natural periplasmic protein which, in this case, was found contaminating the cytoplasmic and outer membrane fractions. In another example, Fisher et al. (2008) used an anti-GroEL antibody for the same purpose. GroEL being a well-known cytoplasmic protein, was used to assess the purity of the periplasmic fractions by the lack of detectable GroEL proteins but in only one of their experiments.

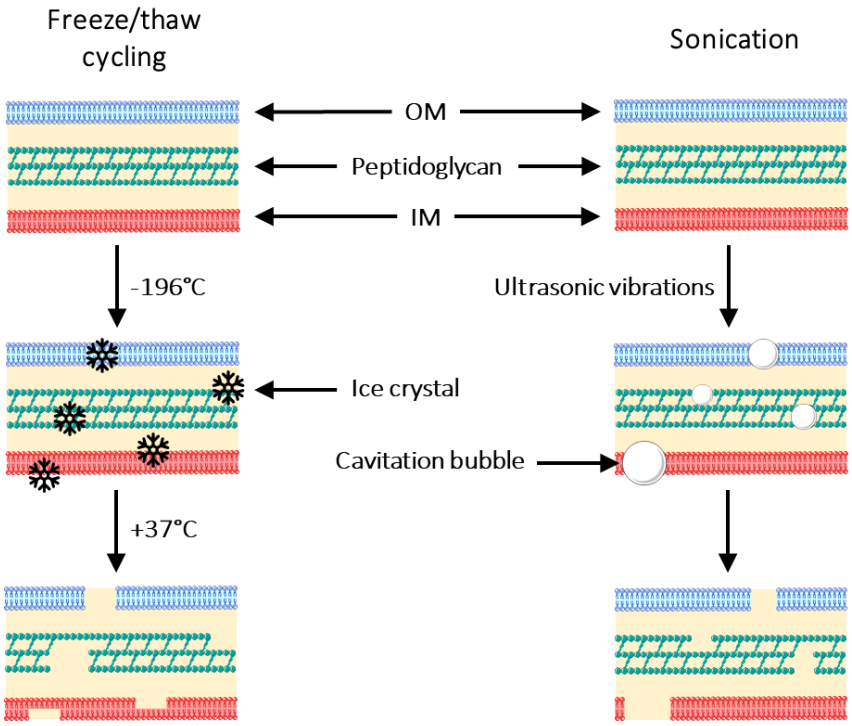


Figure 4-2: Diagram of the action mechanisms of the main cell lysis procedures

On the left, the freeze/thaw cycling strategy leads to creation of ice crystals which disrupt the cells. On the right, the sonication technique applies ultrasonication vibration to the cells creating cavitation bubbles. Their collapse releases mechanical energy disintegrating the membranes. OM: outer membrane, IM: inner membrane.

4.2. Discovery of cross-contamination issues

A clean fractionation method is absolutely necessary in order to detect expressed proteins in their actual respective subcellular compartments. This method must also be robust enough to handle WT strains as well as mutants with envelope debilities such as the control strain MC4100 Tat-null (section 3.6.). During the preliminary stages of this work, contamination was observed in the form of cytoplasmic proteins detected in the periplasmic fraction. His-tagged PhoA (PhoA-His) was expressed in W3110 WT and MC4100 Tat-null strains and targeted to either the cytoplasm (no signal peptide) or periplasm via Sec and Tat export, the latter with the use of specific signal peptides, $OmpA_{SP}$ for Sec or the natural signal peptide of the Tat substrate YcbK (Figure 4-3). As PhoA contains two DSBs, the CyDisCo components, Erv1p and PDI which enable cytoplasmic DSB formation (section 1.4.1.3) were also co-expressed with PhoA as cytoplasmic cMyc-tagged proteins. The cells were grown at

37°C and harvested after 16 h at 18°C post-induction (Table 2-3, condition 4). The fractionation protocol used was method 1 (Table 4-1) and comprised of cold osmotic shock for periplasm extraction and freeze/thaw cycling for cell lysis (Manoil and Beckwith, 1986). Cytoplasmically-expressed PhoA-His (with no signal peptide) was expected to be confined to the cytoplasm. However, as shown in Figure 4-3, significant levels of this protein were detected in both the cytoplasmic and periplasmic fractions when co-expressed with CyDisCo. Without CyDisCo, PhoA-His, expressed without a signal peptide, was not detected in any fraction, indicating that the CyDisCo components are essential for cytoplasmic expression. Similarly, the periplasmic-targeted proteins for either export pathway were observed in the periplasm when expressed in W3110 WT as predicted. Although the Sec-targeted PhoA-His was observed as expected in the periplasm of the MC4100 Tat-null strain which lacks the Tat export machinery, the Tat-targeted PhoA-His was unexpectedly detected in the periplasmic fraction of this strain. Furthermore, the cytoplasmic CyDisCo components, particularly cMyc-tagged PDI was significantly detected in all fractions irrespective of the strain. Although Erv1p-cmyc was present at lower levels in the periplasm, cross-contamination was still obvious. It is also worthwhile to point out that the CyDisCo proteins were detected in a substantial amount in the insoluble fractions irrespective of strain. As the presence of PhoA in the insoluble fractions was minimal, this is likely the result of high levels of overexpression of these proteins and subsequent formation of insoluble inclusion bodies. Although cross-contamination of the insoluble fraction by the presence of un-lysed cells may account for the minimal detection, the insoluble preparation could also capture jamming or periplasmic proteins ‘slowly’ translocating across the inner membrane. Overall, this experiment confirmed that this fractionation protocol needs to be optimised to obtain pure fractions. The contamination of the periplasmic fraction by cytoplasmic proteins revealed that the cold osmotic shock compromised the inner membrane resulting in cytoplasm leakiness. Also, the large amount of proteins detected in the insoluble fractions was a potential risk of artefact from un-lysed cells. Having no cross-contamination and a high recovery of soluble proteins are mandatory to drawing reliable conclusions regarding protein export.

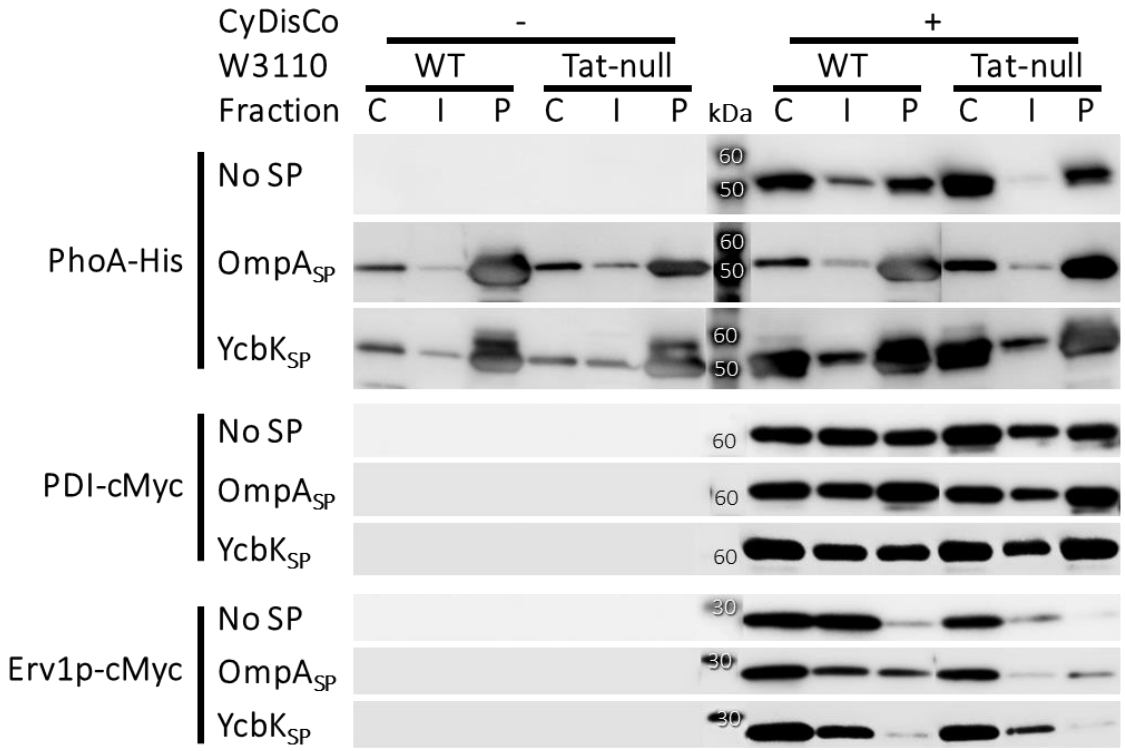


Figure 4-3: Cross-contamination issues using cold osmotic shock and freeze/thaw cycling

His-tagged PhoA (PhoA-His) with no signal peptide (No SP), Sec-targeted signal peptide (OmpA$_{SP}$) or Tat-targeted signal peptide (YcbK$_{SP}$) were expressed with (+) and without (-) cytoplasmic cMyc-tagged Erv1p and PDI in W3110 WT and MC4100 Tat-null strains. Cells were fractionated using method 1 (cold osmotic shock and freeze/thaw cycling) procedure allowing the separation of the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. Proteins were separated by SDS-PAGE and detected following Western-blot by anti-His and anti-cMyc antibodies where appropriate. Molecular weight markers are indicated in kilodalton (kDa). The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

4.3. Identification of a robust fractionation method

As shown in section 4.2, cell fractionation using a published method (Manoil and Beckwith, 1986) resulted in fraction cross-contamination which would have inadvertently led to false conclusions. As it is critically important for data integrity to identify methodologies that would yield clean fractions, several protocols from published literature were compared in parallel (Table 4-1). Method 1, as used in section 4.2., comprised of cold osmotic shock for periplasmic extraction and freeze/thaw cycling for cell lysis and resulted in cross-contamination. Since cold osmotic shock causes inner membrane leakiness, the addition of $MgCl_2$ to the hypotonic solution as suggested by el Yaagoubi et al. (1994) was considered (Table 4-1, method 2). $MgCl_2$ is thought to play a major role in maintaining the

integrity of the outer membrane (Clifton et al., 2015, Neidhardt et al., 1996). Another popular fractionation method utilizes EDTA/lysozyme/cold osmotic shock and ultrasonication for periplasmic and cytoplasmic extraction, respectively (Jones et al., 2016). As described in section 4.1., ultrasonication is not high-throughput unlike freeze/thaw cycling and therefore EDTA/lysozyme/cold osmotic shock was paired with freeze/thaw cycling into method 4 to improve the probability of identifying a robust method that is also high-throughput.

Table 4-1: Cell fractionation methods

Name	Periplasmic extraction	Cell disruption
Method 1	Cold osmotic shock	Freeze/thaw cycling
Method 2	Cold osmotic shock with $MgCl_2$	Freeze/thaw cycling
Method 3	EDTA/lysozyme/cold osmotic shock	Ultrasonication
Method 4	EDTA/lysozyme/cold osmotic shock	Freeze/thaw cycling
PureFrac	Cold osmotic shock with $MgCl_2$	Ultrasonication

To compare these four methods, cytoplasmically-expressed PhoA-His (without a signal peptide) was co-expressed with CyDisCo proteins cMyc-tagged Erv1p and PDI in W3110 WT and MC4100 Tat-null strains. Therefore PhoA-His, Erv1p-cMyc, PDI-cMyc as well as the LacI repressor – which is used to negatively regulate expression of the latter proteins on the expression plasmid – served as cytoplasmic markers. The endogenous cytoplasmic protein, glucose-6-phosphate dehydrogenase (G6PD) and periplasmic host proteins, maltose binding protein (MBP) and protein disulphide isomerase DsbA were used as native host markers to validate fraction purity (Dalbey and von Heijne, 2002, Bassford, 1990, Jonda et al., 1999). Protein localisation and cross-contamination was monitored by Western blot of overexpressed proteins and selected markers.

The transformants were grown as outlined in Table 2-3, condition 4 (growth at 37°C and harvest after 16h at 18°C post-induction) and the cells were fractionated as outlined in Table 4-1, using methods 1 to 4. As shown in Figure 4-4, the periplasmic fraction isolated by cold osmotic shock (method 1) presented minor bands with PDI-cMyc and LacI and, to a lesser extent, PhoA-His and

G6PD in both strains indicating some cross-contamination from the cytoplasm. The markers MBP and DsbA were located in the periplasmic fractions as anticipated, confirming correct localisation of the periplasmic proteins. However, the cytoplasmic contamination observed with method 1 was completely avoided by the inclusion of $MgCl_2$ during the cold osmotic shock (method 2) while the periplasmic markers were still extracted effectively (Figure 4-4) confirming the role of magnesium salts in maintaining outer membrane integrity (section 4.1.). Minor bands of periplasmic MBP were still detected in the cytoplasmic and insoluble fractions (Figure 4-4) presumably from intact cells escaping the periplasmic extraction process and will be addressed later (section 4.5). Despite this slight drawback, the periplasmic extraction by cold osmotic shock with $MgCl_2$ resulted in reproducible data where each protein was detected in its predicted fraction.

In contrast, when using EDTA/lysozyme/cold osmotic shock for periplasmic extraction (Table 4-1, methods 3 and 4), the periplasmic host proteins were not detected as expected. DsbA was not detected at all in any fraction which could be the result of degradation due to instability in the cytoplasmic environment (Figure 4-4). The second periplasmic marker protein, MBP, was detected in majority in the periplasm of the Tat-null cells with method 4 but more distributed in all fractions with method 3 indicating inconsistency of the method. Moreover, this marker was found with a large majority in the cytoplasmic fractions of the WT strain suggesting an ineffective periplasmic extraction (Figure 4-4). Furthermore, all cytoplasmic markers were unexpectedly detected in the periplasmic fractions of the MC4100 Tat-null strain suggesting that this method can be sub-optimal for at least this mutant strain. It is likely that the hypotonic conditions of the extraction method in addition to the effect of cation chelation by the EDTA on the inner membrane (Clifton et al., 2015), and peptidoglycan digestion by the lysozyme all contributed to weakening of the outer and inner membranes and subsequent disruption of the latter. Overall, due to cross-contamination, the EDTA/lysozyme/cold osmotic shock technique can lead to false-negative and false-positive conclusions and therefore is not fit for this purpose.

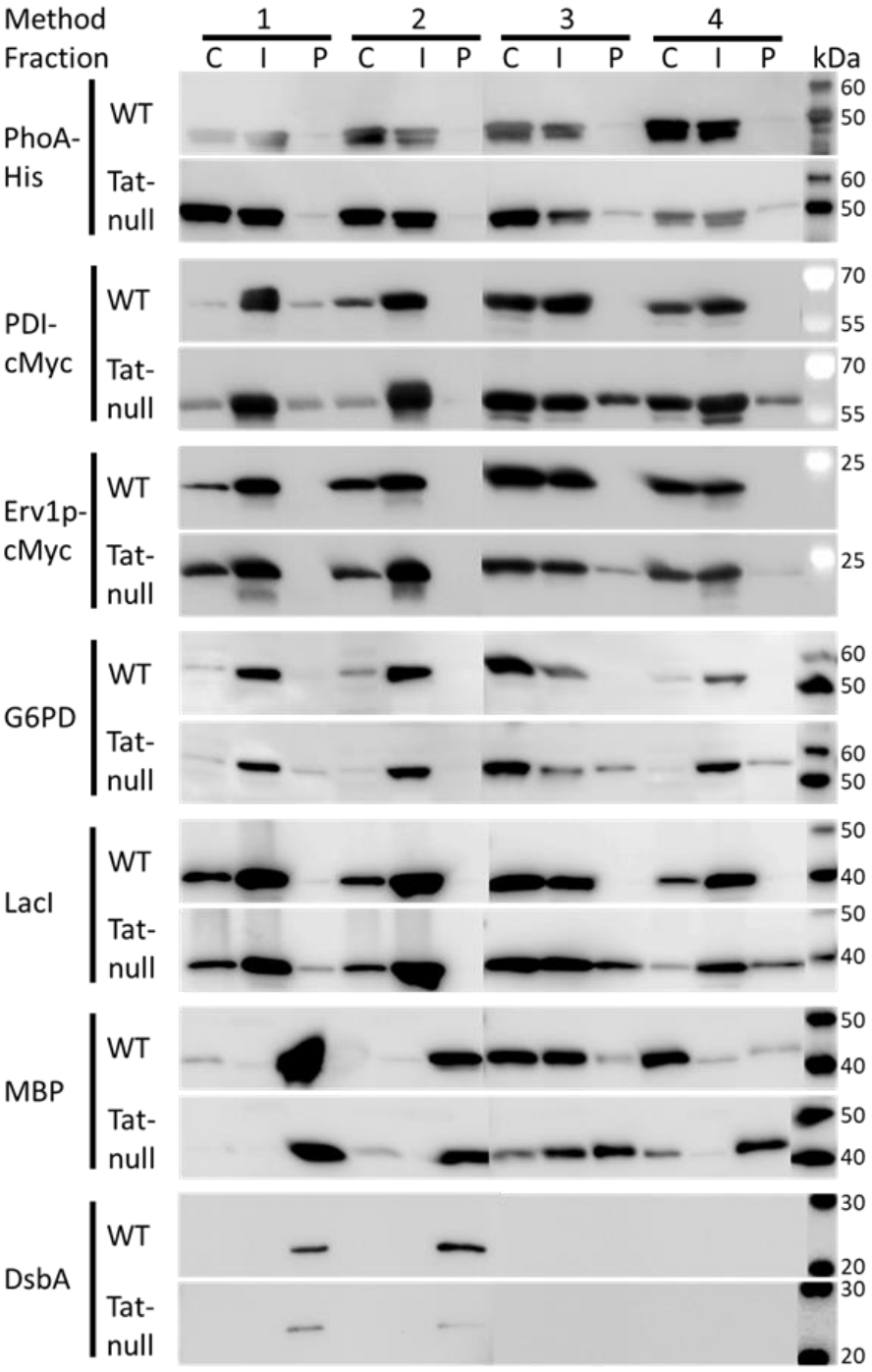


Figure 4-4: Evaluation of periplasmic and cytoplasmic extraction methods

Cytoplasmically expressed His-tagged PhoA, cMyc-tagged Erv1p and PDI proteins were co-expressed in W3110 WT and MC4100 Tat-null strains. The cells were fractionated according to method 1 (cold osmotic shock and freeze/thaw cycling), 2 (cold osmotic shock with MgCl2 and freeze/thaw cycling), 3 (EDTA/lysozyme/cold osmotic shock and sonication) and 4 (EDTA/lysozyme/cold osmotic shock and freeze/thaw cycling). The fractions are represented as cytoplasmic (C), insoluble (I) and periplasmic (P). Western-blots shown (from top to bottom) represent the detection of PhoA-His, PDI-cMyc, Erv1p-cMyc, G6PD, LacI, MBP and DsbA proteins as detected by anti-His, anti-cMyc, anti-G6PD, anti-LacI, anti-MBP and anti-DsbA antibodies respectively. PageRuler Plus Prestained Protein Ladder (Life Technologies) and MagicMark™ (Life Technologies) molecular weight markers are indicated in kilodalton (kDa). The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

For the isolation of cytoplasmic fractions, two methodologies, freeze/thaw cycling and ultrasonication, were compared. Generally, ultrasonication (method 3) appeared to be superior with notably higher recovery of soluble cytoplasmic proteins as indicated by PDI-cMyc, Erv1p-cMyc, G6PD and LacI (Figure 4-4). With the freeze/thaw cycling technique (method 1, 2 and 4), most of the cytoplasmic proteins except PhoA-His were unexpectedly located in the insoluble fraction irrespective of the strain. The inconsistent cytoplasmic proteins recovery from these three methods was most likely due to the periplasmic extraction technique used on the previous steps which left the cells in different states with inner and outer membrane disruption and variable peptidoglycan digestion. The proteins detected in the insoluble fractions likely came from un-lysed cells or carryover from the cytoplasmic fraction (unwashed insoluble pellet). The necessity of purity markers was highlighted here where LacI and MBP gave clear data whereas G6PD and DsbA were more difficult to interpret, if present at all. Hence, the cytoplasmic marker LacI and the periplasmic marker MBP will be used throughout the rest of the study.

In conclusion, cold osmotic shock with $MgCl_2$ appears to be the superior method for periplasmic extraction while ultrasonication gave optimal results for cytoplasmic recovery. These two techniques were subsequently combined in a single new method.

4.4. Validation of the composite method

The optimal method, consisting of cold osmotic shock with $MgCl_2$ and sonication, named PureFrac, is described in Table 4-1 and Appendix 4 (Malherbe et al., 2019). To assess this extraction procedure, the Tat-targeted PhoA-His (fused to a TorA signal peptide, $TorA_{SP}$-PhoA-His) was co-expressed with the cytoplasmic CyDisCo proteins, PDI-cMyc and Erv1p-cMyc in W3110 WT and MC4100 Tat-null strains. PhoA has previously been shown to export into the periplasm as an active protein via the Tat pathway when fused to a TorA signal peptide (Fisher et al., 2008, Matos et al., 2014). Cells were first cultivated under varying growth conditions (Table 2-3) commonly used to overexpress proteins in order to both determine the optimal conditions for Tat export and evaluate cross-contamination

propensity in relation to the environment (Pierce et al., 1997, Kang et al., 2012, Pechsrichuang et al., 2016, Matos et al., 2014, Ellis et al., 2017, Gaciarz et al., 2017, Roth et al., 2017)

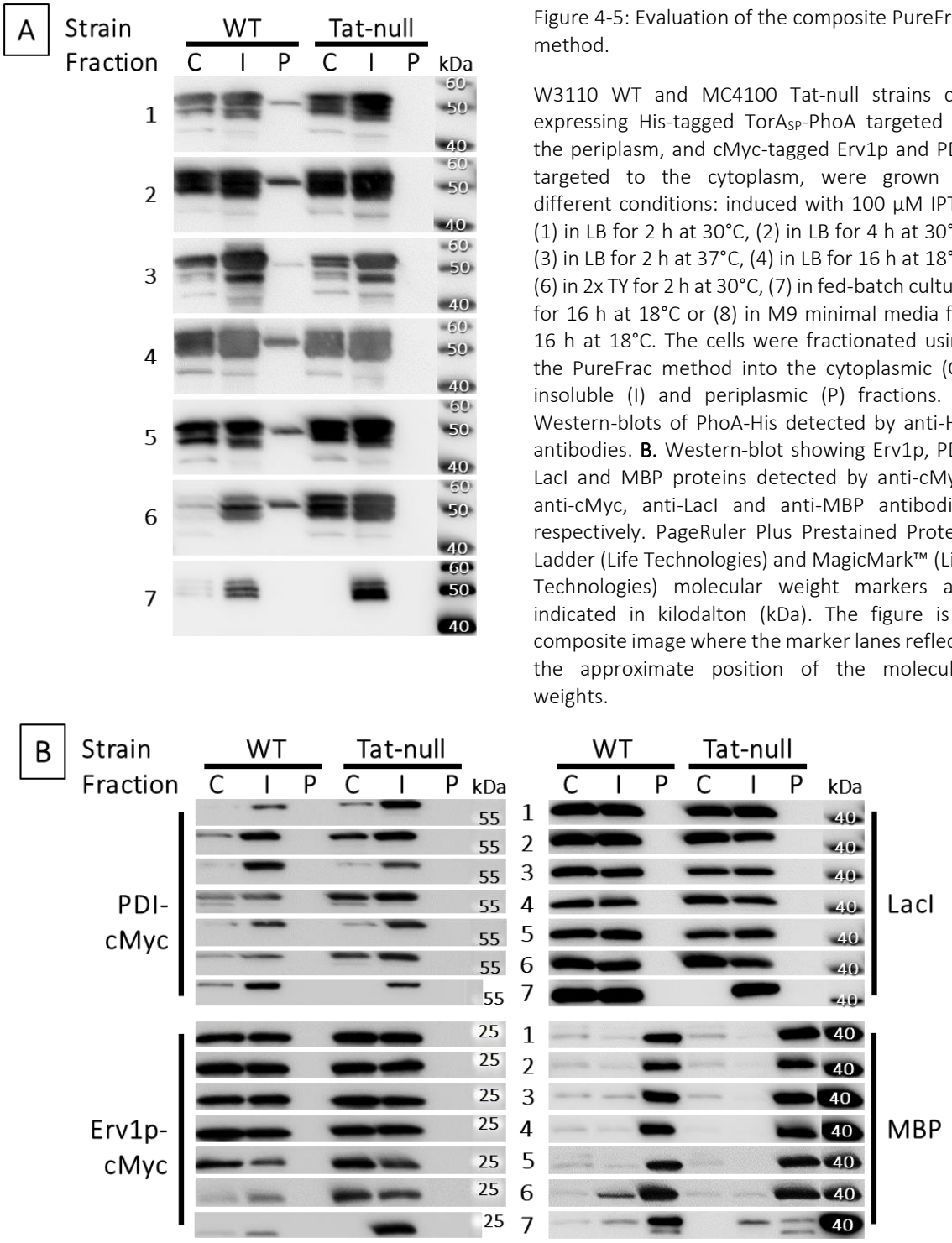


Figure 4-5: Evaluation of the composite PureFrac method.

W3110 WT and MC4100 Tat-null strains co-expressing His-tagged $TorA_{SP}$-PhoA targeted to the periplasm, and cMyc-tagged Erv1p and PDI, targeted to the cytoplasm, were grown in different conditions: induced with 100 µM IPTG (1) in LB for 2 h at 30°C, (2) in LB for 4 h at 30°C, (3) in LB for 2 h at 37°C, (4) in LB for 16 h at 18°C, (6) in 2x TY for 2 h at 30°C, (7) in fed-batch culture for 16 h at 18°C or (8) in M9 minimal media for 16 h at 18°C. The cells were fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. **A.** Western-blots of PhoA-His detected by anti-His antibodies. **B.** Western-blot showing Erv1p, PDI, LacI and MBP proteins detected by anti-cMyc, anti-cMyc, anti-LacI and anti-MBP antibodies respectively. PageRuler Plus Prestained Protein Ladder (Life Technologies) and MagicMark™ (Life Technologies) molecular weight markers are indicated in kilodalton (kDa). The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

As shown in Figure 4-5A, periplasmic PhoA-His was only observed in WT cells indicating Tat-exported protein as expected (Matos et al., 2014) but to varying degrees, this being dependent on the growth

parameters used. Growth condition 2 with a short induction of 4 h at 30°C in LB medium gave the optimal level of exported PhoA-His whereas condition 3 gave the lowest level and condition 8 no export at all (Table 2-3). The cytoplasmic markers PDI-cMyc, Erv1p-cMyc, LacI, and the periplasmic marker MBP were all observed in their predicted fractions of both strains (Figure 4-5B) further supporting the robustness of this method. The condition 2 was therefore selected as the expression conditions for the following studies.

4.5. Lysozyme addition to PureFrac is not a solution

The PureFrac process appeared to be superior to published methods as all control proteins were detected in their predicted fractions. However, this method could potentially be improved using lysozyme to enhance soluble protein recovery. Degrading the peptidoglycan using lysozyme could expand the release of the periplasmic proteins whilst the added magnesium could stabilize the outer membrane disrupted by the osmotic shock. Therefore, a comparison between the use of PureFrac with and without lysozyme was undertaken whilst exploring the Tat-mediated export of PhoA using the signal peptide of another Tat-specific substrate, aldehyde oxidoreductase iron-sulphur-binding subunit A ($PaoA_{SP}$).

$PaoA_{SP}$-PhoA-His was co-expressed with CyDisCo in W3110 WT and MC4100 Tat-null strains and fractionated using PureFrac with and without lysozyme in the hypotonic solution (100 µg/ml). Minor bands of PhoA-His were detected in the periplasmic fractions of the PureFrac method whereas a significant amount of the protein was located in the periplasm when using the lysozyme-complemented PureFrac method (Figure 4-6A). Without lysozyme treatment, the cytoplasmic marker proteins Erv1p-cMyc and LacI were exclusively observed in the cytoplasm whereas a small portion of PDI-cMyc was also found in the periplasm of WT cells. Conversely, significant levels of all these proteins were detected in the periplasm when extracted with lysozyme. This indicated that lysozyme addition to the periplasmic extraction buffer increased cross-contamination risk and the high level of $PaoA_{SP}$-PhoA-His observed in the periplasmic fraction with lysozyme was most likely a

cross-contamination artefact. The detection of this protein in the periplasm of the Tat-null strain further supported the artefactual data. In terms of PhoA activity, this protein appeared to be active in all fractions with close correlation with levels of PhoA found in each fraction (Figure 4-6B). In conclusion, presence of lysozyme appears to cause further cross-contamination and therefore should be avoided.

4.6. The importance of NEM to analyse the disulphide bond status

Some proteins need their DSBs to be formed to reach a native conformation and thus be active. However, the DSB status within a protein can change during sample processing because of air oxidation leading to artefacts and loss of activity (Elena et al., 1998). The use of very weakly alkaline solutions and the multiple pipetting/resuspension steps as required in the PureFrac protocol are optimal conditions for this process to occur. Therefore, preserving the DSB linkages of a protein is of paramount importance.

N-ethylmaleimide (NEM) is a capping agent used to block the free thiol groups of cysteine residues in proteins by reacting with sulfhydryl groups to form alkyl derivatives and thus preventing artefacts (Matos et al., 2014). To test the impact of NEM on DSB formation during extraction procedures, PhoA was expressed in MC4100 WT in fusion with $OmpA_{SP}$. The protein was expressed in LB at 30°C against an empty vector as a control and harvested 2 h after induction as outlined in Table 2-3, method 1. The culture was split into equal volumes and the resulting cell samples after harvest were fractionated using the PureFrac method where 100 µM NEM was added to all the buffers used for one sample. For the other sample, NEM was added at the same concentration to all buffers except to the PBS used to wash the cells in the very first step of the method.

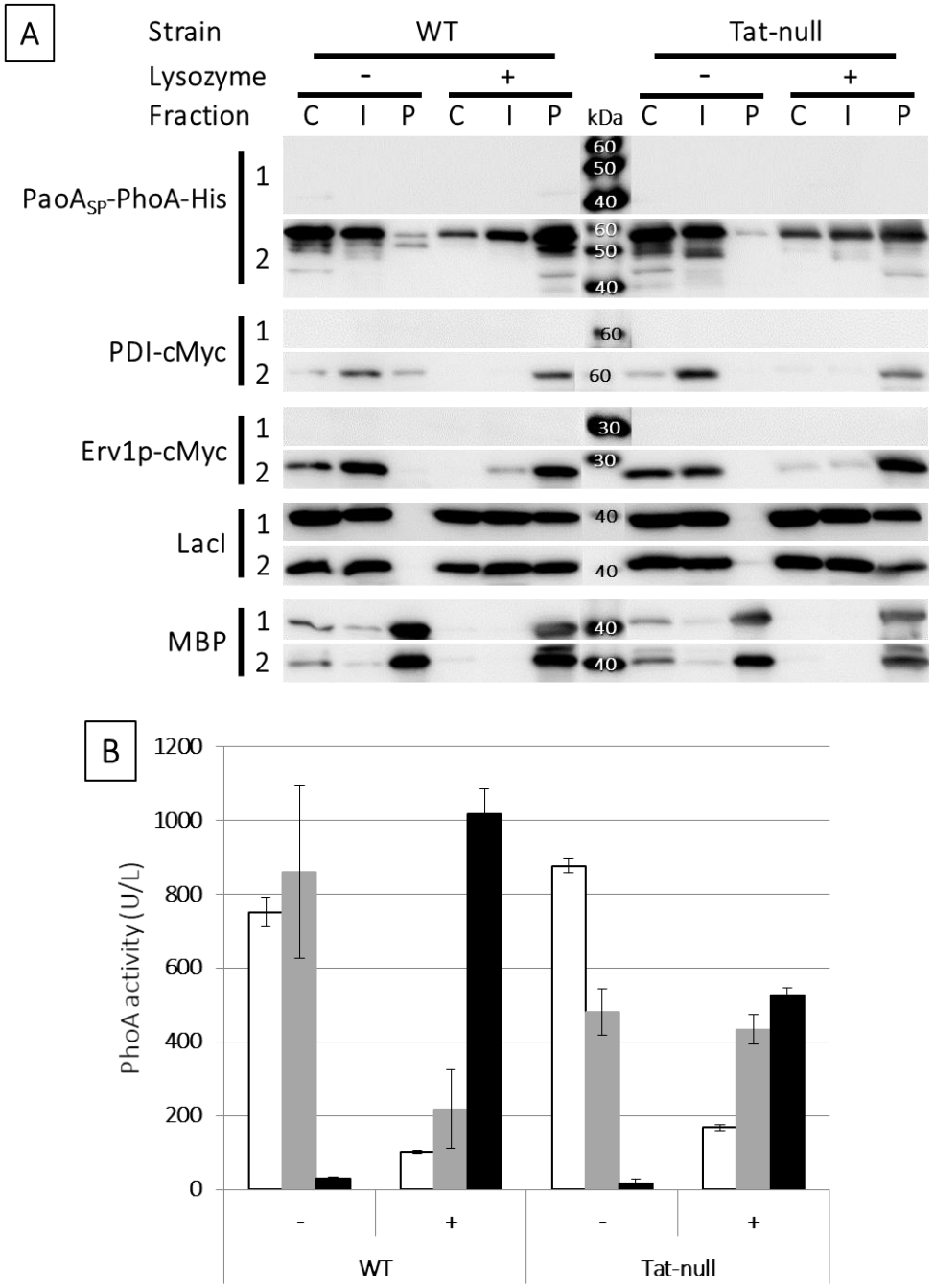


Figure 4-6: Effect of lysozyme treatment on fractionation purity with PureFrac

W3110 WT and MC4100 Tat-null strains consisting of the empty vector (1) as a control or co-expressing Tat-targeted His-tagged PaoA$_{SP}$-PhoA with cytoplasmic CyDisCo (PDI-cMyc and Erv1p-cMyc) were cultured in LB at 30°C and induced with IPTG for 2 h. The cultures were split into two samples and recovered cells were fractionated using the PureFrac method (-) or treated by the addition of 100 µg/ml of lysozyme to the hypotonic solution (+). Fractions include cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. **A.** Western-blots representing PhoA-His, PDI-cMyc, Erv1p-cMyc, LacI and MBP proteins were detected by anti-His, anti-cMyc, anti-LacI and anti-MBP antibodies respectively. Molecular weight markers are indicated in kilodalton (kDa) on the right of each blot. The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights. **B.** PhoA activity was determined using a titration curve of a commercial, purified PhoA and normalized against the empty vector. The bars indicate the cytoplasmic (white), insoluble (grey) and periplasmic (black) fractions. n=3.

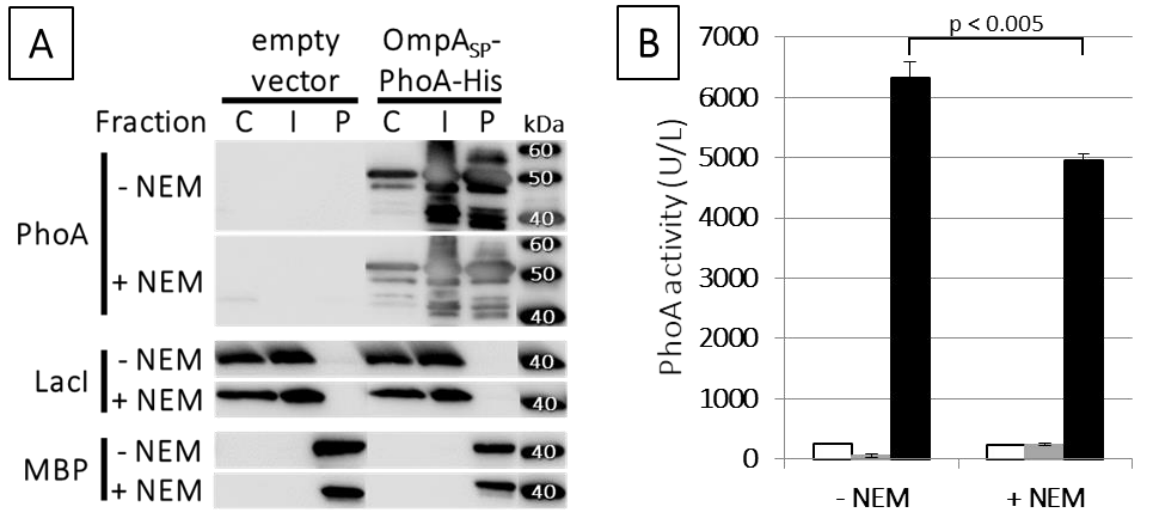


Figure 4-7: Relevance of NEM to assess the propensity of disulphide bond status

W3110 WT strain expressing sec-targeted His-tagged OmpA$_{SP}$-PhoA or only containing an empty vector were incubated for 2 h at 30°C post-induction. The cells were split in two samples and fractionated using PureFrac method into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. One sample was washed in PBS with (+ NEM) and without NEM (-NEM) while all other solutions used contained NEM for both samples. **A.** Western-blots showing PhoA-His, LacI and MBP proteins detected by anti-His, anti-LacI and anti-MBP antibodies respectively. Molecular weight markers are indicated in kilodalton (kDa) on the right of each blot. The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights. **B.** PhoA activity was determined using a titration curve of a commercial, purified PhoA. The bars indicate the cytoplasmic (white), insoluble (grey) and periplasmic (black) fractions. n=3

Figure 4-7A shows high expression of PhoA, albeit clipped products were detected in all fractions including the expected periplasmic fraction indicating export through the Sec pathway, irrespective of NEM addition to PBS. The respective cytoplasmic and periplasmic control markers LacI and MBP indicate no cross-contamination between the soluble compartments further supporting the robustness of the fractionation method. Appreciable levels of PhoA-His and associated clipped product was also detected in the insoluble fractions and likely to be derived from inclusion bodies as a result of high levels of overexpression in rich media.

The activity assays however, revealed a significant difference between the periplasmic fractions (Figure 4-7B). Indeed, without NEM, a 28 % increase in PhoA activity was observed when compared to the condition with NEM. Since PhoA is only active when the DSBs are formed, this strongly supports disulphide shuffling, enabled by air oxidation has taken place. This has resulted in more than 25 % of the extracted periplasmic PhoA-His gaining activity through correct DSB formation in

the absence of NEM. Consequently, the addition of NEM from the very first step of the sample processing, is critical to assess the activity of secreted proteins that contain DSB(s).

4.7. Conclusion

A fractionation protocol was needed for this study, giving pure sub-cellular compartments to reflect the precise localisation of expressed proteins within a cell. From the literature, cold osmotic shock and EDTA/lysozyme/cold osmotic shock are two common methods for periplasmic extraction. In this work, both presented cross-contamination between their soluble fractions. However, when $MgCl_2$ was added to the hypotonic solution of the cold/osmotic shock procedure, uncontaminated periplasmic fractions were obtained. The next step in cell fractionation is to isolate the cytoplasmic preparation. Comparison of ultrasonication and freeze/thaw cycling used to lyse the cell demonstrated that ultrasonication gives superior recovery of soluble proteins. Thus, the optimized cell fractionation method PureFrac, combines both the cold osmotic shock with $MgCl_2$ and ultrasonication procedures and was demonstrated to be a robust technique to isolate proteins with respect to their sub-cellular compartments. This was accomplished without cross-contamination and with superior recovery of soluble proteins. Minor contamination with respect to the periplasmic host marker MBP was observed in the cytoplasmic and insoluble fractions of some cases indicating that no method would give 100 % clean fractions and it is unrealistic to believe so. As the cross-contamination levels are so low in these experiments, very little room is left to misinterpret these observations as false-positives. Conversely, if the cross-contamination was significant, then this would be a major risk. Moreover, adding lysozyme was attempted to enhance soluble periplasmic recovery but led to major cross-contamination and is best avoided for this type of study.

To improve the process, wash steps could be added before the solubilization of the insoluble pellet after ultracentrifugation which should minimize contamination by carryover. If more persists, a slow speed centrifugation step could be added after solubilization of the insoluble materials to separate

the potential un-lysed cells or the ultrasonication step could be further optimized to increase cell lysis.

Since the purpose of this PhD is to investigate Tat export of expressed proteins, only the soluble cytoplasmic and periplasmic compartments are of interest. Therefore, the method to isolate further pools of different protein structures from the insoluble preparation was not optimized.

PureFrac was proven to be robust with both the WT strain and the Tat-null strain which has a compromised envelope. This process was also adequate in all the rich media conditions tested and the isolated proteins were demonstrated to be active. Moreover, the necessity for NEM was demonstrated to cap the free thiol groups and thus, prevent the formation of DSBs by air oxidation during sample processing.

The development of PureFrac has highlighted the need for specific localisation markers, particularly the selection of appropriate and consistent protein markers to monitor cross-contamination events. Hence, through the whole project, Western-Blots against LacI and MBP were performed to justify of fractions purity for every single experiment.

Overall, PureFrac presents several advantages over the compared methods from literature. It allows extraction of the periplasmic fraction without contamination by cytoplasmic endogenous (G6PD) or recombinant (PhoA-His, PDI-cMyc, Erv1p-cMyc and LacI) whereas EDTA/lysozyme and cold osmotic shock procedures led to such contamination. Moreover, PureFrac enable a higher recovery of soluble cytoplasmic proteins by sonication than by using freeze/thaw cycling. Such improvements are not only valid for WT *E. coli* strain but also for a mutant strain with impaired cell wall demonstrating the robustness of the method. Furthermore, the addition of NEM in PureFrac allow extraction of proteins which DSB status is unaltered during sample processing. The purity of the soluble fractions could be further demonstrated by running proteomics analysis on the samples from the different fractionation techniques. Finally, quantitative measurement of the benefit of PureFrac over other techniques can be achieved via such proteomics study but also via peptide mapping, His-tag purification followed by size exclusion chromatography and mass spectrometry on the fractions.

5. EVALUATION OF NATURAL TAT SUBSTRATES AS SOLUBLE PARTNER

5.1. Initial expression of the five selected candidates

The suitability of the selected five NTS as outlined in section 3.3. as an export and soluble partner in the fusion design was tested. To this aim, all five His-tagged candidates (AmiA, HyaA, PaoA, TorA and YcbK) were expressed in W3110 WT, MC4100 WT and MC4100 Tat-null strains in LB at 30°C induced for 2 h at 30°C (Table 2-3, condition 1). The cells were fractionated according to the PureFrac method (section 4.4, Appendix 4) and the expression and export profiles analysed by SDS-PAGE and Western blot. As shown in Figure 5-1A, the expression of the 5 NTS was measured through various degrees of success. YcbK was not expressed even though the plasmid sequence was confirmed before and after the experiment to rule out nucleic acid mutations. The growth profile of YcbK was similar to those of the other NTS in the WT strains with respect to that during the induction phase (Appendix 5), a lower OD_{600} was reached compared to the control strain (empty vector). This suggests that YcbK was expressed but bearing in mind the results on analysis, was instable in the cytoplasm and so was degraded. YcbK is currently of unknown function, is not in an operon structure with other genes which could highlight its role or the necessity of partners or chaperones essential for expression, folding and stability (section 3.3.). Computational analysis using Virtual Footprint of the region upstream of the open reading frame of *ycbK* within the chromosome revealed a conserved -10 and -35 promoter region as well as RpoD, CRP, ArgR and Sigma70 binding sites, which strongly support that this gene is at least transcribed (Munch et al., 2005, Solovyev, 2011). The absence of YcbK expression in this study could be due to several reasons: (i) requirement of unknown chaperones and/or partners, (ii) requirement of an unknown metal ion which was not found in the supplied LB, (iii) instability brought by the aerobic conditions as some Tat substrates are involved in anaerobic growth. Furthermore, AmiA, HyaA and PaoA were expressed but most of the proteins were detected in the insoluble fractions demonstrating their instability irrespective of the strain. TorA was the only candidate to express in a significant amount and showed partial stability as demonstrated by the intensity of the detected bands in the soluble fractions (Figure 5-1A).

Regarding export, the three highly unstable candidates AmiA, HyaA and PaoA showed no sign of crossing the inner membrane. AmiA has been overexpressed previously but no evidence of the subcellular localisation have been sought using isolated fractions (Ize et al., 2003, Bernhardt and De Boer, 2003). HyaA is known to form a dimer with HyaB and this complex is believed to be Tat exported using the signal peptide of HyaA (Rodrigue et al., 1999), although no direct evidence have been published. Moreover, evidence of the chaperone HyaE binding the HyaA signal peptide have been highlighted by Dubini and Sargent (2003). However, no evidence of HyaE being required for HyaA export has been presented. It is plausible that the chaperone HyaE and/or the partner HyaB may need to be co-expressed in order for the complex to be exported by the Tat mechanism.

In relation To PaoA, a more complex story has emerged from other authors. A cytoplasmic variant of PaoA has been expressed in the cytoplasm where it was found to interact with the PaoB and PaoC partners to form the PaoABC complex (section 3.3.) (Neumann et al., 2009, Correia et al., 2016). Lee et al. (2014) confirmed the formation of this complex but also proved that both partners are required for Tat export of the complex as "hitchhikers" thanks to the signal peptide of PaoA. The team also demonstrated the implication of the PaoD chaperone in improving the complex stability and its requirement to gain activity. In this case, PaoA would require co-expression of the partners PaoB and PaoC as well as potentially, the chaperone PaoD for Tat export. Due to the success of export reported by Lee et al. (2014), this highlighted that the PaoA experiment needed further optimization which was performed and reported in section 5.2.

The only successful candidate, with regards to export in the first experiment, was TorA, as a significant amount of proteins was detected in the periplasmic fraction of the WT strains. Conversely, no proteins were detected in the periplasm of the Tat-null strain (Figure 5-1A). Since the control markers LacI and MBP were both detected in their respective fractions, this confirms TorA was specifically exported by Tat and supports previous published data (Weiner et al., 1998, Jack et al., 2004). TorA, even without its chaperone TorD, is known to fold to a near native conformation although at a slower rate (Pommier et al., 1998, Ilbert et al., 2003).

The fractionation controls represented in Figure 5-1B indicate the purity of the isolated fractions, as confirmed by the absence of LacI and MBP contaminating proteins in the respective periplasmic and cytoplasmic fractions. Overall, only TorA was exported by Tat without any further optimization. However, the DNA sequence coding for this protein rendered the plasmid unstable which made TorA technically impractical. Therefore, none of the five selected candidates were fit to use as soluble partner as such and would require optimization.

5.2. Optimization of PaoA for Tat export

5.2.1. PaoA stability assessment

In section 5.1., PaoA was shown to not export via Tat and revealed a relatively low stability when expressed while bearing its own signal peptide. Expression of PaoA alone has never been published hence the following experiment was designed to assess the cytoplasmic expression and periplasmic export of PaoA alone and with its partners and /or chaperone. PaoA was expressed with no signal peptide to assess the stability in the cytoplasm, with $OmpA_{SP}$ to test the suitability for Sec export, with its own signal peptide as a reference and with the $TorA_{SP}$ to test the improvement of the stability or export. These proteins were expressed in the W3110 WT strain and the cells were grown in LB at 30°C and induced for 2 h (Table 2-3, condition 2) before fractionation using the PureFrac method (Appendix 4).

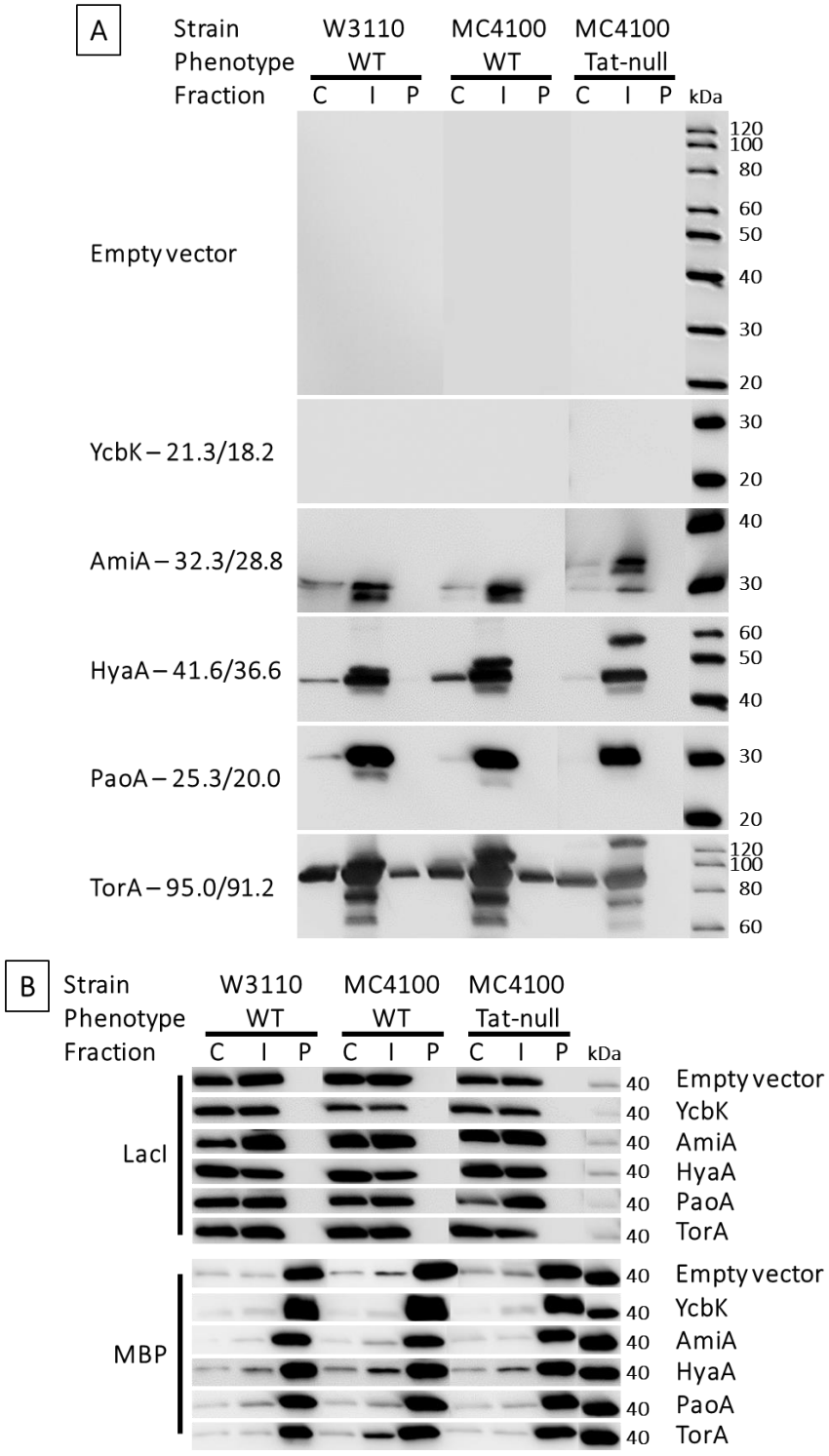


Figure 5-1: Initial screen of the selected natural Tat substrates for suitability as a soluble partner

The His-tagged natural Tat substrates AmiA, HyaA, PaoA, TorA and YcbK with their own signal peptides and an empty vector as a control, were expressed in the W3110 WT, MC4100 WT and MC4100 Tat-null strains. The cells were grown at 30°C and induced for 2 h prior to cell fractionation using the PureFrac method, into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. **A.** The western-blots represent AmiA, HyaA, PaoA, TorA, YcbK and an empty vector detected by an anti-His antibody. The molecular weights of proteins before/after theoretical cleavage of their respective signal peptide are indicated in kDa on the left. **B.** The western-blots represent the LacI (38.7 kDa) and MBP (43.4/40.7 kDa) detected on the same samples by anti-LacI and anti-MBP antibodies respectively. Molecular weight markers are indicated in kilodalton on the right (kDa). The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

In Figure 5-2, PaoA was revealed to be unstable when expressed in the cytoplasm without a signal peptide. This further supports the conclusion from Neumann et al. (2009) and Correia et al. (2016) stating that cytoplasmic PaoA is unstable without its partners PaoBC (section 3.3.). Further evidence of instability is demonstrated with its own signal peptide (25.3 kDa) and $TorA_{SP}$ (24.2 kDa), where no export was detected, and both presented expressed material in the insoluble fraction. The secondary bands observed in these insoluble fractions have an apparent larger size than the mature protein (see No SP lane) and hence, are likely to be degradation products. All the bands correlated to PaoA appeared to have migrated around 5 kDa higher than their expected sizes. This discrepancy was not observed in the work of Neumann et al. (2009) and Lee et al. (2014) but was consistently present in this work (Figure 5-1, Figure 5-2, Figure 5-3, Figure 5-4, Figure 5-5, Figure 5-7 and Figure 5-8).

With $OmpA_{SP}$, the premature protein (22.1 kDa) is present in the periplasm which suggest cytoplasmic contamination. Indeed, Sec-exported proteins are released in the periplasm upon cleavage of their signal peptide by the LepB peptidase (section 1.3.1.). The protein may be targeted to the Sec pathway but cannot be fully processed, hence the signal peptide still present. This idea is supported by the presence of presence of mature protein in the insoluble fraction which could come from PaoA jammed into Sec. The wedging of Sec would compromise the cells which is supported by the two species of the periplasmic marker MBP found in all three fractions. Furthermore, the insoluble fraction also contains mature proteins (20.0 kDa) which were properly exported via Sec but were unstable in the periplasm due to the absence of its partners and therefore, formed inclusion bodies. Overall, the data suggests that the expression of PaoA in the cytoplasm or targeted to the periplasm via Sec or Tat, resulted in protein aggregation and therefore confirms that PaoA is not stable on its own.

5.2.2. Co-expression of the partners and chaperone of PaoA

Since the instability of PaoA alone was confirmed in the section 5.2.1., co-expressing the chaperone and its partners was tested to replicate the data of Neumann et al. (2009), Lee et al. (2014) and Correia et al. (2016) as validation. According to Lee et al. (2014), PaoA only requires its partners PaoB

and PaoC to form the PaoABC complex and export via Tat into the periplasm of *E. coli*. But according to Neumann et al. (2009), the cytoplasmic chaperone PaoD improves the complex stability and is required for its activity. Therefore, two new plasmids were created for polycistronic expression of PaoB and PaoC and of PaoB, PaoC and PaoD (Table 2-2). All three proteins were C-terminally V5-tagged. These plasmids were designed to enable co-transformation with the pTTO plasmid expressing PaoA, and therefore, needed to have compatible origins of replication. Hence, the pBR replication origin was selected since it offers a copy number of 15-20 close to the 18-22 copies for p15A, the origin of replication of pTTO (Chang and Cohen, 1978). Additionally, the new plasmids contained an ampicillin resistance gene for specific selection.

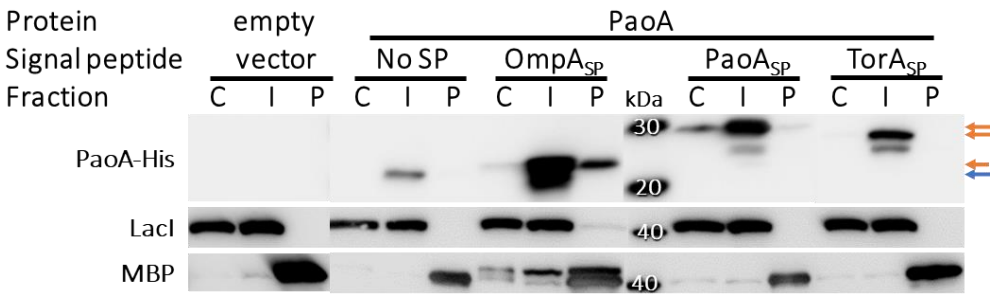


Figure 5-2: Assessing the stability of PaoA

An empty vector ant the natural Tat substrate PaoA-His fused with no signal peptide (No SP), the signal peptide of OmpA (OmpA$_{SP}$), PaoA (PaoA$_{SP}$) or TorA (TorA$_{SP}$) were expressed in the W3110 WT strain. The cytoplasmic (C), insoluble (I) and periplasmic (P) fractions were obtained using the PureFrac method. The Western-blots represent the detection of PaoA-His, LacI and MBP using an anti-His, anti-LacI and anti-MBP antibodies respectively. The bands matching the sizes of the premature and mature proteins are indicated by orange and blue arrows respectively on the right side of the relevant blots. Molecular weight markers are indicated in kilodalton (kDa). The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

Due to their relatively similar sizes, PaoB-V5 (35.3 kDa) and PaoD-V5 (36.4 kDa) could not be distinguished on anti-V5 Western-blots and have therefore been annotated as PaoB-V5/PaoD-V5 whilst PaoC-V5 (79.5 kDa) was well separated. PaoA-His, with and without its own signal peptide or an empty vector, was co-expressed with PaoBC or PaoBCD from two different plasmids to replicate the export reported by Lee et al. (2014). The W3110 strain were used and the cells were fractionated

according to the PureFrac method (Appendix 4) into the cytoplasmic, insoluble and periplasmic fractions.

The results, represented in Figure 5-3, showed no significant improvement of PaoA-His export with or without PaoBC nor PaoBCD. PaoA still remained largely in the insoluble fraction with a small portion located soluble in the periplasm. However, PaoA was also detected in a similar amount in the periplasm when expressed without a signal peptide indicating that the periplasmic localisation of this proteins is not the result of translocation by Tat. Moreover, PaoC-V5 and PaoB-V5/PaoD-V5 were detected in the periplasm when co-expressed with PaoA-His (no signal peptide) which was unexpected since neither of these proteins bear a signal peptide. Moreover, PaoB-V5/PaoD-V5 were again detected in the periplasm when co-expressed with the empty vector. No proteins are significantly detected in the periplasm without PaoD-V5, and the marker proteins are present in their respective fractions demonstrating the absence of cross-contamination (Figure 5-3). Therefore, PaoD-V5 somehow enabled export of PaoB-V5/PaoD-V5 and PaoC-V5 (only when co-expressed with PaoA-His) most likely through its chaperone activity. Furthermore, soluble cytoplasmic PaoC-V5 appeared to be more stable when co-expressed with PaoD-V5 as published by Neumann et al. (2009). Altogether, PaoD-V5 is needed to increase the stability of PaoC-V5 and export of PaoC-V5 and PaoB-V5/PaoD-V5 but its expression was still not enough to obtain export of the complex.

5.2.3. Effect of the molybdenum cofactor on export of PaoABC

In the section 5.2.2., PaoABCD were co-expressed in LB as described by Lee et al. (2014) but PaoA was not significantly exported. Otrelo-Cardoso et al. (2014) published that in the PaoABC complex, the PaoA subunit binds to two iron-sulphur clusters and the PaoC subunit binds to a molybdenum cofactor as identified by small-angle X-ray scattering. This data was later confirmed by Correia et al. (2016) by crystallography. Therefore, a new attempt to export PaoA was performed using the same conditions as the section 5.2.2. except that sodium molybdate and iron chloride were added to the culture medium to a final concentration of 1 mM (Neumann et al., 2009) and 0.5 mM respectively (section 2.3.1.).

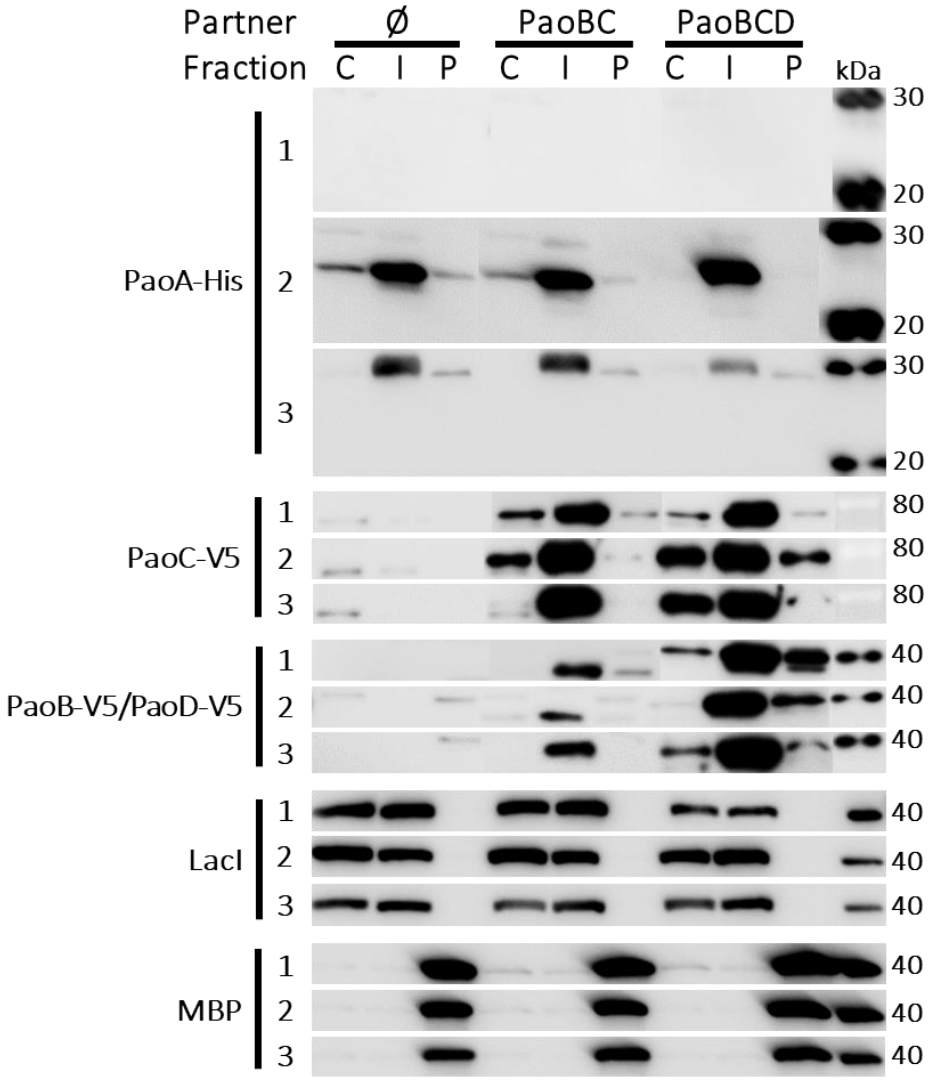


Figure 5-3: Co-expression of PaoA with its partners PaoBC(D)

An empty vector as control (1), PaoA-His without a signal peptide (2) or $PaoA_{SP}$-PaoA-His (3) were expressed alone (Ø) or co-expressed with the partners PaoB-V5 and PaoC-V5 (PaoBC) or with the partners and the chaperone PaoD-V5 (PaoBCD) in the W3110 WT strain. The cells were grown on LB at 30°C, induced for 2 h at 30°C and fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. The Western-blots represents the detection of the PaoA-His, PaoC-V5 and PaoB-V5/PaoD-V5 using anti-His and anti-V5 antibodies respectively. Molecular weight markers are indicated in kilodalton on the right. The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

No significant improvement was observed on the export status of the PaoABCD proteins compared with the expression without the cofactors (Figure 5-4). The cofactors were thought to improve the stability of the complex but surprisingly, cytoplasmic PaoABCD were detected in a lower amount. The LacI and MBP purity control markers, however, indicated extraction of nearly all periplasmic

proteins, no contamination by cytoplasmic proteins and a high recovery of soluble cytoplasmic proteins.

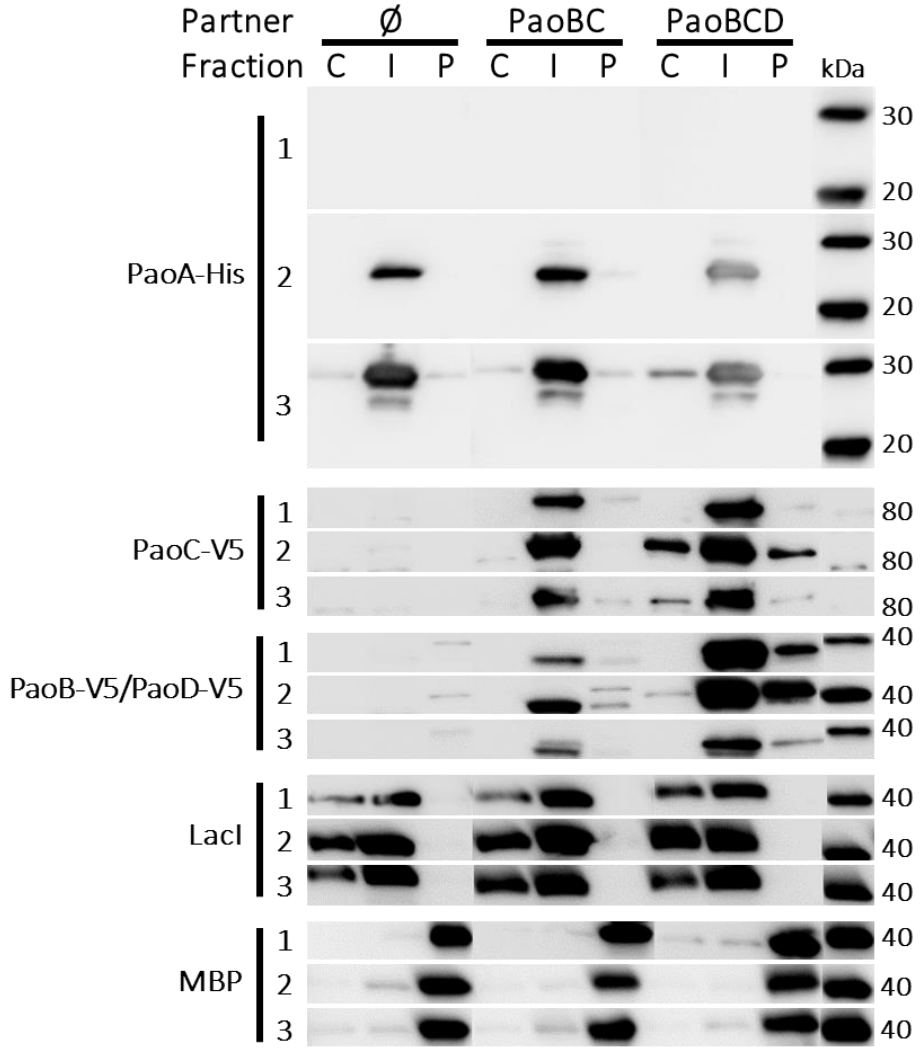


Figure 5-4: Effect of the molybdenum cofactor on export of PaoABC

An empty vector as control (1), PaoA-His without a signal peptide (2) or $PaoA_{SP}$-PaoA-His (3) were expressed alone (Ø) or co-expressed with the partners PaoB-V5 and PaoC-V5 (PaoBC) or with the partners and the chaperone PaoD-V5 (PaoBCD) in the W3110 WT strain. The cells were grown on LB supplemented with $FeCl_3$ (0.5 mM) and Na_2MoO_4 (1 mM) at 30°C, induced for 2 h at 30°C and fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. The Western-blots represents the detection of the PaoA-His, PaoC-V5, PaoB-V5/PaoD-V5, LacI and MBP using anti-His, anti-V5, anti-LacI an anti-MBP antibodies respectively. Molecular weight markers are indicated in kilodalton on the right. The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

5.2.4. Culture conditions and strains assessment for export of PaoABC

In the next attempt to reproduce the data obtained by Lee et al. (2014), the divergence of strains and culture conditions were considered. The MC4100 WT strain was used by the Lee team while

W3110 WT was used in these experiments and even if no difference were observed in terms of protein localisation when studying the NTS between the two strains, they were compared to test whether the strain mattered. It is interesting to note here that the *paoABCD* operon is present in W3110 WT strain whereas it has been deleted in the MC4100 WT strain (section 3.3.). With respect to experimental methodology, very little information regarding the particular expression conditions used have been stated by Lee et al. (2014), and despite efforts to discern this information from attempted communication with the authors. However, experimental methods for the cytoplasmic expression of PaoABC have been provided by Neumann et al. (2009) and cited since by other authors. Therefore, the culture conditions referenced from the latter publication and subsequent citations thereof, are hence believed to be a 16 h induction at 22°C with 100 rpm agitation or a 2 h induction at 30°C with 220 rpm. Consequently, W3110 WT and MC4100 WT were compared for Tat export of PaoABC in three culture conditions: 2 and 4 h induction at 30°C with 220 rpm agitation (Table 2-3, condition 1 and 2) and 16 h incubation at 22°C with 100 rpm agitation.

Figure 5-5 indicated no export of PaoA-His in the tested conditions meaning that the genetic differences between the strains have no impact on localisation of the protein of interest but also that these culture conditions do not influence protein localisation. Figure 5-6 revealed export of PaoC-V5 and PaoB-V5/PaoD-V5 after 2 h post-induction in W3110 WT but could be due to fractionation artefacts since LacI was observed contaminating the periplasmic fraction in this condition. Similarly, the large periplasmic band observed 16 h post-induction in the same strain with PaoABCD and signal peptide is non-conclusive because of the presence of contaminating LacI in the periplasm. With the other conditions, PaoC-V5 was detected in the cytoplasm only when co-expressed with PaoD-V5 in the MC4100 WT strain whereas it was observed in the cytoplasm of W3110 WT in all conditions. This difference indicated that PaoD improved the solubility of PaoC through its chaperoning effect. The expression of endogenous PaoD might be the reason why this protein is not required in W3110 to solubilize PaoC-V5. The localisation of LacI in the other conditions than cited above and MBP validated the fractionation robustness. The upper and lower bands present in the insoluble fractions when expressing $PaoA_{SP}$-PaoA-His matched the premature and

mature proteins respectively and could be the results of signal peptide cleavage in inclusion bodies. Overall it seems that the PaoABC subunits are more stable in solution when PaoD was co-expressed but none of the growth conditions tested allowed export of PaoABC complex.

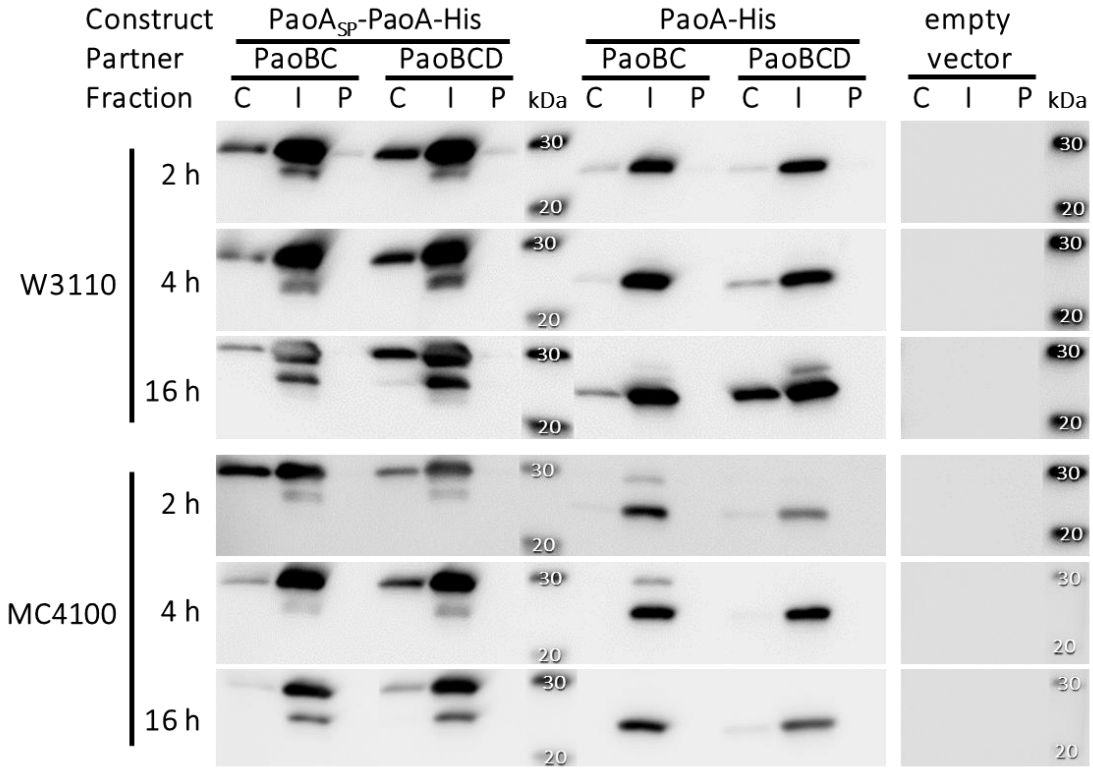


Figure 5-5: Culture condition and strain screen to achieve export of PaoABC – Part I

PaoA-His without a signal peptide or PaoA$_{SP}$-PaoA-His were co-expressed with the partners PaoB-V5 and PaoC-V5 (PaoBC) or with the partners and the chaperone PaoD-V5 (PaoBCD) in the W3110 WT and MC4100 strains. The cells were grown on LB supplemented with $FeCl_3$ (0.5 mM) and Na_2MoO_4 (1 mM) at 30°C, induced for 2 h or 4 h at 30°C or for 16 h at 22°C and fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. The Western-blots represents the detection of PaoA-His using anti-His antibody. Molecular weight markers are indicated in kilodalton. The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

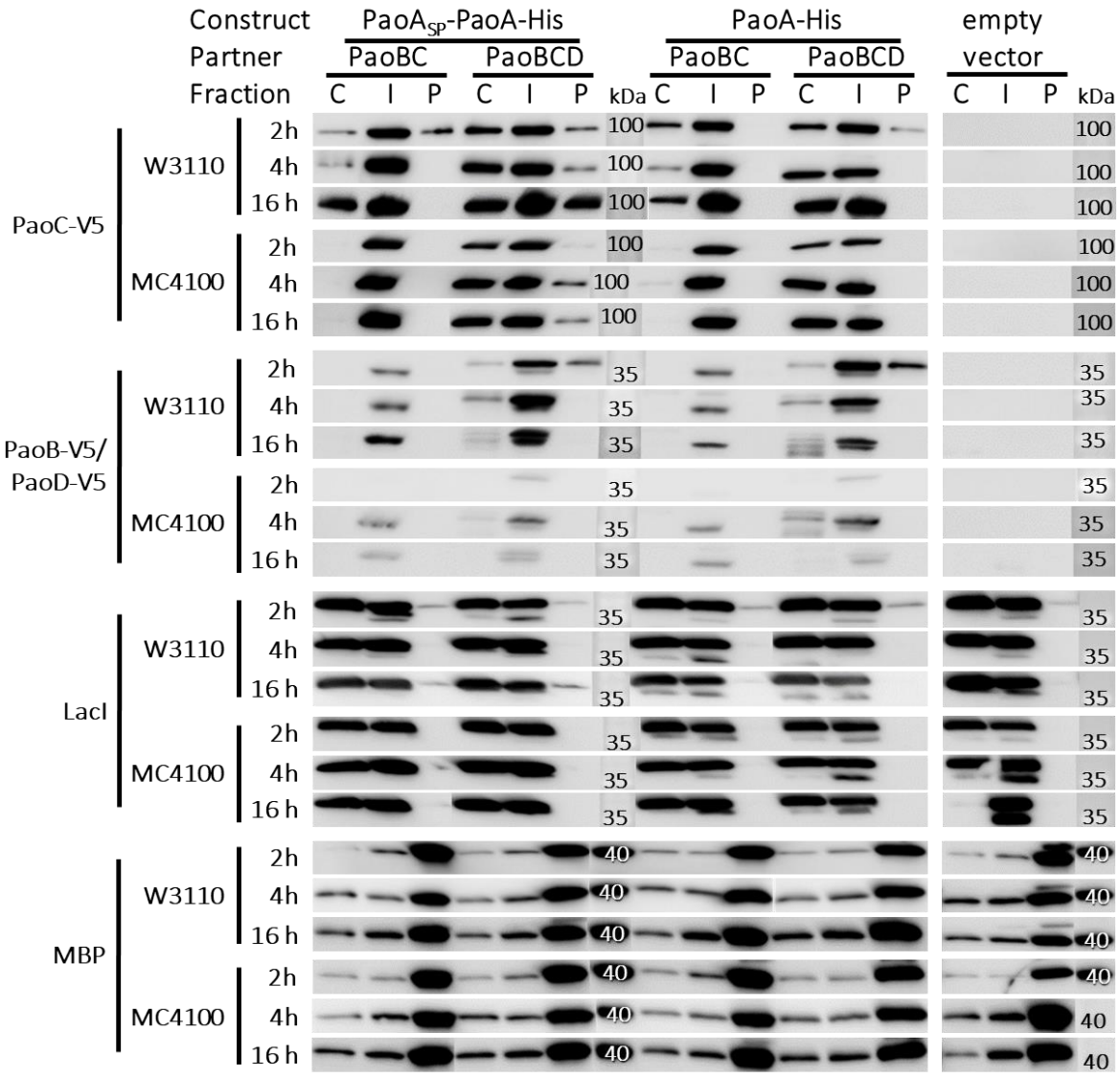


Figure 5-6: Culture condition and strain screen to achieve export of PaoABC – Part II

The samples used in this figure are the same as the ones used in Figure 5-5. PaoA-His without a signal peptide or PaoA$_{SP}$-PaoA-His were co-expressed with the partners PaoB-V5 and PaoC-V5 (PaoBC) or with the partners and the chaperone PaoD-V5 (PaoBCD) in the W3110 WT and MC4100 strains. The cells were grown on LB supplemented with FeCl3 (0.5 mM) and Na2MoO4 (1 mM) at 30°C, induced for 2 h or 4 h at 30°C or for 16 h at 22°C and fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. The Western-blots represents the detection of PaoC-V5, PaoB-V5/PaoD-V5, LacI and MBP using anti-V5, anti-LacI and anti-MBP antibodies respectively. PageRuler Plus Prestained Protein Ladder (Life Technologies) and MagicMark™ (Life Technologies) molecular weight markers are indicated in kilodalton (kDa). The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

5.2.5. Single plasmid expression of PaoABCD in several growth conditions

The strain and the tested growth conditions made no difference in terms of export of PaoA-His.

Therefore, more culture conditions were evaluated with the objective of reproducing the data

published by Lee et al. (2014). Another difference from their work to this tentative remained in their expressions of the PaoABC proteins from a single plasmid whereas two plasmids were used here so far. Two plasmids induce more stresses upon the cell than a single plasmid due to the presence of two antibiotics to maintain the selection. Also, the copy numbers were slightly different from one plasmid to the other which might have led to differential expression of the PaoABC complex proteins creating an unbalanced ratio. Therefore, two new pTTO-based vectors were designed to co-express the PaoABC or the PaoABCD proteins from a single vector (Table 2-2). This modification was used as an opportunity to change the detection tag on PaoD for the HA tag to be able to distinguish it from the similar sized protein PaoB. To remain as close as possible from Lee et al. (2014) conditions, the culture media were supplemented with $FeCl_3$ (0.5 mM) and Na_2MoO_4 (1 mM) in the growth conditions described in Table 2-3.

Unfortunately, no export of PaoA-His was observed in the conditions tested which seem to have no impact on the protein localisation (Figure 5-7A). The same observations were made on PaoB-V5 (Figure 5-7B) and PaoC-V5 showed signs of periplasmic export but no localisation difference between the conditions. PaoD-HA appears to be exported to the periplasm as well but at very low and fluctuating levels. The Figure 5-7C proves the robustness of the fractionation by the absence of LacI contaminating the periplasm and the full extraction of MBP into the periplasmic fraction. The only growth condition where changes observed in terms of protein localisation was the condition 7 with minimal media. This condition is shown once again to lead to poor recovery of soluble materials and a less robust fractionation purity.

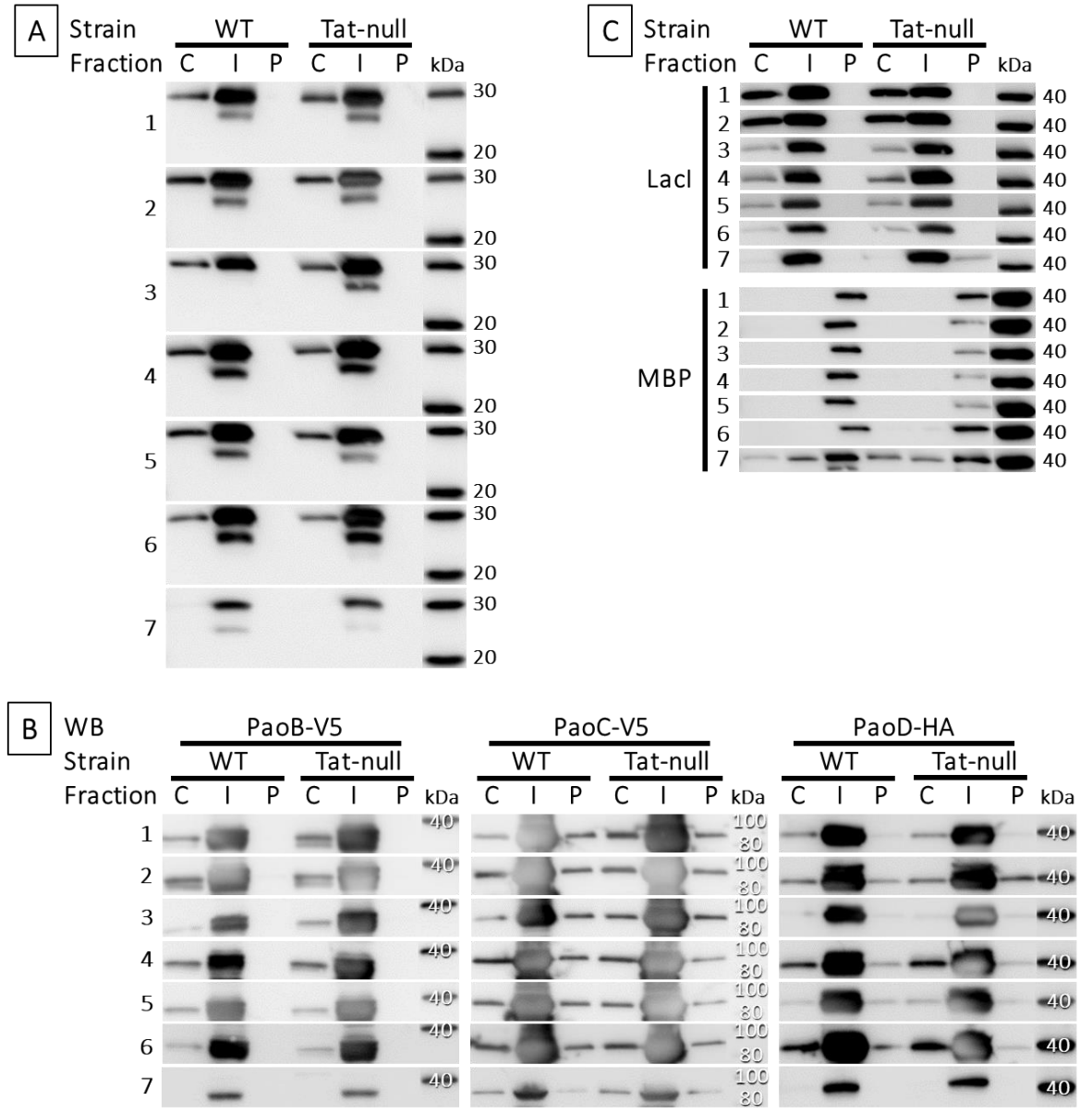


Figure 5-7: Expression of the PaoABCD proteins from a single plasmid in different growth conditions

$PaoA_{SP}$-PaoA-His, PaoB-V5, PaoC-V5 and PaoD-HA were co-expressed in the MC4100 WT and Tat-null strain. The cells were grown in different conditions: induced (1) in LB for 2 h at 30°C, (2) in LB for 4 h at 30°C, (3) in LB for 2 h at 37°C, (4) in LB for 16 h at 18°C, (5) in 2x TY for 2 h at 30°C, (6) in fed-batch culture for 16 h at 18°C or (7) in minimal media for 16 h at 18°C. Each media was supplemented with $FeCl_3$ (0.5 mM) and Na_2MoO_4 (1 mM) and the cells were fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. The Western-blots (WB) represent the detection of the PaoA-His (**A**), PaoC-V5, PaoB-V5 and PaoD-V5 (**B**) and LacI and MBP (**C**) using anti-His, anti-V5, anti-HA, anti-LacI and anti-MBP antibodies respectively. PageRuler Plus Prestained Protein Ladder (Life Technologies) and MagicMark™ (Life Technologies) molecular weight markers are indicated in kilodalton (kDa). The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

5.2.6. Influence of the induction level with IPTG on PaoA export

The partners, chaperone, cofactors, strain, vector and growth condition were optimized for the export of PaoA but the results from Lee et al. (2014) are still not reproducible. The last parameter which wasn't optimized is the induction level by ITPG. Indeed, the article specifies that IPTG was added at the time of induction to a working concentration of 100-500 µM. To appreciate the impact of IPTG level on the export of PaoA-His, a titration was performed using 0, 10, 100 and 500 µM of IPTG. The closest expression conditions were selected: in MC4100 WT, in LB supplemented with $FeCl_3$ (0.5 mM) and Na_2MoO_4 (1 mM) at 30°C, induced for 16 h at 22°C, with the partners PaoB-V5 and PaoC-V5, with the chaperone PaoD-HA and in a single vector.

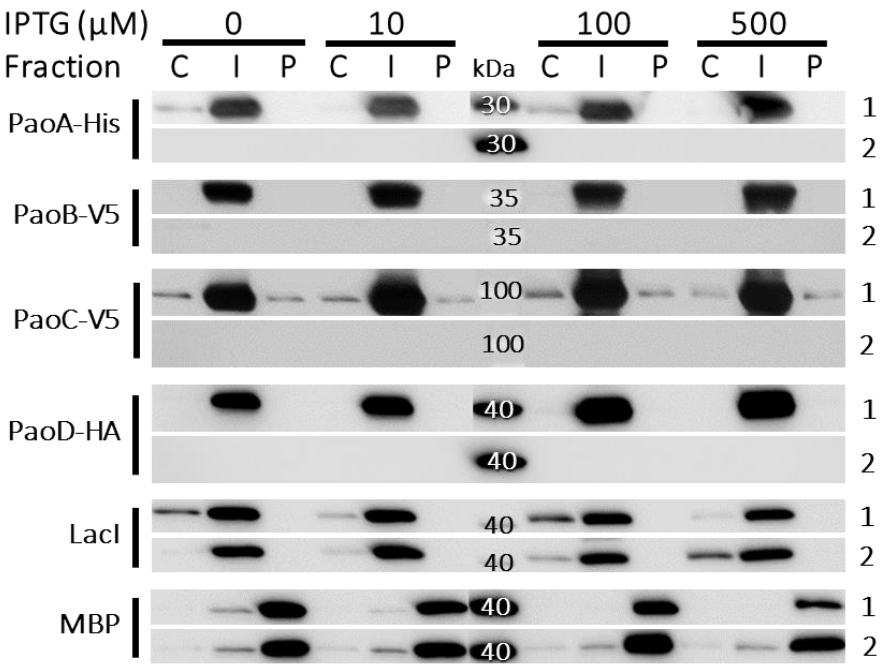


Figure 5-8: IPTG titration to export PaoA

$PaoA_{SP}$-PaoA-His (1) or an empty vector (2) were co-expressed with the partners PaoB-V5 and PaoC-V5 and the chaperone PaoD-V5 in the MC4100 WT strain. The cells were grown on LB supplemented with FeCl3 (0.5 mM) and Na2MoO4 (1 mM) at 30°C, induced for 16 h at 22°C with 0, 10, 100 or 500 µM of IPTG and fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. The Western-blots represents the detection of the PaoA-His, PaoB-V5, PaoC-V5, PaoD-HA, LacI and MBP using anti-His, anti-V5, anti-HA, anti-LacI and anti-MBP antibodies respectively. PageRuler Plus Prestained Protein Ladder (Life Technologies) and MagicMark™ (Life Technologies) molecular weight markers are indicated in kilodalton (kDa). The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

The absence of PaoA-His in the periplasmic fractions (Figure 5-8) indicated that none of the IPTG levels tested influenced its export. PaoB-V5 and PaoD-HA were found only in the insoluble fractions whereas PaoC-V5 was found soluble in both the cytoplasm and the periplasm even if the majority was detected in the insoluble fraction irrespective of the level of induction. The large bands of PaoABCD proteins with 0 µM of IPTG were the results of the leakiness of the *tac* promoter controlling their expression (section 2.2.10.1.).

5.3. Conclusion on the use of a natural Tat substrate as a soluble partner

To use as a soluble partner enhancing both the folding of the protein of interest and the targeting to the Tat pathway, an NTS was initially considered to fuse to the recombinant protein. To identify suitable applicants, five NTS were selected using criteria such as size and oligomeric status. From the initial screen of these five proteins, only $TorA_{SP}$-TorA-His showed Tat export. The translocation failure from the other candidates could be due to missing partner(s) or chaperone(s) or yet unknown specific conditions. However, TorA is a large protein (94.5kDa) and only small size proteins could be fused to it for Tat export since protein with a maximum known of 70 Å can be processed by this pathway (Palmer and Berks, 2012).

The NTS PaoA was investigated more thoroughly since its successful Tat export was reported (Lee et al., 2014). The authors demonstrated the requirement of the partners PaoB and PaoC to form the pre-export complex PaoABC. Also, Neumann et al. (2009) established the need to co-express the chaperone PaoD to stabilize PaoC and confer activity to the PaoABC complex. However, Lee et al. (2014) published export of PaoABC without co-expressing this chaperone.

Here, the candidate was observed in its premature and mature forms in the insoluble fractions suggesting that the signal peptide can be cleaved off in inclusion bodies. Interestingly, with the dual plasmid expression, PaoB-V5/PaoD-V5 were observed in the periplasm when co-expressed with the empty vector or the leaderless PaoA-His controls while the fractionation purity was proven via the LacI and MBP blots (sections 5.2.2., 5.2.3. and 5.2.4.). Since this event was not reported using the

MC4100 strain (Figure 5-6), the export could be the result of association of plasmid-encoded PaoBC with endogenous PaoA in the W3110 strain since the MC4100 does not possess the *pao* operon (section 3.6.). In the end, efforts to replicate the exact conditions used to express periplasmic PaoABC were unsuccessful and thus the export data remains unreproducible.

Overall, the initial candidates tested proved to be unsuitable to use as a soluble carrier protein in the fusion design. Optimisation of the expression conditions might lead to their export but the story of PaoA troubleshooting indicated no guaranteed success. Going back to the list of NTS (Appendix 1) could be an option but based on this work, it would likely be laborious, difficult to interpret data and very time consuming as little is known about these proteins. Therefore, it was a better use of time to look for a carrier from an alternative source.

6. EVALUATION OF REPORTER PROTEIN CANDIDATES

6.1. Introduction to reporter proteins

Many proteins are used as reporter for gene expression and regulation studies (Tsien, 1998, Thorne et al., 2010, Juers et al., 2012). Applications are wide and cover transcriptional control element testing, identification of interacting proteins, monitoring transfection efficiency, viral assays and mechanism of action, promotor region identification as well as many other processes (Schenborn and Groskreutz, 1999). Because the results are observed through the reporter protein's activity, they need to provide a measurable readout which are available with different properties: enzymatic activity (e.g. β-galactosidase, PhoA, chloramphenicol acetyltransferase), fluorescence (e.g. GFP), luminescence (e.g. luciferase) (Miraglia et al., 2011, Juers et al., 2012).

The experiment helps determinate the most qualified reporter protein for the desired task. β-galactosidase is easy to assay, widely used for *in-situ* staining but is a large protein due to its homotetrameric form (465 kDa) and there can be endogenous activity (Schenborn and Groskreutz, 1999). PhoA is a naturally exported enzyme in *E. coli* and its activity can be assessed with an easy and sensitive assay but potential endogenous background activity need to be subtracted (Schenborn and Groskreutz, 1999). GFP was derived in various colours which can report of several activities in a single assay. It is a highly stable protein and require no substrate (Schenborn and Groskreutz, 1999, Day and Davidson, 2009). Luciferases are enzymes capable of luminescence which can be measured through a fast and easy assay with a high sensitivity and a wide linear range but it requires a special equipment (luminometer) and the protein is labile (Schenborn and Groskreutz, 1999).

Here, proteins of interest were thought to be first replaced by a reporter protein to validate the fusion design as describe in section 3.4. By selecting suitable NTS and reporter protein which both export specifically through the Tat pathway, the fusion of these two proteins should also lead to translocation. Here, the objective of the reporter protein was to inform onto the export of the fusion protein and its native folding through its activity. Selecting such reporter proteins is not trivial for *E. coli* due to their limited amount and to their lack of validity in different cellular compartments.

Nevertheless, β-galactosidase, PhoA, sfGFP and hGH were selected and their evaluation for Tat export and periplasmic activity are described in this chapter.

6.2. β-galactosidase

β-galactosidase is a natural cytoplasmic enzyme of *E. coli*, composed of 116 kDa monomers which form a 465 kDa homotetramer. It is encoded by the *lacZ* gene from the *lacZYA* operon (Juers et al., 2012). A simple and quick activity assay is available whereby the amount of o-nitrophenol (yellow) released from the hydrolysis of o-nitrophenyl-β-D-galactopyranoside (ONPG) by β-galactosidase, can be measured. It does not possess any DSBs and therefore, does not require the use of a DSB enabling technology (section 1.4.1.). With respect to export in *E. coli*, β-galactosidase was reported to not export via Sec because of multiple regions within the amino acid sequences as well as N-terminal positively charged amino acids (Lee et al., 1989). Furthermore, Tat export has not been reported to date in *E. coli*. However, successful translocation of β-galactosidase through the Tat pathway has been demonstrated in *Bacillus subtilis* using the signal peptide of alkaline phosphatase D (Xia et al., 2010, Ren et al., 2016). Due to the large size of the complex, it was believed that the monomeric components are primarily exported prior to assembly into active complexes after crossing the membrane. It is important to note here that the reason why this enzyme was secreted by Xia et al. (2010) and Ren et al. (2016) might lie in the different Tat mechanism of *Bacillus subtilis* as detailed in section 1.3.2.6., the different sequence of their engineered β-galactosidase, the use of the signal peptide from PhoD, the requirement of TatAd-Cd co-expression and that the alkaline phosphatase D is not found in Gram-negative bacteria. Thus, it is plausible to speculate that β-galactosidase at least in monomeric form, can be exported by the Tat machinery of *E. coli* and therefore would make it a suitable reporter partner to test.

A preliminary experiment was performed to confirm the absence of β-galactosidase export via the Sec pathway. To this end, the *E. coli lacZ* gene was fused to a C-terminal His tag for detection (coding for LacZ-His) and fused to the N-terminal $OmpA_{SP}$ for export. A cytoplasmic control with no signal

peptide was also used. During this experiment, Tat export was also investigated by N-terminal fusion of the five selected NTS signal peptides (AmiA, HyaA, PaoA, TorA and YcbK) described in Chapter 5. All proteins were expressed in the MC4100 WT and Tat-null strains in LB at 30°C with induction of 2 h at 30°C (Table 2-3, condition 1) and fractionated using the PureFrac method (Appendix 4). As expected, cytoplasmic LacZ-His expressed well and was localised in the cytoplasm and insoluble fraction, presumably as inclusion bodies (Figure 6-1A). Moreover, the insoluble materials showed greater activity than the cytoplasmic proteins (Figure 6-1B). But, the latter observation is most likely to be the result of a higher concentration of enzyme in the insoluble fraction than in the cytoplasmic fraction as the activity presented is not specific to the amount of protein. The activity in the insoluble fraction is suspected to come from the inclusion bodies as they have been shown to be capable of retaining the proteins activities (García-Fruitós et al., 2005, de Groot and Ventura, 2006). Nevertheless, the fact that the enzyme is active reflected the stability of the monomers and the correct assembly into an active tetrameric form.

Similarly, when fused to $OmpA_{SP}$ and in support of published data (Lee et al., 1989), LacZ-His did not export via the Sec pathway. Surprisingly, no export via Tat was evident with respect to the five NTS signal peptides tested, in terms of periplasmic localisation or periplasmic activity. However, cytoplasmic stability appears to be signal peptide dependent where $AmiA_{SP}$, $PaoA_{SP}$ and $YcbK_{SP}$ enable polypeptides to form stable β-galactosidase complexes in the cytoplasm whereas $HyaA_{SP}$ and $TorA_{SP}$ do not, irrespectively of the strain (Figure 6-1B). This difference showed that signal peptides do not perform in the same way and can lead to various extent of instability of the recombinant protein. The differences highlighted in the set of signal peptides in terms of net charge, hydrophobicity or length (Appendix 2), do not correlate to any behaviour given to β-galactosidase. The Western-blots revealed no signs of double bands which is a usual indication of signal peptide cleavage. But the western-blot resolution prevents from identifying them as premature or mature proteins. Since β-galactosidase presented activity in the cytoplasm, N-terminal signal peptides do not interfere with the folding of the monomers nor with the assembly of the active tetramers. With the signal peptides, activity was also demonstrated in the insoluble fractions (Figure 6-1B). Since a similar activity of

about 50 kU/L was observed in the absence of a signal peptide, it is suspected to come from inclusion bodies rather than membrane-associated proteins due to their signal peptide.

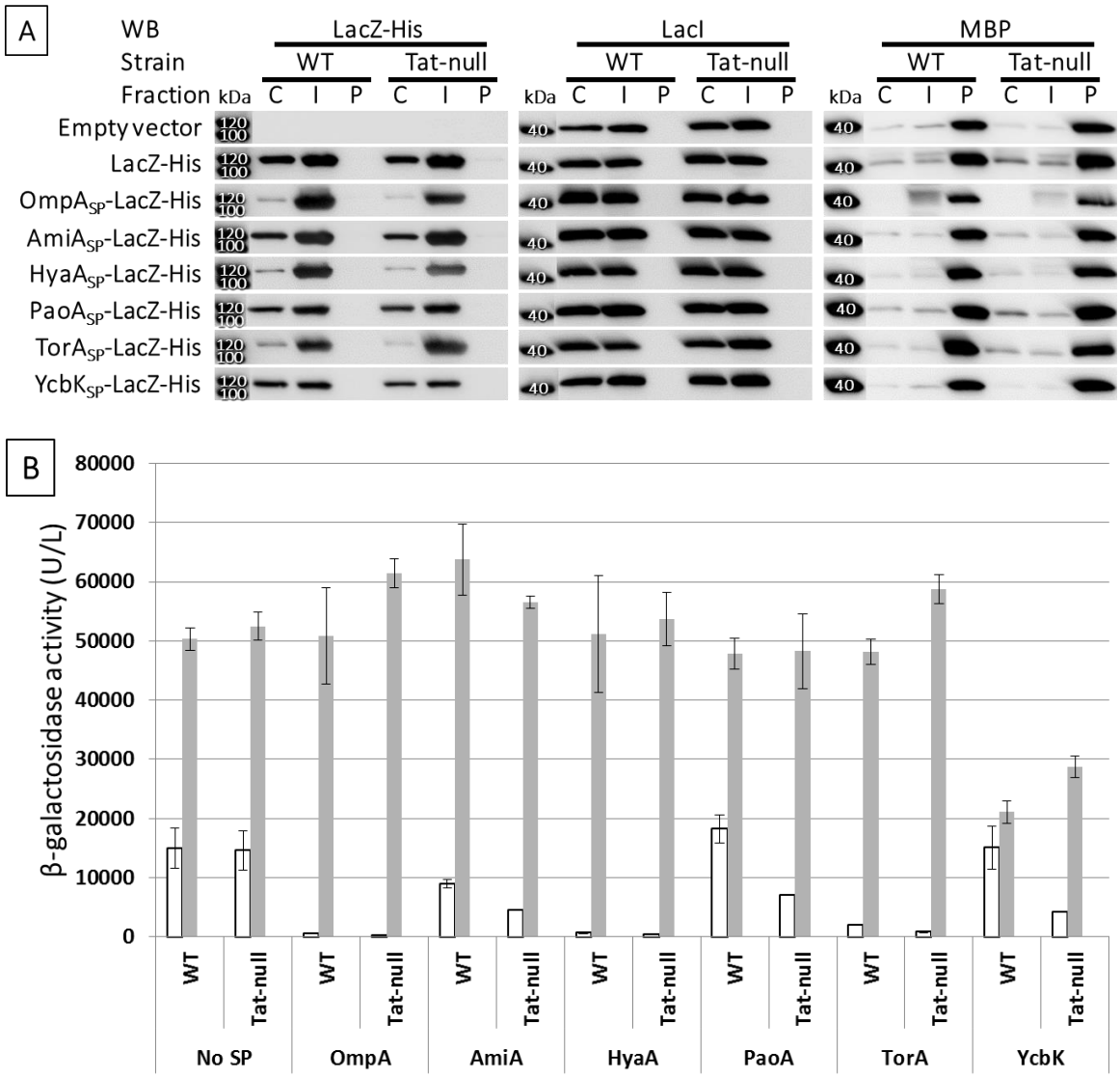


Figure 6-1: Cellular localisation and activity profiles of β-galactosidase expressed in *E. coli*

His-tagged LacZ with no signal peptide (No SP) or fused with $OmpA_{SP}$, $AmiA_{SP}$, $HyaA_{SP}$, $PaoA_{SP}$, $TorA_{SP}$ or $YcbK_{SP}$ were expressed in the MC4100 WT and Tat-null strains. The cells were harvested after 2h induction at 30°C and fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. **A.** Western-blots represent the detection of LacZ-His, LacI and MBP using anti-His, anti-LacI and anti-MBP antibodies respectively. Molecular weight markers are indicated in kilodalton (kDa). The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights. **B.** β-galactosidase activity assays were performed in triplicate and measured against a titration curve of a commercially available β-galactosidase as a standard. The background signals were eliminated by subtracting the activity value from the empty vector's fractions to each corresponding fraction. The bars indicate the cytoplasmic (white), insoluble (grey) and periplasmic (black) fractions.

The localisation results are strengthened by the LacI and MBP controls which proved the purity of the periplasmic fraction while the periplasmic proteins are still extracted (Figure 6-1A). Overall, β-galactosidase was not a suitable reporter protein because it showed no translocation via Tat likely because it folded into its tetrameric structure in the cytoplasm before reaching the Tat translocon preventing export. To better understand the expression, secretion and stability behaviours of β-galactosidase with the different signal peptides, analysing the specific activity of the sample would have been more informative. However, since the reporter protein is only a tool in this study and β-galactosidase presented no sign of Tat export, no further work was performed on this recombinant protein.

6.3. Alkaline phosphatase

PhoA is a zinc homodimeric enzyme (99 kDa) formed by two non-covalently linked subunits (49 kDa) (Fisher et al., 2008). This protein is naturally located in the periplasm of *E. coli* thanks to its Sec signal peptide (Michaelis et al., 1983). PhoA requires two DSB on each subunit to be correctly folded and active which explains its natural need for periplasmic export (Derman and Beckwith, 1991). These DSB were found to be required for Tat export and could be cytoplasmically formed by CyDisCo (Hatahet et al., 2010, Matos et al., 2014). In these cases, the signal peptide of TorA was used to direct the disulphide bonded PhoA to the periplasm via Tat making this protein a potential candidate as reporter protein.

6.3.1. Evaluation of alkaline phosphatase as a reporter protein

During the optimisation of the fractionation method (Chapter 4), $TorA_{SP}$-PhoA-His was shown to successfully export into the periplasm via the Tat pathway when co-expressed with CyDisCo. This experiment highlighted condition 4 (Table 2-3) as the culture parameters required for optimal Tat export: growth at 30°C and 4 h induction at 30°C. Following on from this discovery, the export of PhoA was tested with a restricted set of signal peptides in the presence or absence of CyDisCo for

DSB formation. The β-galactosidase case (section 6.2.) demonstrated that the use of signal peptides is not straight forward. It was therefore decided that selecting the signal peptides which have been already proven successful was a better approach. Hence, the signal peptide of TorA and PaoA were selected based on previous publications detailed in section 3.4. and 5.2. Additionally, $TorA_{SP}$-PhoA-His and $PaoA_{SP}$-PhoA-His, PhoA-His with no signal peptide and the Sec signal peptide $OmpA_{SP}$ completed the set.

Expectedly, PhoA-His was Tat-dependently exported and active in the periplasmic fraction when fused to $TorA_{SP}$ and co-expressed with CyDisCo (Figure 6-2, Figure 6-3). In this condition, proteins were also detected in the cytoplasm and insoluble fractions, the latter presumably from inclusion bodies or membrane-associated proteins. The activity assays indicated that the material in the insoluble fraction is indeed inactive and therefore not folded correctly, but also surprisingly, the proteins in the cytoplasmic fraction did not show activity (Figure 6-3). Although, a higher quantity of cytoplasmic proteins was detected with CyDisCo suggesting that this technology improved protein stability supposedly through its DSB formation activity. This indicated that although PhoA-His was expressed and stable in the cytoplasm via CyDisCo, it was not correctly folded presumably due to inability to form native DSBs or to form DSBs in the first place. This implied that PhoA was exported in an inactive state by Tat and gained its activity in the periplasm likely due to the action of the natural Dsb pathway (section 1.4.1.1.).

Without CyDisCo, more soluble material was observed in the cytoplasm with $PaoA_{SP}$ than with $TorA_{SP}$ demonstrating that signal peptides can influence specifically the expression of recombinant proteins. With $PaoA_{SP}$, PhoA-His was exported to the periplasm via Tat but a small amount was observed in the periplasmic fraction of the Tat-null strain (Figure 6-2). This indicates that this export is not wholly Tat specific and correlates with data observed previously (Figure 4-6). It appears that two forms of periplasmic PhoA-His are present in the WT strain compared to Tat-null, which correlate with the molecular weights of the premature ($PaoA_{SP}$-PhoA-His, 53.7 kDa) and mature (PhoA-His, 48.0 kDa) proteins. Since a similar profile was observed with $OmpA_{SP}$ and $TorA_{SP}$, it was evident that the lack of signal peptide cleavage by the signal peptidase (section 1.3.) was not related to the secretion

pathway. The cause may come from the accessibility and number of the signal peptidase not being optimal leading to the partial export of recombinant proteins by an unknown mechanism (discussed in section 1.3.4.). PhoA is a natural Sec substrate and when fused to the non-natural Sec signal peptide $OmpA_{SP}$ was observed to give the highest export yield of active proteins as anticipated (Figure 6-2, Figure 6-3). Export was observed in both WT and Tat-null strains confirming translocation via the general and Tat-independent Sec pathway. Unexpectedly though, in the presence of CyDisCo, increased protein levels were detected in the cytoplasmic and insoluble fractions reinforcing the theory that this technology improves protein expression. However, cytoplasmic activity was not detected consistent with the hypothesis that DSBs may be shuffled into non-native states (Figure 6-3). The exported PhoA was observed to be active in the WT strain, presumably due to post-translocation modifications as discussed previously.

In the absence of a signal peptide and CyDisCo, PhoA-His was only recovered in the insoluble fraction in a very faint band (Figure 6-2). This result was expected as PhoA expressed without signal peptide was reported to be inactive in the cytoplasm and to quickly degrade (Michaelis et al., 1983, Manoil and Beckwith, 1985, Hoffman and Wright, 1985). With CyDisCo, however, the recombinant protein appeared more stable in solution and remained in the cytoplasm as anticipated. Here again, no activity was observed from the cytoplasmic proteins.

PhoA-His was exported specifically by the Tat pathway when fused to $TorA_{SP}$ and co-expressed with CyDisCo and was active in the periplasm as published (Matos et al., 2014). However, a significant amount of proteins was found in the cytoplasm which were inactive (Figure 6-3). Unfortunately, Matos et al. (2014) did not analyse the cytoplasmic activity. This suggest that the Tat exported PhoA-His became active in the periplasm after translocation. In this case, PhoA-His became sufficiently folded by the help of CyDisCo to be accepted by Tat but was still in a non-native conformation indicating a failure from the proofreading mechanism as discussed in section 1.3.2.5. Although CyDisCo worked partially, it provided the proteins to acquire a more stable conformation to be accepted by the proofreading system of Tat.

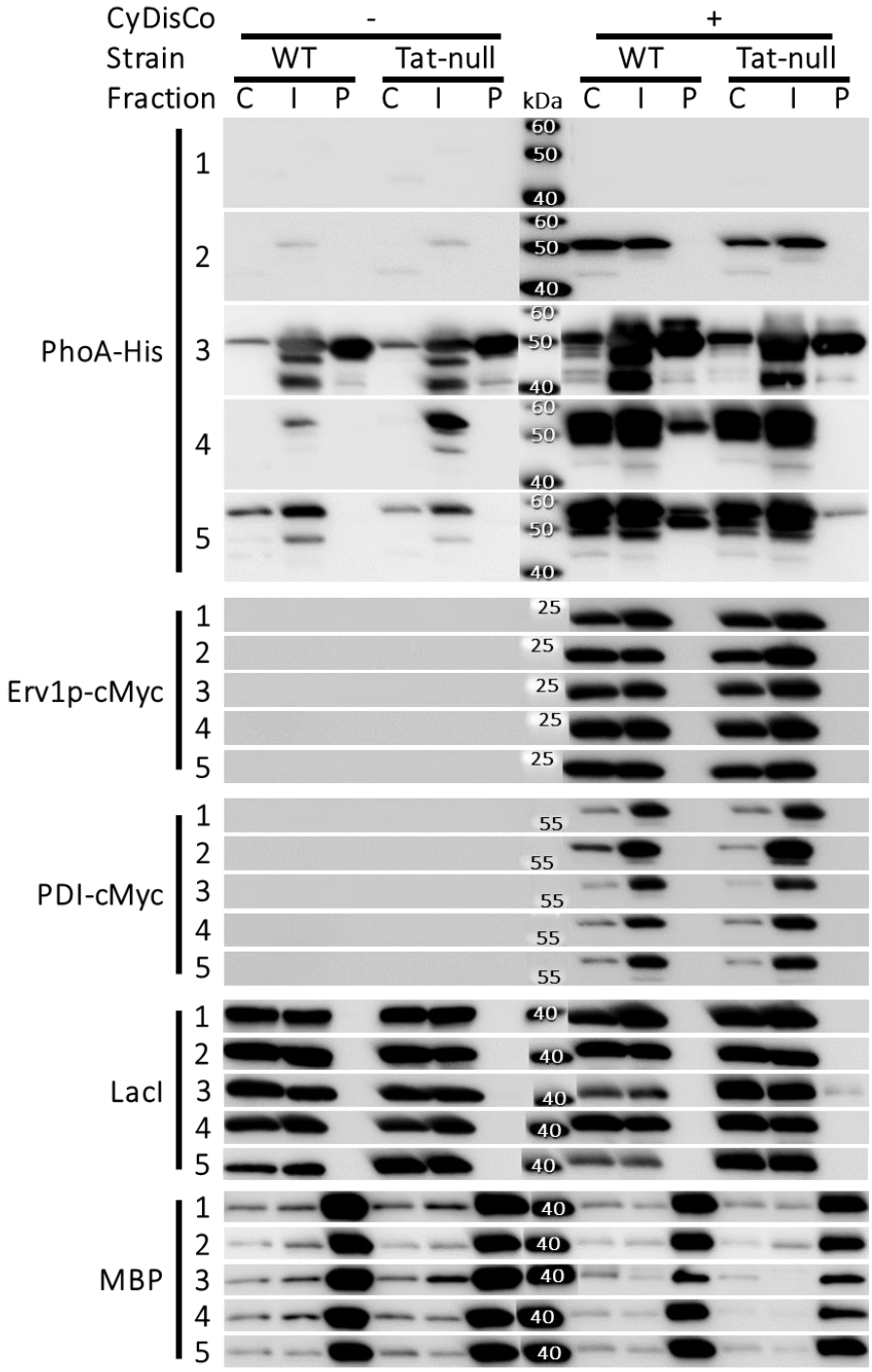


Figure 6-2: Cellular localisation of PhoA expressed in *E. coli*

An empty vector (1) or His-tagged PhoA with no signal peptide (2) or fused to $OmpA_{SP}$ (3), $TorA_{SP}$ (4) or $PaoA_{SP}$ (5) were expressed without (-) or with CyDisCo (+) in the MC4100 WT and Tat-null strains. The cells were harvested after 4 h induction at 30°C and fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. A. Western-blots of PhoA-His, Erv1p-cMyc, PDI-cMyc, LacI and MBP detected by anti-His, anti-cMyc, anti-LacI and anti-MBP antibodies respectively. PageRuler Plus Prestained Protein Ladder (Life Technologies) and MagicMark™ (Life Technologies) molecular weight markers are indicated in kilodalton (kDa). The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

Hatahet et al. (2010) found that purified PhoA was active in the cytoplasm of BL21(DE3) strain when co-expressed with Erv1p only. The difference with the present study could come from the co-expression of isomerase PDI which could theoretically shuffle the DSB into a non-native conformation explaining the absence of activity from the cytoplasmic PhoA-His. Moreover, when co-expressing $TorA_{SP}$-PhoA-His with Erv1p and DsbC, Matos et al. (2014) obtained active protein but did not specifically looked at the cytoplasmic activity. DsbC is an isomerase from *E. coli* (section 1.4.1.1.) whereas PDI is a human isomerase. The different origin between these enzymes could explain the still hypothetic shuffling issue. To test this hypothesis, the strain SHuffle was used since it expresses cytoplasmic DsbC (section 1.4.1.2. and Table 2-1).

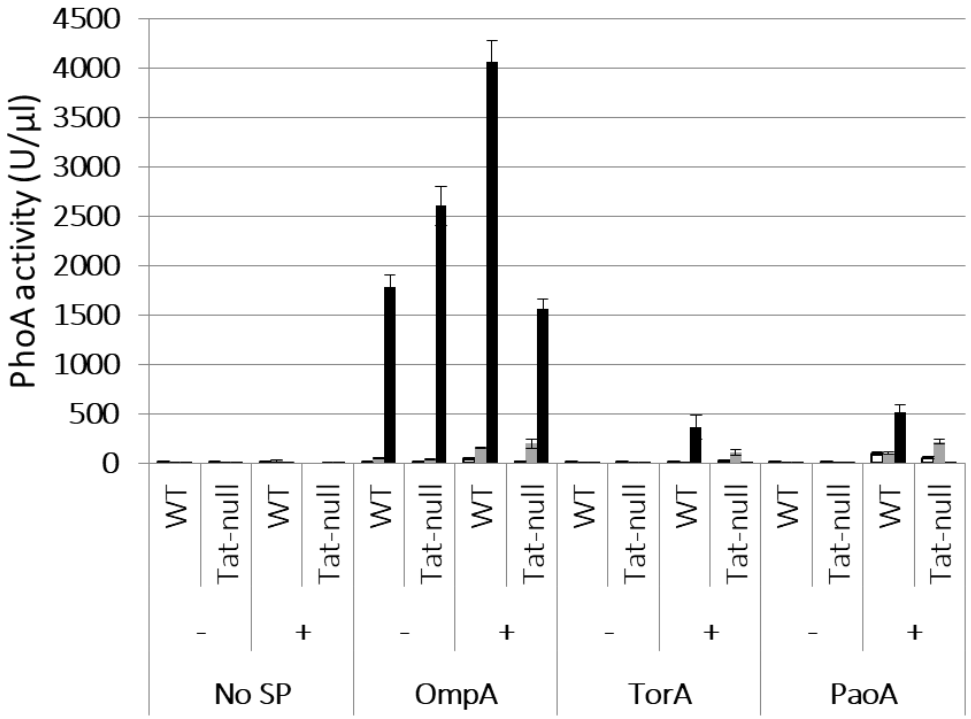


Figure 6-3: Activity of PhoA expressed in *E. coli*

The samples collected in Figure 6-2 were also used to determinate the activity of PhoA. Briefly, an empty vector (1) or His-tagged PhoA with no signal peptide (2) or fused to $OmpA_{SP}$ (3), $TorA_{SP}$ (4) or $PaoA_{SP}$ (5) were expressed without (-) or with CyDisCo (+) in the MC4100 WT and Tat-null strains. The cells were harvested after 4 h induction at 30°C and fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. The PhoA activity assays were performed in triplicate and measured against a titration curve using a commercially available *E. coli*-derived PhoA as a standard. The background signals were eliminated by subtracting the activity value from the empty vector's fractions to each corresponding fraction. The bars indicate the cytoplasmic (white), insoluble (grey) and periplasmic (black) fractions.

6.3.2. The impact of CyDisCo on the disulphide bonds of alkaline phosphatase

In section 6.3.1., the Tat export of PhoA-His was observed with the Tat signal peptide $TorA_{SP}$ and in the presence of the CyDisCo components Erv1p and PDI. However, disparities in PhoA activity demonstrated in the cytoplasmic and periplasmic fractions (Hatahet et al., 2010, Matos et al., 2014) suggest that CyDisCo is not responsible for the active state of these proteins. To test the origin of the cytoplasmic activity of PhoA-His, the strain SHuffle was used since it expresses cytoplasmic DsbC (section 1.4.1.2. and Table 2-3). The proteins locations were validated with the LacI and MBP controls as well as PDI-cMyc and Erv1p-cMyc located in their respective fractions (Figure 6-4).

Effective expression of cytoplasmically targeted PhoA-His (no signal peptide) required co-expression of either cytoplasmic Erv1p and PDI (CyDisCo) or DsbC (SHuffle, Figure 6-4). The WT MC4100 strain showed no soluble cytoplasmic PhoA-his and only traces of insoluble protein. In all situations, no meaningful PhoA activity was detected (Figure 6-5). Hence, neither CyDisCo nor cytoplasmic DsbC appear to achieve the properties claimed for them.

The Sec signal peptide of OmpA directs expressed PhoA-His to the periplasm and the periplasmic material has high levels (3.8-7.3 kU/L) of enzymatic activity (Figure 6-4, Figure 6-5). These results were expected since PhoA is a natural Sec substrate of *E. coli*. Moreover, the lower bands present on anti-His do not correlate to the mature protein since $OmpA_{SP}$ weights 2.1 kDa, but are likely to be truncation products.

With $TorA_{SP}$, no periplasmic export of PhoA-His was achieved in the WT cells without CyDisCo (Figure 6-4). However, co-expression of cytoplasmic DsbC or CyDisCo or DsbC and CyDisCo resulted in detectable periplasmic PhoA-His, with activity being seen in all three situations. For the first time, soluble cytoplasmic enzyme activity was detected in the SHuffle cells (Figure 6-5), but it was not established whether this was related to the premature or the mature forms of the protein. This suggest that conferring activity in the cytoplasm is signal peptide dependent. To understand how the presence of $TorA_{SP}$ led to PhoA-His activity in the SHuffle strain without CyDisCo, a screen of signal peptides and $TorA_{SP}$ mutants might shed light on the impact of the signal peptide on DsbC's activity.

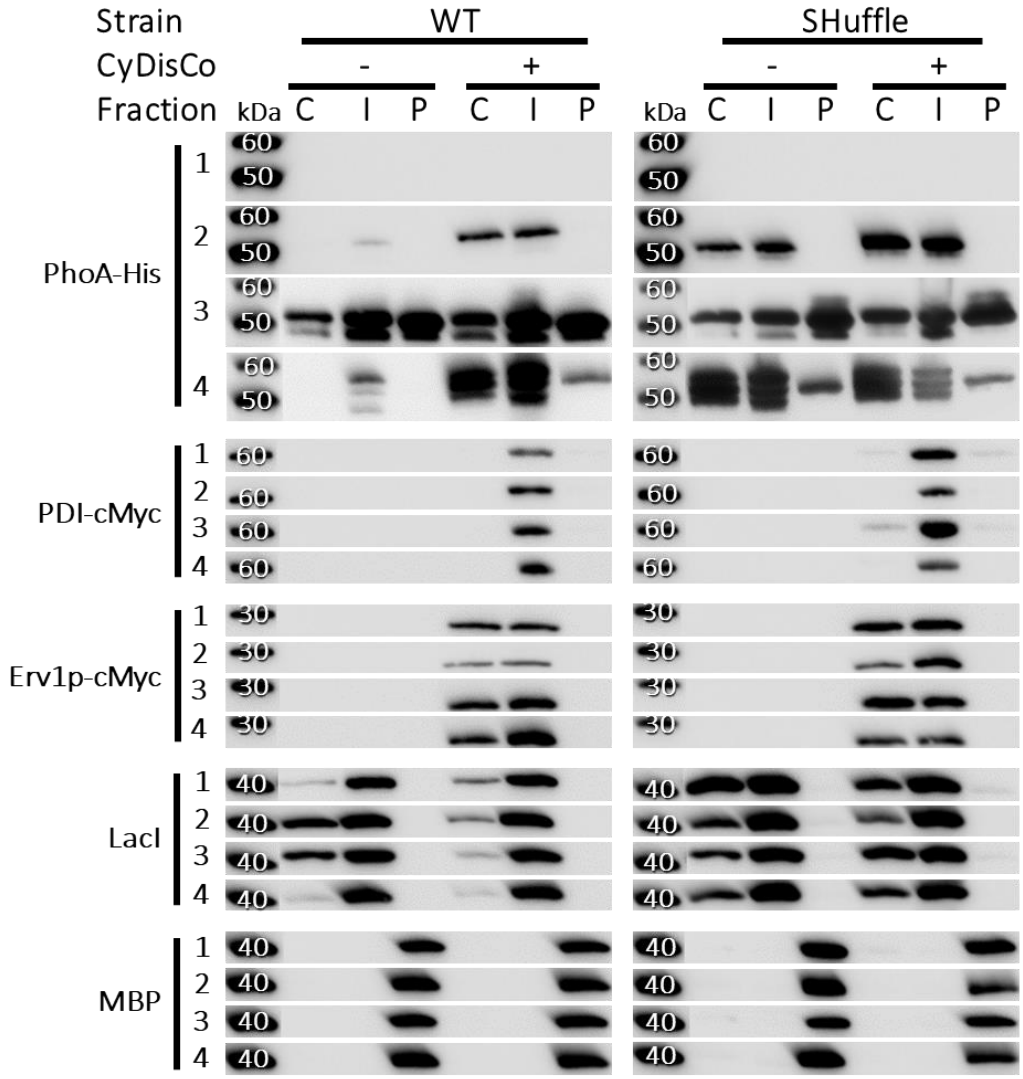


Figure 6-4: Effect of the cytoplasmically expressed disulphide isomerase DsbC on the localisation of PhoA

An empty vector (1) or His-tagged PhoA with no signal peptide (2) or fused to $OmpA_{SP}$ (3) or $TorA_{SP}$ (4) were expressed without (-) or with CyDisCo (+) in the MC4100 WT and SHuffle strains. The cells were harvested after 4 h induction at 30°C and fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. A. Western-blots of PhoA-His, PDI-cMyc, Erv1p-cMyc, LacI and MBP were detected by anti-His, anti-cMyc, anti-LacI and anti-MBP antibodies respectively. Molecular weight markers are indicated in kilodalton (kDa). The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

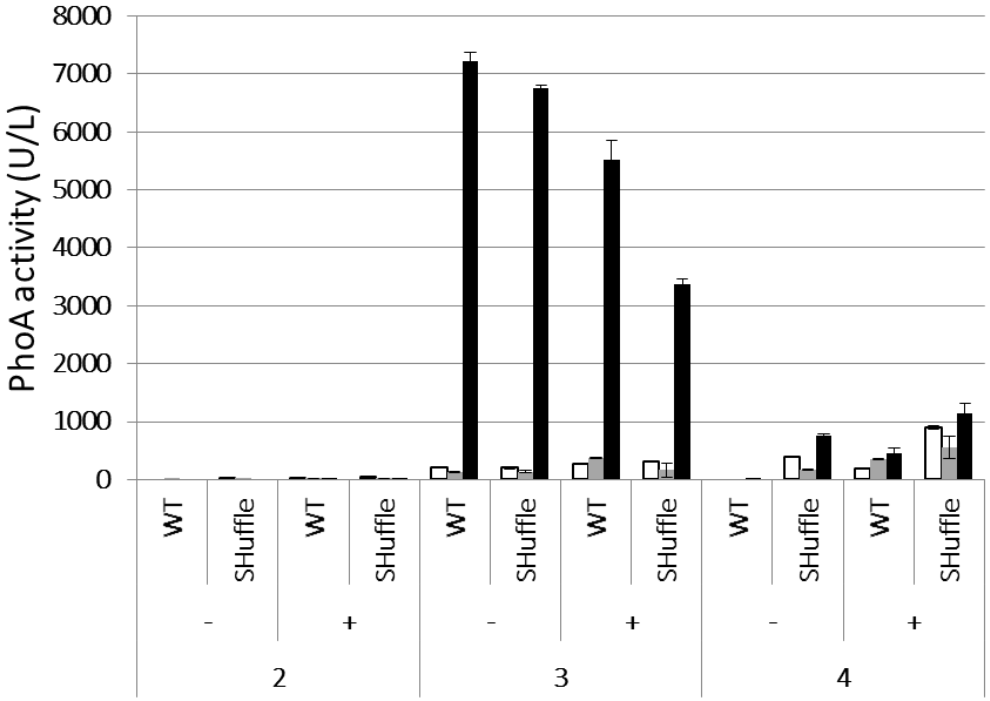


Figure 6-5: Effect of the cytoplasmically expressed disulphide isomerase DsbC on the activity of PhoA

The samples collected in Figure 6-4 were also used to determinate the activity of PhoA. Briefly, an empty vector (1) or His-tagged PhoA with no signal peptide (2) or fused to $OmpA_{SP}$ (3) or $TorA_{SP}$ (4) were expressed without (-) or with CyDisCo (+) in the MC4100 WT and SHuffle strains. The cells were harvested after 4 h induction at 30°C and fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. The PhoA activity assays were performed in triplicate and measured against a titration curve using a commercially available *E. coli*-derived PhoA as a standard. The background signals were eliminated by subtracting the activity value from the empty vector's fractions to each corresponding fraction. The bars indicate the cytoplasmic (white), insoluble (grey) and periplasmic (black) fractions.

6.4. Superfolder green fluorescent protein

GFP is a widely used reporter protein due to its fluorescence activity which can be observed in living cells for gene expression (Chalfie et al., 1994) and protein localisation studies (Feilmeier et al., 2000). GFP was first identified from the jellyfish *Aequorea Victoria* and its exceptional properties quickly raised interest (Prasher et al., 1992). This protein has a unique β-barrel structure which allows it to emit a green light when excited at the appropriate wavelength (section 2.5.2.3.). It has no DSB, is monomeric and has a size of 26.9 kDa. In *E. coli*, GFP was reported to export via the Sec pathway but remained inactive in the periplasm whereas the cytoplasmic protein was active (Feilmeier et al., 2000). However, successful Tat export leading to active periplasmic GFP has been reported in *E. coli* and optimized to up to 1.1 g/L of active protein purified from the periplasm after fermentation

(Matos et al., 2012). Furthermore, a GFP variant was created, named superfolder green fluorescent protein (sfGFP), to improve the solubility of the recombinant partner in reporter fusions thanks to six mutations (Pedelacq et al., 2006). This sfGFP mutant was exported to the periplasm via the Sec pathway (Aronson et al., 2011, Dinh and Bernhardt, 2011) but no report of Tat translocation has been published. However, Zhang et al. (2017) described the "auto-secretion" phenomenon of sfGFP where the protein localised in the cytoplasm, periplasm, outer membrane and extracellular medium without a signal peptide in *E. coli* (detailed in section 1.3.4.). The experiment was pursued here to establish the reproducibility of the data and the suitability of the sfGFP as a reporter protein.

In a preliminary experiment to investigate whether sfGFP can be stably produced in our expression system, His-tagged sfGFP were built with no signal peptide or in fusion with $OmpA_{SP}$, $PaoA_{SP}$ or $TorA_{SP}$. They were expressed in the MC4100 WT and Tat-null strains at 30°C, induced for 2 or 4 h at 30°C. They were immediately observed under a fluorescence microscope (section 2.6.) and these cells were fractionated using the PureFrac method including the media fraction.

sfGFP-His with no signal peptide, was observed in all cellular fractions including the periplasm irrespective of strain, in support of published data (Figure 6-6) (Zhang et al., 2017). The exported protein was found to be active whereas the cytoplasmic fraction was surprisingly not active for reasons unknown (Figure 6-7). Fluorescence microscopy of induced cells at 2 h and 4 h showed no distinctive features to differentiate between cytoplasmic and exported sfGFP (Figure 6-8). The only difference observed was morphological in relation to WT and Tat-null strains where Tat-null was observed in chains consistent with published data (section 3.6.). When the signal peptides $OmpA_{SP}$ or $PaoA_{SP}$ targeting the respective Sec and Tat export pathways was fused at the N-termini of sfGFP-His, a similar result to sfGFP-His with no signal peptide was observed with respect to protein localisation, activity and fluorescence as expected for $OmpA_{SP}$-sfGFP-His (Dinh and Bernhardt, 2011, Aronson et al., 2011). The exception is with $OmpA_{SP}$, where sfGFP-His was also found active in the media fraction after 4h. However, LacI, a cytoplasmic host marker was also present in the periplasmic and media fractions indicating cellular lysis. With $PaoA_{SP}$, the bands corresponding to the mature size of sfGFP-His in the periplasmic fractions are significantly less intense in the Tat-null than in the WT

strains. This suggests that these proteins were exported by Tat which allowed the signal peptides to be processed. The bands corresponding to the premature size in the periplasmic fractions were most likely exported by the same unknown mechanism which treated the signal peptide-less version of sfGFP-His. The same observation goes with $OmpA_{SP}$ where a band matching the size of the premature proteins were located in the periplasmic fractions.

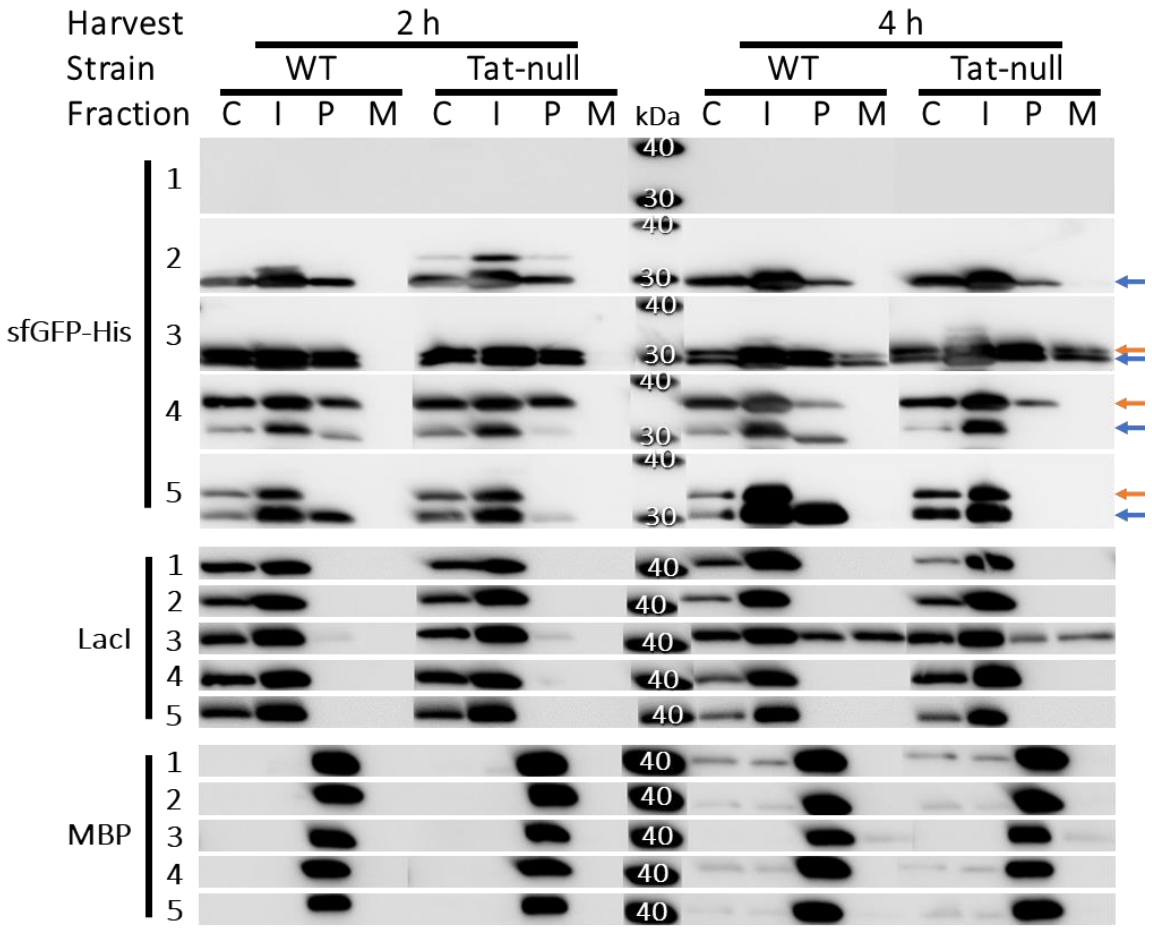


Figure 6-6: Western-blot analysis for the evaluation o sfGFP as a reporter

An empty vector (1) or His-tagged sfGFP with no signal peptide (No SP) (2) or fused to $OmpA_{SP}$ (3), $PaoA_{SP}$ (4) or $TorA_{SP}$ (5) were expressed in the MC4100 WT and Tat-null strains. The cells were harvested after 2 and 4 h induction at 30°C and fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I), periplasmic (P) and media (M) fractions. Western-blots represent the detection of PhoA-His, LacI and MBP using anti-His, anti-LacI and anti-MBP antibodies respectively. The arrows on the right indicate the mature (black) and premature (grey) forms of the proteins. Molecular weight markers are indicated in kilodalton (kDa). The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

Conversely with $TorA_{SP}$, a different behaviour was observed. sfGFP-His was not exported by the unknown mechanism when fused to $TorA_{SP}$ since no band matching the premature size was detected in the periplasmic fractions (Figure 6-6, Table 6-1). However, regarding the mature size bands, active sfGFP-His was found in the periplasmic fraction of the WT cells and a considerably lower amount of protein was detected in the Tat-null strain but appeared to be active (Figure 6-6, Figure 6-7). These observations revealed that most of the periplasmic proteins were exported via the Tat pathway but the remaining exported proteins most likely came as cytoplasmic mature proteins through the unknown mechanism. Moreover, the fluorescence was localised as discrete bodies around the cellular periphery and in particular, at the poles (Figure 6-8). Polar localisation of other Tat-targeted GFP variants were reported before with a periplasmic distribution of the fluorescence but a stronger signal at the poles (Santini et al., 2001, Ize et al., 2002). The localisation of Tat substrates at the pole of *E. coli* was speculated to be due to the thicker periplasmic space at the cell poles (Bernhardt and De Boer, 2003) or due to the polar localisation of the majority of the Tat translocase components (Berthelmann and Bruser, 2004, Ray et al., 2005).

Table 6-1: sfGFP protein sizes

Sizes before (premature) and after (mature) signal peptide cleavage of the sfGFP proteins expressed in Figure 6-6, Figure 6-7 and Figure 6-8.

Number	Construct		Size (kDa)	
	Signal peptide	Gene	Premature	Mature
2	No SP	sfGFP-His	N/A	27.5
3	$OmpA_{SP}$	sfGFP-His	29.6	27.5
4	$PaoA_{SP}$	sfGFP-His	33.2	27.5
5	$TorA_{SP}$	sfGFP-His	32.2	27.5

Pedelacq et al. (2006) observed a loss of GFP fluorescence due to the sonication used to lyse the whole cells which could explain the absence of activity in the cytoplasmic and insoluble fractions in the present study since the fractionation method employed on the periplasmic-extracted cells

involved sonication (Appendix 4). Overall, sfGFP is not a valid candidate to use as a reporter protein as it was confirmed to export via an unknown mechanism which did not appear to involve Tat.

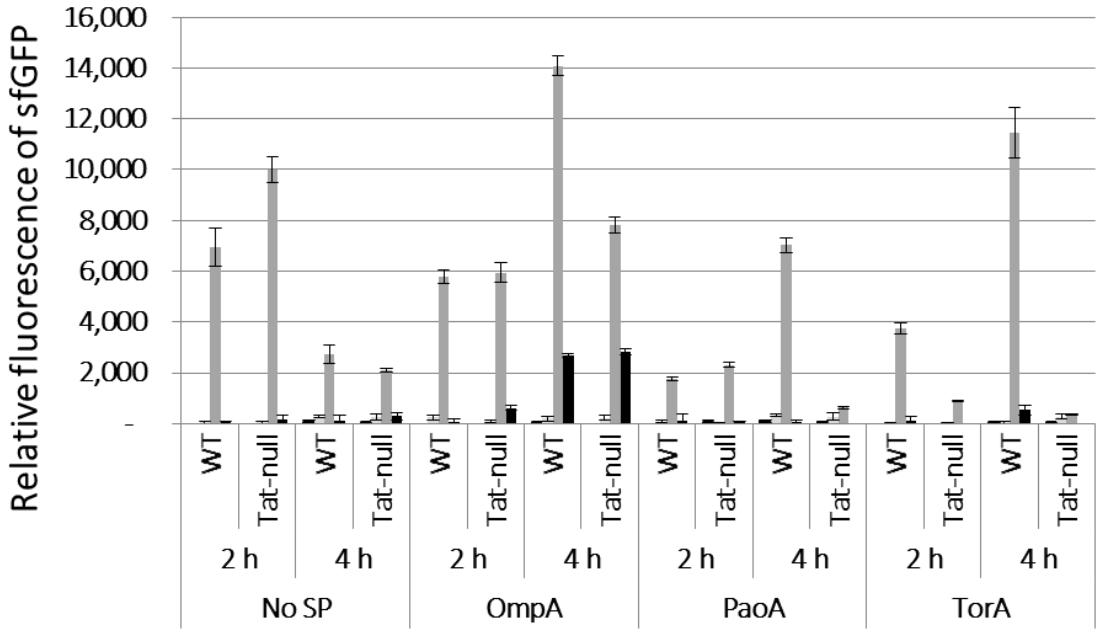


Figure 6-7: Activity analysis for the evaluation of sfGFP as a reporter

The activity assays were performed on the samples obtained as detailed in Figure 6-6 were in triplicate and represented the relative fluorescence compared to the empty vector. The bars indicate the cytoplasmic (white), insoluble (light grey), periplasmic (dark grey) and media (black) fractions.

6.5. Human growth hormone

hGH is a 24.9 kDa secreted protein possessing two DSB and is monomeric in solution but dimerizes upon binding to its receptor (Li, 1982, Walsh et al., 2003). This protein was historically produced in the cytoplasm of *E. coli*. However, this method was limited by the accumulation of the difficult to resolubilise inclusion bodies, a refolding step necessary to form the DSB and the remaining N-terminal methionine increased immunogenicity in patients treated with this hormone (Becker and Hsiung, 1986, Glasbrenner, 1986, Soares et al., 2003). To counter these issues, hGH was produced in the periplasm of the bacteria where the DSB formed thanks to the Dsb pathway (section 1.4.1.1.) and hence, less inclusion bodies were observed (Gray et al., 1985). Sec export of hGH was reported with $OmpA_{SP}$ to 10-15 mg/L (Becker and Hsiung, 1986). More recently, translocation of this protein was published via the Tat pathway in *E. coli* when fused to $TorA_{SP}$ (Alanen et al., 2015). Interestingly,

the authors demonstrated that hGH was translocated even without its DSB suggesting that the apo-protein adopted a near-native conformation accepted by the Tat proofreading mechanism (Bewley et al., 1969, Youngman et al., 1995). Alanen et al. (2015) also demonstrated that hGH acquired its DSB in the periplasm after translocation but that the Tat transport event was not impaired if the DSB were cytoplasmically formed. This protein is therefore an interesting candidate as a reporter protein for Tat expression.

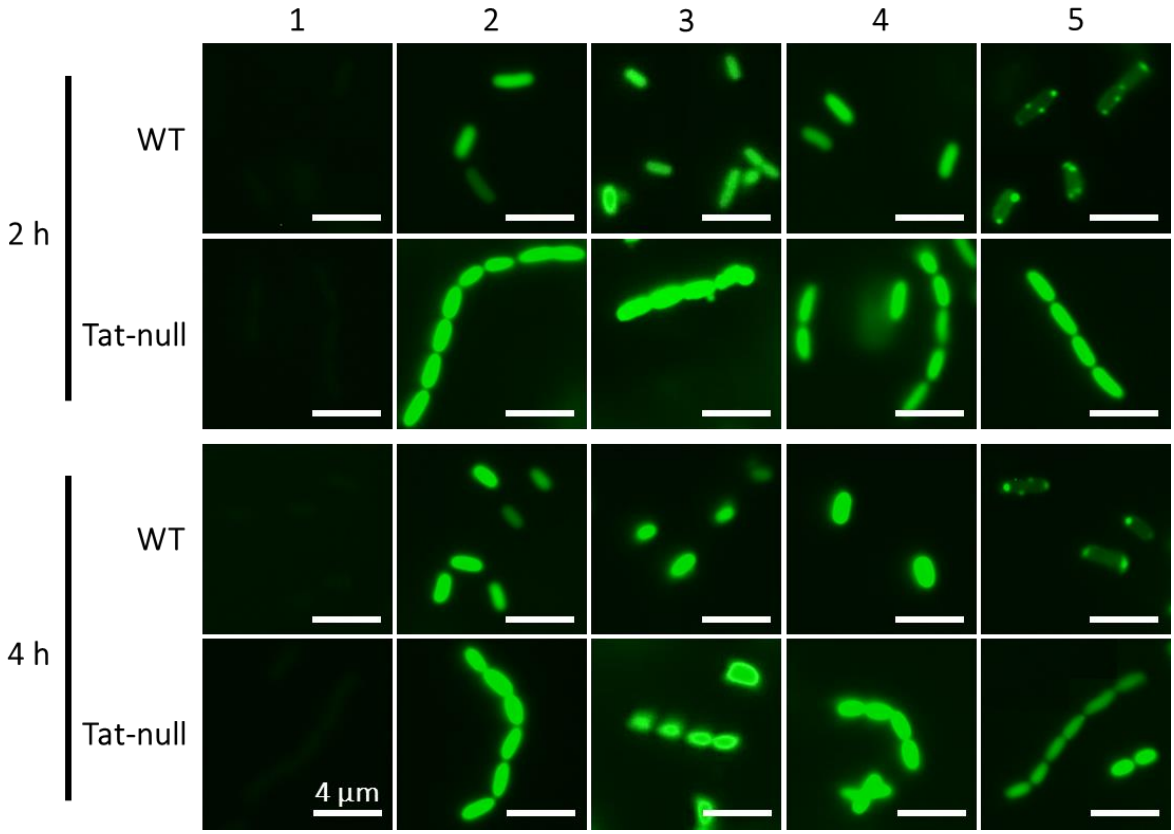


Figure 6-8: Microscopic observations of *E. coli* expressing sfGFP

An empty vector (1) or His-tagged sfGFP with no signal peptide (2) or fused to $OmpA_{SP}$ (3), $PaoA_{SP}$ (4) or $TorA_{SP}$ (5) were expressed in the MC4100 WT and Tat-null strains. The cells were harvested after 2 and 4 h induction at 30°C and immediately observed under a fluorescence microscope with a 1000-fold magnification. The scale bar is 4 µm.

To test its suitability, the gene coding for the hGH protein (*gh1*) was cloned into pTTO without its natural 26 bp mammalian signal peptide and with a C-terminal His tag (Table 2-2). Variants were then built to bear the N-terminal signal peptides $OmpA_{SP}$ or $TorA_{SP}$. These three plasmids including an empty vector as the negative control were expressed in MC4100 WT and Tat-null strains, with

and without CyDisCo. After 2 h induction at 30°C (Table 2-3, condition 1), the cells were fractionated according to the PureFrac protocol into the cytoplasmic, insoluble and periplasmic fractions (Figure 6-9). Surprisingly, hGH-His with no signal peptide was detected in the periplasm of all conditions with an apparent higher amount with CyDisCo. This suggested the involvement of an unknown transport mechanism since no N-terminal export signal peptide was identified with the prediction software SignalP (Nielsen, 2017a). No previous report of this hGH "auto-export" was found and this phenomenon was not artefactual since the localisation of the Erv1p, PDI, LacI and MBP controls indicated correct fractionation. A parallel can be made with the "auto-secretion" of sfGFP (section 6.4.) but no evidence so far suggested that both proteins were transported via the same mechanism. This discovery is worth investigating further to see if a protein of interest can be piggybacked into the periplasm when fused to hGH-His but was outside the scope of this project. Furthermore, the candidate protein appeared in the periplasmic fractions in all tested conditions with $OmpA_{SP}$ indicating Sec export was not influenced by CyDisCo. But with this signal peptide, fraction impurities were observed where the periplasm was contaminated by the cytoplasmic PDI, Erv1p and LacI proteins. The fact that such cross-contamination was not observed in any other condition led to the conclusion that the expression of hGH-His with $OmpA_{SP}$ compromised the cell envelope which was consequently more sensitive to the fractionation. The periplasmic hGH-His could hence come from a compromised envelope, the unknown mechanism observed with the no signal peptide variant or Sec export. With $TorA_{SP}$, hGH-His were detected in the periplasmic fractions of the WT but not the Tat-null strains irrespective of CyDisCo confirming that hGH-His can be Tat-dependently exported even when the DSB are not formed prior to translocation Overall, Tat-specific export was proven thanks to $TorA_{SP}$ making this protein a suitable reporter protein for Tat export and promising transport alternative to control with no signal peptide.

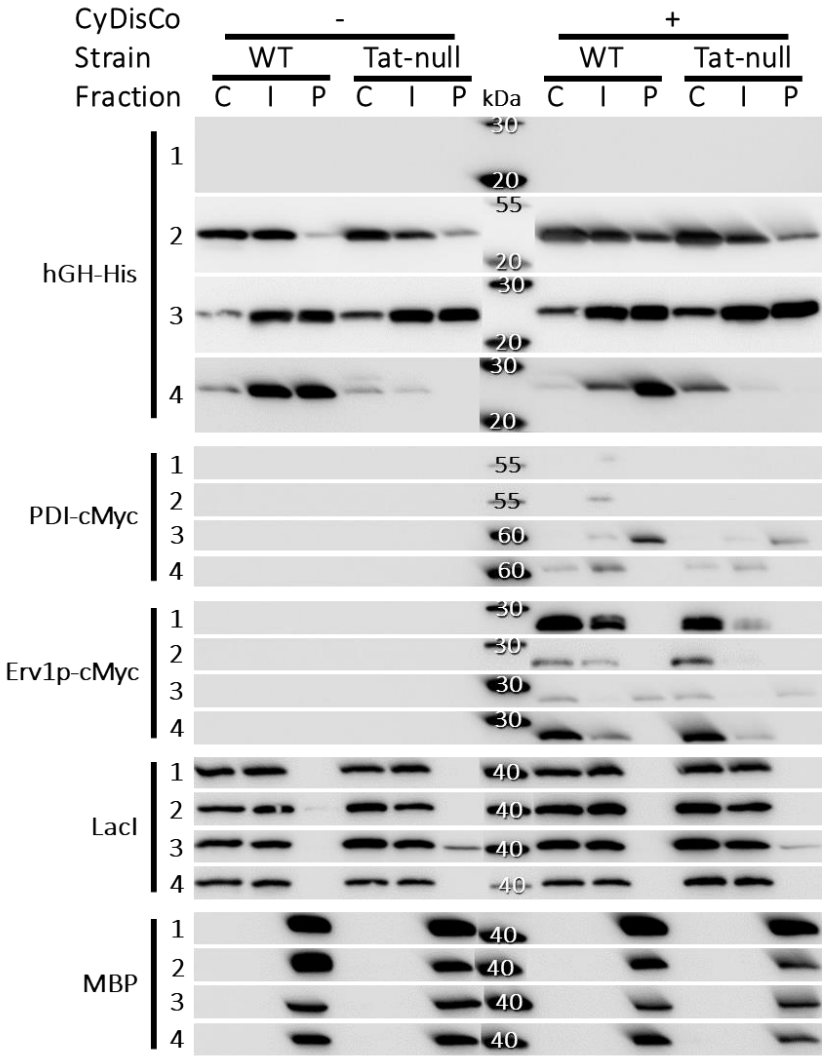


Figure 6-9: Tat-independent and dependent exports of human growth hormone

An empty vector (1) or His-tagged hGH with no signal peptide (2) or fused to $OmpA_{SP}$ (3) or $TorA_{SP}$ (4) were expressed in the MC4100 WT and Tat-null strains. The cells were harvested after 2 h induction at 30°C and fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I), periplasmic (P) and media (M) fractions. The Western-blots represents the detection of hGH-His, LacI and MBP using anti-His, anti-LacI and anti-MBP antibodies respectively. PageRuler Plus Prestained Protein Ladder (Life Technologies) and MagicMark™ (Life Technologies) molecular weight markers are indicated in kilodalton (kDa). The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

6.6. Conclusion on valid reporter proteins

Out of the four proteins tested, only two came out as valid reporter proteins. PhoA was exported specifically via Tat using when fused to the signal peptide of TorA and found active in the periplasm but only when co-expressed with CyDisCo. It appeared that the homodimer form of PhoA was not an issue for translocation and PhoA was considered to be exported via Tat as a monomer since all

periplasmic PhoA appeared to be leaderless. The second valid candidate was hGH which has shown Tat-dependant export with $TorA_{SP}$. Interest was also raised from a different feature: hGH was observed to export without the need for a signal peptide. Both characteristics made this protein an interesting candidate for periplasmic secretion. On the other hand, β-galactosidase and sfGFP were also considered but not selected. The former showed no evidence of export with any of the tested signal peptides. In the case of the Tat pathway, β-galactosidase was believed to assemble into a multimeric structure prior to reaching the translocase. However, including the signal peptide of alkaline phosphatase D from *Bacillus subtilis* in this screen would have led to a more complete set of data since it was reported to allow export of β-galactosidase in this strain by Xia et al. (2010) and Ren et al. (2016). The sfGFP protein demonstrated a Tat-specific export profile with the $TorA_{SP}$ but only at a specific harvest time. This protein, similarly to hGH, presented periplasmic export even without a signal peptide and evidence suggested that sfGFP fused to $OmpA_{SP}$ and $PaoA_{SP}$ was partially exported independently to the Sec and Tat pathway respectively. Moreover, this protein has already been assessed to carry proteins of interest across the inner membrane of *E. coli* (Zhang et al., 2017). To summarize, PhoA and hGH were selected to use as reporter protein in the fusion strategy.

7. EVALUATION OF NON-NTS PROTEINS AS CARRIER TO FACILITATE TAT EXPORT

7.1. Modifying the strategy by using PhoA and hGH as soluble carrier proteins

The original hypothesis was to investigate whether an NTS could be used as a fusion partner to facilitate the correct folding and subsequent export of a C-terminally linked reporter or recombinant protein into the periplasm via the Tat pathway (Chapter 3). As described in Chapter 5, the five initial NTS candidates selected for study were either poorly expressed and/or could not be exported into the periplasm via Tat when linked to their own signal. Furthermore, strain selection, optimisation of growth conditions, cofactors addition, co-expression of partners/chaperone, expression vector and induction level did not improve export efficiency of the NTS in the case of PaoA (section 5.2.) and thereby highlighted that this process is low throughput, time consuming and does not guarantee success. Successful Tat export of PaoA was reported (Hatahet et al., 2010, Matos et al., 2014) but this published data was found to be not reproducible in the present study despite exploring several avenues for optimisation. The remaining NTS of *E. coli* mentioned in Appendix 1, did not pass the selection criteria. Gathering the fundamental information necessary to export these NTS via Tat via informative deductions and technical evidence would require more time than allocated for this project.

On the other hand, experiments to select potential reporter proteins identified two proteins, PhoA and hGH that both presented properties that would inadvertently be considered as good characteristics of a potential soluble partner (Chapter 6). Thus, the initial hypothesis was modified to utilize these reporter proteins as the soluble partner instead of an NTS. Both PhoA and hGH proteins are relatively small (49.4 and 24.9 kDa respectively), allowing the fusion of a broad range of recombinant protein of varying size with limits for export. Moreover, both were specifically exported as soluble proteins via Tat thanks to the $TorA_{SP}$ and in the case of PhoA, also as active proteins.

Indeed, PhoA and hGH were detected in uncontaminated periplasmic fractions in WT strain when fused to $TorA_{SP}$ while no export was observed in the Tat-null strain (section 6.3. and 6.5.).

Thus, design strategy was revised to export proteins of interest in fusion with PhoA or hGH as the soluble partner bearing the signal peptide of TorA. First, the validated soluble partners were fused to one another (hGH-PhoA and PhoA-hGH) as positive controls to test whether the independently-exported proteins can be translocated as a fusion as well as the fusion design itself. Then, relevant proteins of interest from UCB Celltech were applied: the human fatty acid binding protein 4 (FABP4) and two camelid-derived single domain VHH antibodies.

7.2. Evaluation of the fusion strategy

7.2.1. Design of the fusion proteins

The ability of PhoA (with CyDisCo) and hGH to be exported by Tat, as revealed by the lack of export with the Tat-null strain, indicated their capability to fold in the cytoplasm in a state accepted by the Tat proofreading mechanism (Chapter 6). Preliminary experiments were performed to evaluate whether these proteins could be exported when fused to one another as a control for the new fusion strategy. However, although PhoA was found to strictly require a signal peptide for export, hGH export is complicated by the observation that it was also translocated without a leader sequence by an unknown mechanism. Selected recombinant proteins were then used with hGH and PhoA as the soluble carrier to test Tat export of these fusions. The design of the fusion constructs can be found in Figure 7-1.

7.2.2. Export capability of PhoA using hGH as a soluble partner

The first fusion construct expressed was PhoA fused to hGH as the carrier protein as described in Figure 7-1, part 1. Ideally, the PhoA-hGH fusion (Figure 7-1, part 2) would also be expressed in parallel along with hGH-His and PhoA-His alone to validate export of the fusion partners individually. Unfortunately, the number of samples exceeded the maximum which can be handled in one run. Therefore, the fusions were expressed in separate experiments with their respective C-terminal

protein as export control. In this section, hGH-PhoA and PhoA-His with no signal peptide or fused with $OmpA_{SP}$ or $TorA_{SP}$ were expressed in MC4100 WT, with and without CyDisCo, at 30°C and induced for 2 h at the same temperature (Table 2-3, condition 1). The harvested cells were fractionated using the PureFrac method into cytoplasmic, insoluble and periplasmic fractions and analysed by reducing SDS-PAGE and immunological detected with relevant antibodies. The shifts in the band sizes between in conditions 2, 3 and 4 (Figure 7-2) are due to the theorical molecular weight sizes of the protein with no signal peptide, $OmpA_{SP}$ (2.1 kDa) and $TorA_{SP}$ (4.2 kDa) respectively (Appendix 7). With respect to detection of the exported proteins with anti-His and anti-FLAG, the pre-mature and mature proteins are indicated where possible. The latter indicates cleavage of the signal peptide by presumably the inner-membrane located signal peptidase, LepB. The presence of truncated species with both signal peptides in the cytoplasmic and periplasmic fractions is also notable in this data and will be discussed in section 7.5.

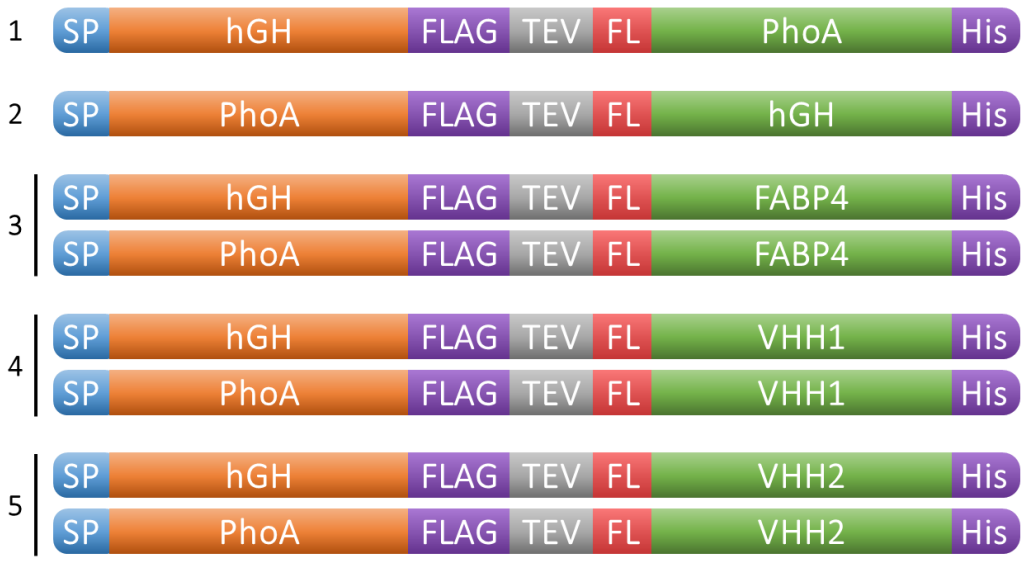


Figure 7-1: Design of the fusion constructs following the new hypothesis

The constructs were designed from N-terminal to C-terminal as follows: signal peptide (SP), either hGH or PhoA as soluble carrier protein, FLAG tag, TEV cleavage site, short flexible linker (FL), recombinant protein and His tag. The recombinant proteins tested were either PhoA, hGH, FABP4, VHH1 or VHH2.

The localisation (Figure 7-2, panel A) and activity of PhoA-His (Figure 7-3, panel A) were consistent with previous PhoA-His data as reported in section 6.3.1. Briefly, PhoA-His with no signal peptide was

degraded without CyDisCo whereas it was stabilized in the presence of CyDisCo (Figure 7-2, panel A, His, 2) although inactive (Figure 7-3, panel A), presumably through incorrect DSB formation. The protein localisation was validated since the Erv1p, PDI, LacI and MBP control proteins were all detected in their respective fractions (Appendix 6).

Effective expression and translocation of PhoA-His was achieved in fusion to $OmpA_{SP}$ (Figure 7-2, panel A, 3) suggesting that the signal peptide improved the expression of this protein. Indeed, Singh et al. (2013) demonstrated that signal peptides modulate the stability of recombinant proteins in a sequence-dependent manner. Moreover, the SecB chaperone binds to the signal peptide and the emerging polypeptide from the ribosome (section 1.3.1.2.). Thus, this chaperone prevents conformational changes through its antifolding activity (Huang et al., 2016) and ultimately leads to its rapid export via Sec. In addition, chaperones such as GroEL, DnaK and trigger factor (TF), were reported to be recruited by SecB and involved in OmpA preprotein processing by direct interaction, thus contributing to the antifolding activity and targeting of the preprotein to the inner membrane (Castanie-Cornet et al., 2014). $OmpA_{SP}$-PhoA-His was exported irrespective of CyDisCo and was active in the periplasm (Figure 7-2, panel A, 3; Figure 7-3, 3). As PhoA is a natural Sec substrate of *E. coli*, all soluble proteins were expected to be exported and here, the presence of cytoplasmic proteins suggests that the Sec pathway is saturated presumably due to overexpression of this protein. PhoA activity was only detected in the periplasmic fractions, indicating correct folding of the protein exported via the Sec pathway (Figure 7-3, panel B, 3).

With $TorA_{SP}$, PhoA-His was exported only when co-expressed with CyDisCo and was also active in the periplasm (Figure 7-2, panel A, His, 4 and Figure 7-3, panel A, 4). However, this protein was not active in the cytoplasmic fraction and suggests that PhoA was exported via Tat in a folded but non-native state and acquired its native conformation in the periplasm (detailed in section 6.3.). This strongly suggests that CyDisCo did not form cytoplasmic DSBs or that they were in a non-native state.

With respect to the hGH-PhoA fusion protein, without a signal peptide, this protein was detected in the periplasm by both anti-His and anti-FLAG (Figure 7-2, panel B, His and FLAG, 2). This was not wholly unexpected as previously it was shown that without a signal peptide, hGH was carried across

the inner membrane by an unknown mechanism (section 6.5.). The protein was translocated into the periplasm irrespective of CyDisCo, although CyDisCo did improve overall expression and amplify export (Figure 7-2, panel B, His, 2). The marker proteins LacI and MBP were detected in their respective fractions validating fraction purity (Appendix 6). Moreover, PhoA, in the C-terminal region of the fusion, presented no activity in any fraction (Figure 7-3, panel B). This suggests that although the full-length fusion protein adopted a stable conformation, this was not native for PhoA.

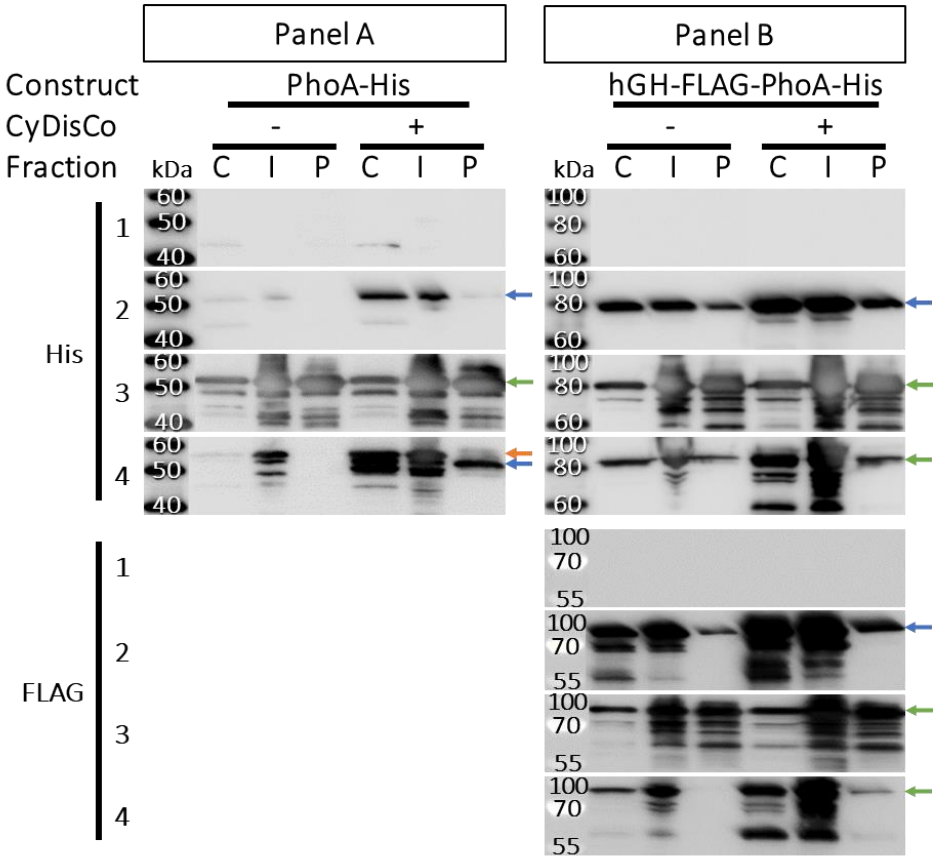


Figure 7-2: Localisation of hGH fusions with PhoA as the protein of interest

His-tagged PhoA or the fusion hGH-PhoA with no signal peptide (2) or fused with $OmpA_{SP}$ (3) or $TorA_{SP}$ (4) as well as an empty vector (1) were expressed with (+) and without (-) CyDisCo in the MC4100 WT strain. The cells were harvested after 2h induction at 30°C and fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. The Western-blots represent the detection of His-tagged and FLAG-tagged proteins using anti-His and anti-FLAG antibodies respectively. Arrows indicate the bands of premature (orange) and mature proteins (blue) when corresponding to the theoretical molecular weight size (Appendix 7). Green arrows indicate full-length protein where the presence of the absence of the signal peptide is not discernible. PageRuler Plus Prestained Protein Ladder (Life Technologies) and MagicMark™ (Life Technologies) molecular weight markers are indicated in kilodalton (kDa). The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

With fusion to the Sec signal peptide of OmpA, hGH-PhoA was detected in all fractions with both anti-His and anti-FLAG, indicating export irrespective of CyDisCo (Figure 7-2, panel B, 3). Though as the leaderless variant of this protein was shown to export via an unknown mechanism, it is difficult to confirm whether export was solely via Sec. Indeed, identifying exported premature and mature proteins is challenging due to the presence of truncated species and the large size of the fusion (Figure 7-2, panel B, 3; Appendix 7). Interestingly, the periplasmic proteins showed some PhoA activity (Figure 7-3, panel B, 3) although the levels were significantly lower (> 5-10 fold) compared to the corresponding independently expressed PhoA-His controls. However, due to a range of truncated species detected by anti-FLAG in the periplasmic fraction, it is difficult to confirm whether PhoA activity can be attributed to correctly folded, full-length proteins, or to truncated species that contain an intact PhoA (Figure 7-2, panel B, 3). The presence of periplasmic preproteins suggest cytoplasmic cross-contamination, however this is unlikely as the control proteins LacI and MBP were only found in their natural locations, although Erv1p and PDI were not (Appendix 6). The periplasmic localisation of the CyDisCo components when co-expressed with an $OmpA_{SP}$-recombinant protein has been a recurring and consistent observation throughout this study and will be discussed further in section 9.4. (Appendix 6).

In contrast, expression of this fusion with $TorA_{SP}$ led to the periplasmic localisation of full-length protein when co-expressed with CyDisCo (Figure 7-2, panel B, 4; Appendix 7). Due to the close molecular weights of pre-(87.9 kDa) and mature (84.7 kDa) proteins, it is difficult to ascertain the quality of the periplasmic bands (Appendix 7). Whilst in the cytoplasm, the truncated products are indicative of a lower stability, and only the full-length protein is exported. However, although the latter appears to be stable and fully intact, it is completely inactive, indicating a non-native conformation (Figure 7-3, panel B, 4). Without CyDisCo, the full-length protein was detected in the periplasmic and, interestingly, also in the cytoplasmic fraction (Figure 7-2, panel B, 4). The absence of truncated species in the cytoplasm contrasts starkly with co-expression of this fusion with CyDisCo. This suggests that the presence of these chaperones resulted in the reiteration of many and often unstable conformations, destined presumably for proteolytic cleavage. However, all proteins were

also inactive and point to non-native conformations (Figure 7-4, panel B, 4). To confirm whether the pathway used during this translocation event was Tat or the same unknown pathway as with the leader-less version, a test could be conducted comparing export in the WT versus the Tat-null strains. The amount of periplasmic proteins was higher with $OmpA_{SP}$ compared to no signal peptide or with $TorA_{SP}$ suggesting that the Sec pathway contributed to high level export of the Sec signal peptide fusion proteins.

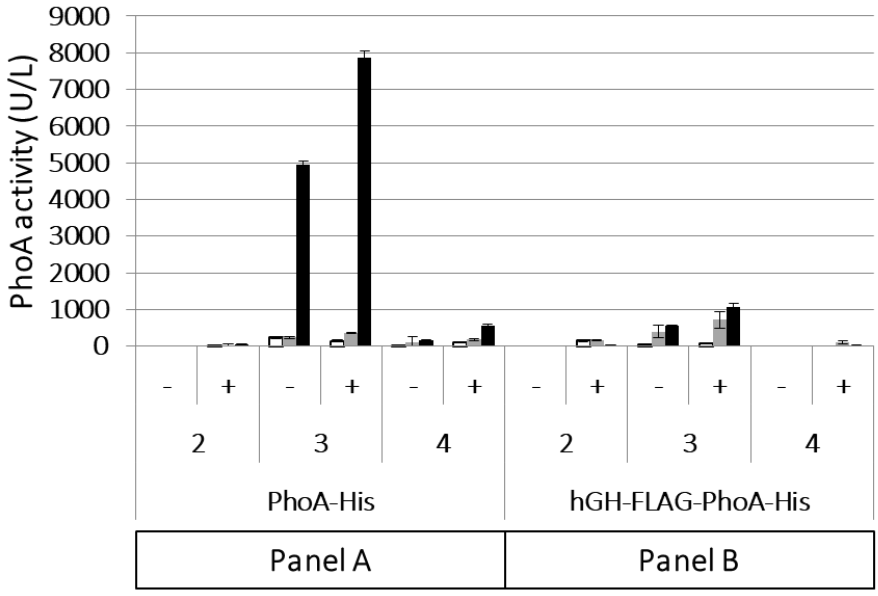


Figure 7-3: PhoA activity of hGH fusions with PhoA as the protein of interest

The PhoA activity assays were performed in triplicate and measured against a titration curve using a commercial enzyme. The background signals were eliminated by subtracting the activity value from the empty vector's fractions to each corresponding fraction. PhoA (panel A) and hGH-PhoA fusions (panel B) are represented as no signal peptide (2), fused with $OmpA_{SP}$ (3) or with $TorA_{SP}$ (4), and with (+) and without (-) CyDisCo. The bars indicate the cytoplasmic (white), insoluble (grey) and periplasmic (black) fractions.

7.2.3. Export capability of hGH using PhoA as a soluble partner

A similar set of constructs were built and expressed using hGH as the protein of interest and PhoA as the soluble partner (Figure 7-1, part 2). In this case, the aim was to test the export of hGH carried by PhoA. The localisation and PhoA activity results are represented in Figure 7-4 and Figure 7-5 respectively. The localisation profile of hGH-His which was used as a control (Figure 7-4, Panel A, His), was identical to previous data obtained in section 6.5. Briefly, hGH-His showed minimal traces

of the "auto-export" feature with no signal peptide and exported at a higher yield with $OmpA_{SP}$ and $TorA_{SP}$ in what appears to be the premature form (Appendix 7). The marker proteins were all located in their expected fractions justifying the purity of the isolated compartments except for the periplasmic-localised CyDisCo proteins which has been observed previously.

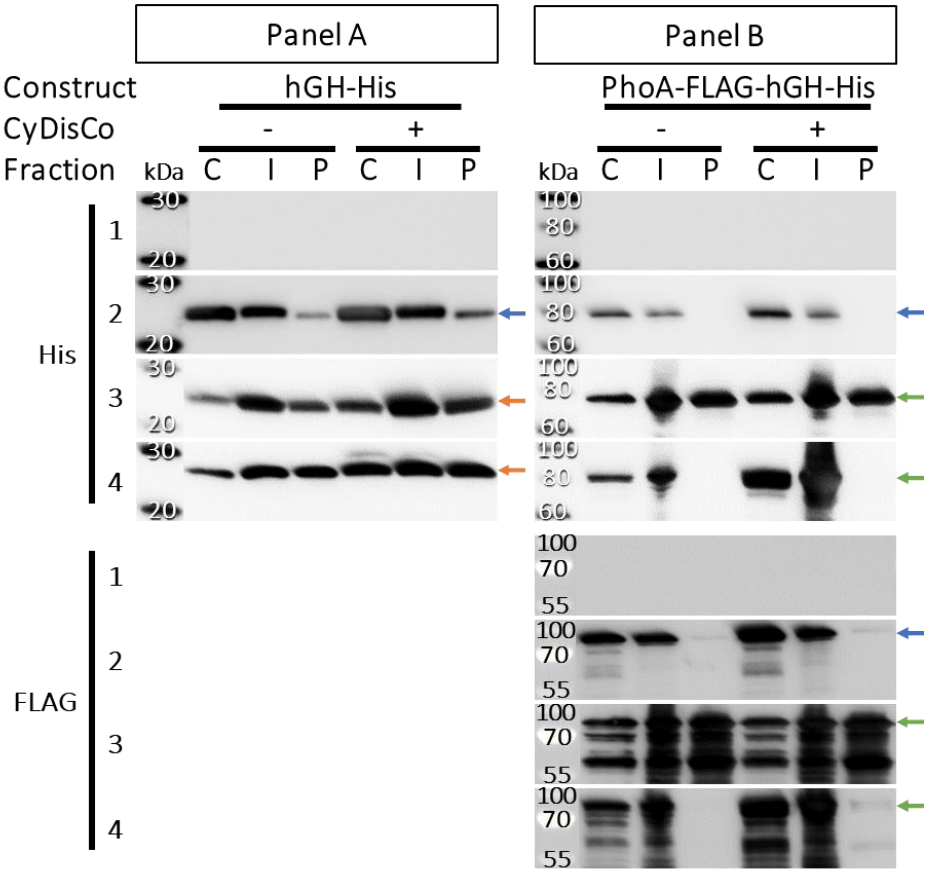


Figure 7-4: Localisation of PhoA fusions with hGH as the protein of interest

His-tagged hGH or the fusion PhoA-hGH with no signal peptide (2) or fused with $OmpA_{SP}$ (3) or $TorA_{SP}$ (4) as well as an empty vector (1) were expressed with (+) and without (-) CyDisCo in the MC4100 WT strain. The cells were harvested after 2h induction at 30°C and fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. The Western-blots represent the detection of His-tagged and FLAG-tagged proteins using anti-His and anti-FLAG antibodies respectively. Arrows indicate the bands of premature (orange) and mature proteins (blue) when corresponding to the theoretical molecular weight size (Appendix 7). Green arrows indicate full-length protein where the presence of the absence of the signal peptide is not discernible. PageRuler Plus Prestained Protein Ladder (Life Technologies) and MagicMark™ (Life Technologies) molecular weight markers are indicated in kilodalton (kDa). The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

With respect to the PhoA-hGH fusion, with no signal peptide, no export was detected, and soluble proteins were only detected in the cytoplasm with anti-His and anti-FLAG (Figure 7-4, Panel B, 3).

The PhoA component of these fusions was found to be inactive irrespective of CyDisCo (Figure 7-5, panel B, 2).

With fusion to $OmpA_{SP}$, the full-length fusion protein was found in the cytoplasm and periplasm (Figure 7-4, panel 2, His, 3) irrespective of CyDisCo. This was detected by anti-His and anti-FLAG, although detection by anti-FLAG, highlights major truncated species and therefore fusion instability (Figure 7-4, panel B, 3). As full-length protein (Appendix 7) was observed with anti-His (Figure 7-4, Panel B, His, 3), it appears that the fusion was truncated at multiple positions after the FLAG tag. In terms of activity, the protein was active in the periplasmic but not in the cytoplasmic fractions (Figure 7-5, panel B, 3). This indicates that a native conformation was achieved in the periplasm, presumably thanks to the Dsb pathway (section 1.3.1.), but in the cytoplasm, most of the hGH-PhoA conforms to a stable but non-native conformation facilitated by the co-expression of CyDisCo. However, since PhoA-containing truncated products are present in the periplasm, it is not possible at this stage to determine whether the activity originates from full-length protein. Similar to exported hGH-PhoA (section 7.2.2.), it was difficult to confirm whether the periplasmic fraction contained premature and/or mature protein due to a size difference of only 2.1 kDa (Appendix 7). PDI, a member of CyDisCo, was also detected in the periplasmic fraction (Appendix 8) as observed previously with $OmpA_{SP}$-hGH and $OmpA_{SP}$-hGH-PhoA (section 7.2.2.). Cell lysis was ruled out since LacI was not detected in the periplasm. This support the idea that the combination of $OmpA_{SP}$ fusions with CyDisCo correlates with specific export of at least one of the CyDisCo components by a means yet unknown.

In contrast, $TorA_{SP}$ did not allow export of the PhoA-hGH fusion (Appendix 7) irrespective of CyDisCo (Figure 7-4, panel B, His, 4). The cytoplasmic proteins were both detected as full-length and also truncated downstream of the FLAG-tag but were found to be inactive (Figure 7-5, panel B, 4). This indicates that although these fusions have achieved a stable conformation in the cytoplasm, they are not native for activity nor can bypass the proofreading ability of Tat. Furthermore, this suggests that either PhoA, hGH or both are not folded to the same conformation as their independently expressed counterparts (section 6.3. and 6.5.).

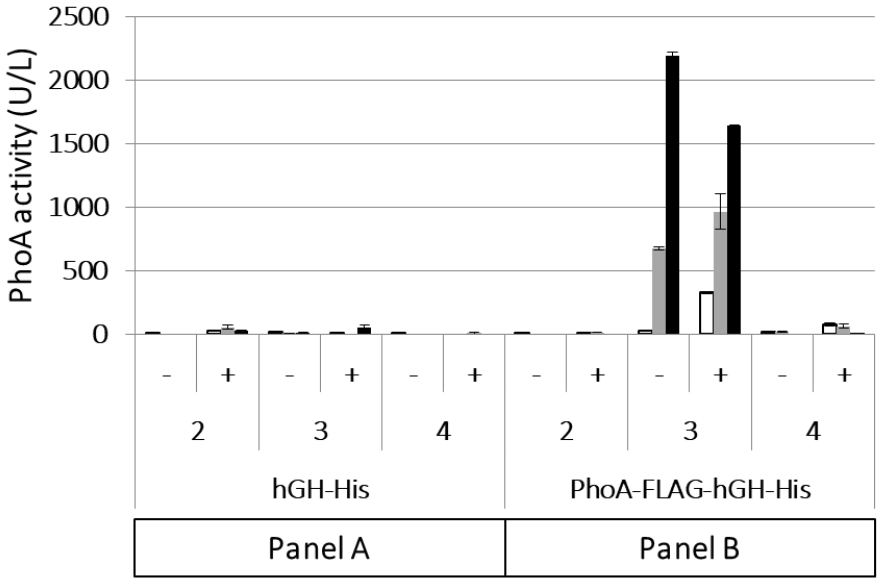


Figure 7-5: PhoA activity of fusions with hGH as the protein of interest

The PhoA activity assays were performed in triplicate and measured against a titration curve using a commercial enzyme. The background signals were eliminated by subtracting the activity value from the empty vector's fractions to each corresponding fraction. PhoA-hGH fusions are represented as no signal peptide (2), fused with $OmpA_{SP}$ (3) or with $TorA_{SP}$ (4), and with (+) and without (-) CyDisCo. The bars indicate the cytoplasmic (white), insoluble (grey) and periplasmic (black) fractions.

7.3. Fusions with the fatty acid binding protein 4

The fatty acid binding protein 4 (FABP4), also known as aP2 and A-FABP, is a human protein localised in the nucleus (Cao et al., 2013). It possesses no DSBs, has a native size of 14.7 kDa and is monomeric. This protein is a chaperone involved in many immunometabolic diseases by reversibly binding to specific lipids and thus is a therapeutic target (Burak et al., 2015, Miao et al., 2015, González and Fisher, 2015). Soluble FABP4 was successfully expressed in-house in the cytoplasm of *E. coli* to the decent yields of ≈ 50-100 mg/L in shake flasks (D. McMillan, unpublished data). This protein is therefore not in need for an alternative way of production but was selected to evaluate Tat export as it readily folds correctly in the cytoplasm. FABP4 alone as well as fusions with hGH (hGH-FLAG-FABP4-His) or PhoA (PhoA-FLAG-FABP4-His) were cloned with no signal peptide, $OmpA_{SP}$ and $TorA_{SP}$ (Table 2-2). The plasmids containing these genes were transformed, along with an empty vector, into the MC4100 WT strain and protein expressed following a 2 h induction at 30°C (Table 2-3, condition

1). The harvested cells were fractionated into the cytoplasmic, insoluble and periplasmic fractions according to the PureFrac method. The localisation and PhoA activity results are represented in Figure 7-6/Appendix 9 and Figure 7-7 respectively.

When FABP4 was expressed alone and without a signal peptide, stable FABP4-His was detected in the cytoplasmic and, unexpectedly, in the periplasmic fractions irrespective of CyDisCo (Figure 7-6, panel A, 2). Similar to sfGFP and hGH (sections 6.4. and 6.5.), this indicates export via an unknown mechanism(s). The reliability of this data can be justified by the PDI-cMyc, Erv1p-cMyc, LacI and MBP controls which confirmed the fractions purity (Appendix 9). Analysis of the N-terminal protein sequence of FABP4 by the prediction software SignalP failed to identify a conserved Sec-like signal peptide motif (Nielsen, 2017a), although it is naturally secreted by adipocytes to control liver glucose metabolism via a non-classical pathway (Cao et al., 2013). FABP4 shares a β-barrel structure with sfGFP (Appendix 10) which could be linked to a common unknown transport mechanism. However, hGH, which also presented evidence of non-classical export (sections 6.5. and 7.2.3.), has a helix bundle structure (Figure 9-1), suggesting that there might be more than one unknown pathway. It would be interesting to investigate whether the tertiary structure of the substrate plays a role in determining export by this unknown pathway, as eluded to by Zhang et al. (2017). Lastly, CyDisCo had no effect on protein localisation of the leaderless FABP4, but this is not unexpected since FABP4-His does not contain any DSBs, although improved expression was evident.

FABP4 is exported without a signal peptide and therefore, subsequent interpretation of the data when this protein is fused to a signal peptide might be challenging. Indeed, with $OmpA_{SP}$ and without CyDisCo, this protein was present in the periplasm but in the premature form (Figure 7-6, panel A, His, 3, orange arrow, Appendix 7). The fractionation controls confirm that this result is not due to cross contamination artefacts (Appendix 9). A prerequisite for Sec export of periplasmic proteins is the cleavage of the signal peptide by LepB prior to release into the periplasm (section 1.3.1.). LepB is an essential endopeptidase and therefore reports of a non-functional or attenuated enzyme has never been published. It can be argued that those polypeptides that exit the Sec translocon pore, but bear uncleaved signal peptides, will presumably remain tethered to the plasma membrane and

consequently be detected in the insoluble fraction. Therefore, this data suggests that the export occurred via an unknown mechanism where the premature protein does not interact with LepB nor is the result of the Sec pathway. It is not known if this is the same pathway used by $OmpA_{SP}$-hGH-PhoA (Figure 7-2, panel A, 3), although there is strong correlation with the latter which was also exported and released into the periplasm with an intact leader sequence. Indeed, the lower band detected in the periplasm does not correspond to the theoretical molecular weight size of the mature protein but is rather considered a proteolytically-cleaved product (Appendix 7).

With CyDisCo however, the absence of export of $OmpA_{SP}$-FABP4 (Figure 7-6, panel A, 3) was unexpected with respect to export profiles of leaderless FABP4 which was exported irrespective of CyDisCo. Although FABP4 does not possess any DSBs, it does have two free cysteine residues (González and Fisher, 2015) as indicated in magenta in Appendix 10. These residues are internally located within the β-barrel structure of FABP4 and presumably are inaccessible to CyDisCo once FABP4 is correctly folded. As FABP4 was exported without a leader, and with $TorA_{SP}$ in the presence of CyDisCo (Figure 7-6, panel A, 4), it is clear that a unique interaction between CyDisCo and the Sec-targeted FABP4 has occurred. It is therefore possible that SecB recruitment by $OmpA_{SP}$ delays folding of the nascent polypeptide and this allows CyDisCo to force a DSB bond. The resulting, presumably unstable and non-natively folded proteins are subsequently unsuitable for either export pathway and are either degraded or aggregated. The lack of any stable protein in the cytoplasmic fraction and presence of less aggregate in the insoluble preparation compared to the profile without CyDisCo, lends support to this possibility. Further confirmation with cysteine mutants of FABP4 may elucidate the validity of this theory.

Lastly, with $TorA_{SP}$, export of FABP4 was comparable irrespective of CyDisCo (Figure 7-6, panel A, His, 4) and suggests that $TorA_{SP}$ nor the nascent polypeptide interacts with CyDisCo. It is possible that these polypeptides fold rapidly similar to FABP4 expressed without a signal peptide, and before CyDisCo can interact. The exported proteins correlate to the theoretical molecular weight size of the mature form (Appendix 7) thereby suggesting that this is specifically Tat export. It is less likely that export is due to the unknown mechanism utilised by the leaderless FABP4 since it is suspected to not

trigger signal peptide cleavage. However, to prove that this is exclusively Tat-specific export, further confirmation is needed using the Tat-null strain.

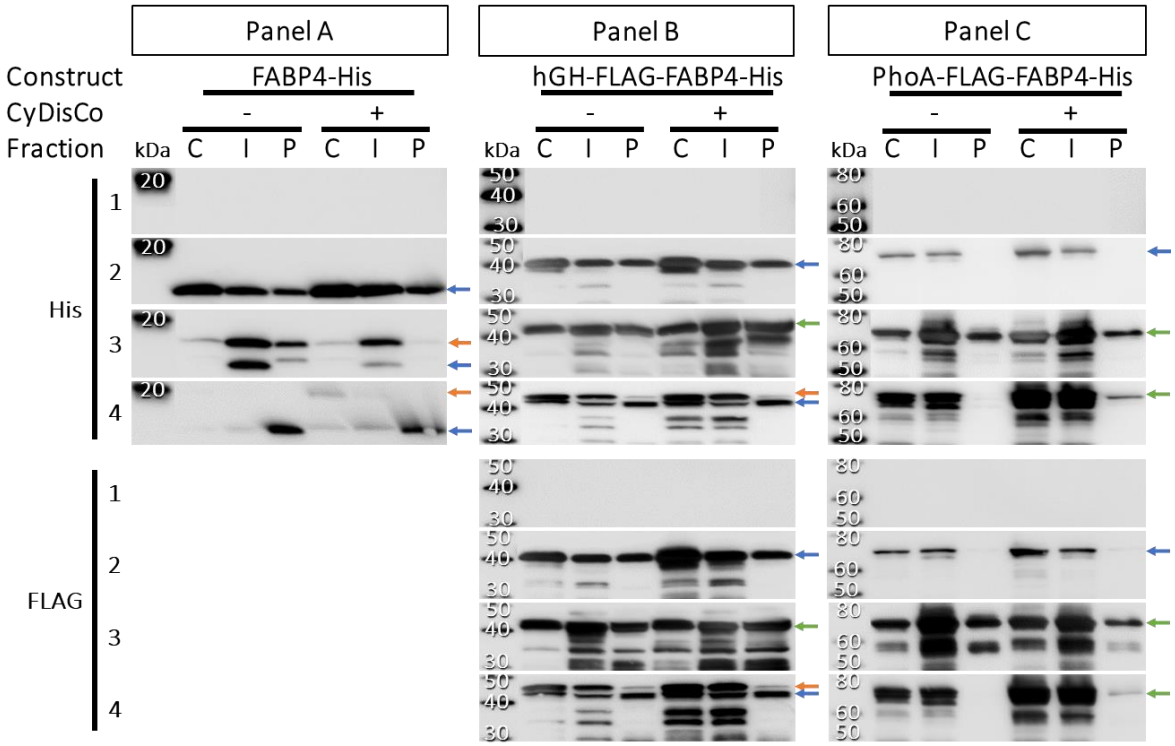


Figure 7-6: Localisation of fusions with FABP4 as protein of interest

His-tagged FABP4-His, the fusion hGH-FABP4-His or the fusion PhoA-FABP4-His with no signal peptide (2) or fused with $OmpA_{SP}$ (3) or $TorA_{SP}$ (4) as well as an empty vector (1) were expressed with (+) and without (-) CyDisCo in the MC4100 WT strain. The cells were harvested after 2h induction at 30°C and fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. The Western-blots represent the detection of His-tagged and FLAG-tagged proteins using anti-His and anti-FLAG antibodies respectively. Arrows indicate the bands of premature (orange) and mature proteins (blue) when corresponding to the theoretical molecular weight size (Appendix 7). Green arrows indicate full-length protein where the presence of the absence of the signal peptide is not discernible. PageRuler Plus Prestained Protein Ladder (Life Technologies) and MagicMark™ (Life Technologies) molecular weight markers are indicated in kilodalton (kDa). The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

With regards to the FABP4-His fused to hGH as the soluble partner, the fusion protein was observed in all periplasmic fractions irrespective of a signal peptide or CyDisCo. hGH was able to carry C-terminally-fused FABP4 across the inner membrane without any signal peptide (Figure 7-6, panel 2, His and FLAG, 2) similar to the hGH-PhoA fusion (Figure 7-2, panel B, 2). It is also clear that whist proteolytically-cleaved species were detected in both the cytoplasmic and the insoluble fractions

(Figure 6, panel B, 2), only full-length protein was exported (Appendix 7). However, due to this undefined export, the interpretation of data when the fusion is expressed with a signal peptide, is challenging.

Indeed, the export mechanism of the translocated proteins observed with $OmpA_{SP}$ cannot be easily deduced (Figure 7-6, panel B, 3). This protein appears to be pre-mature, though it is difficult to assign a mature form due to the presence of proteolytically cleaved products (Figure 7-6, panel B, 3). This further suggests that the fusion with $OmpA_{SP}$ is unstable. However, as the localisation profiles are similar irrespective of CyDisCo, this suggests that fusion of FABP4 to hGH prevents an interaction with CyDisCo as observed previously with $OmpA_{SP}$-FABP4 (Figure 7-6, panel A, 3).

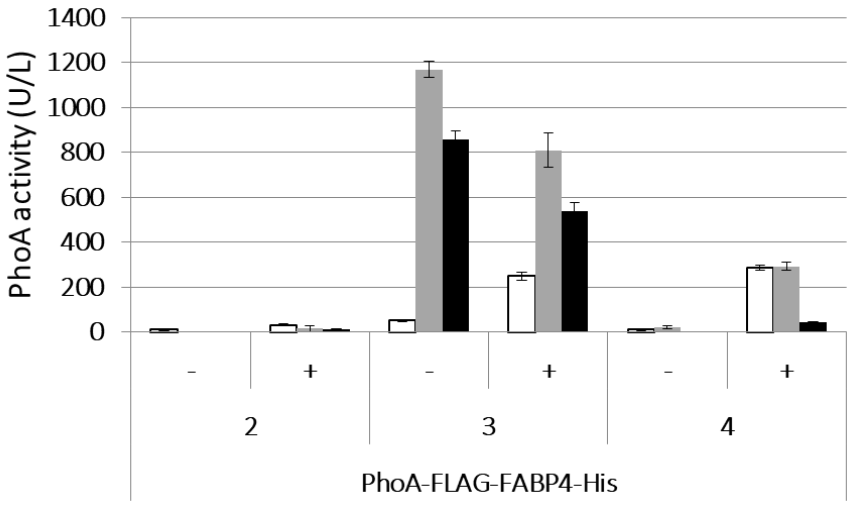


Figure 7-7: PhoA activity of fusions with FABP4 as protein of interest

PhoA activity assays were performed in triplicate and measured against a titration curve using a commercial PhoA enzyme. The background signals were eliminated by subtracting the activity value from the empty vector's fractions to each corresponding fraction. PhoA-FABP4 fusions are represented as no signal peptide (2), fused with $OmpA_{SP}$ (3) or with $TorA_{SP}$ (4), and with (+) and without (-) CyDisCo. The bars indicate the cytoplasmic (white), insoluble (grey) and periplasmic (black) fractions.

With $TorA_{SP}$, the exported proteins are mostly full-length proteins and correspond to the theoretical molecular size of the mature proteins (Figure 7-6, panel B, 4; Appendix 7). However, proteins corresponding to the both premature and mature forms were detected in the cytoplasm. It is unclear if these are proteolytically-cleaved products by a non-specific protease(s) or specifically cleaved at

the signal peptide, a reaction that can be driven by TatC in the absence of TatB (section 1.3.2.4.). It is therefore difficult to understand the role of the Tat pathway in relation to export of this fusion.

Conversely, when FABP4 is fused to PhoA, this product behaved in a specific way towards the export pathways compared to the hGH-FABP4 fusion. With regards to the leaderless and $OmpA_{SP}$-fused proteins, the export and activity profiles correlate with to the corresponding fusion where hGH was used as the protein of interest (Figure 7-4 and Figure 7-5, panel B, 2 and 3). Export of FABP4 with the natural Sec substrate PhoA at the N-terminus was not detected without a signal peptide, as expected (Figure 7-6, panel C, 2). Conversely, with $OmpA_{SP}$, the fusion was comparably exported, irrespective of CyDisCo, and demonstrably active in terms of PhoA activity (Figure 7-6, panel C, 3 and Figure 7-7, 3). It is not clear whether this protein is the mature or premature form due to their similarity in molecular weight size. Interestingly without CyDisCo, activity of soluble PhoA was only detected in the periplasmic fractions, presumably from translocated proteins via the Sec pathway and their subsequent native folding in the periplasm. In contrast with CyDisCo, activity was also detected in the cytoplasmic fraction, suggesting that a portion of the preprotein is folded in the cytoplasm with correct DSB formation within the PhoA portion, presumably via CyDisCo. It is also notable that the fusion is significantly truncated (Figure 7-6, FLAG, 4), and therefore the activity could be attributed to a proportion of proteolytically cleaved products that bear an intact PhoA as well as the full-length fusion.

With the Tat leader $TorA_{SP}$, the fusion protein showed export albeit minor, only with CyDisCo (Figure 7-6, panel C, 4). It is not known whether this form is the premature or mature due to their similar predicted molecular weight sizes (Appendix 7). The periplasmic proteins presented insignificant PhoA activity (Figure 7-7, 4) but is not unexpected based on the poor level of export. PhoA, the N-terminal soluble carrier of the fusion, was confirmed to be active mostly in the cytoplasmic and insoluble fractions and only with CyDisCo (Figure 7-7, 3). This suggests that when the correct DSBs of the carrier PhoA are formed, the fusion protein is accepted by the proofreading of Tat and translocated. With CyDisCo, however, the cytoplasmic proteins showed similar PhoA activity with the corresponding $OmpA_{SP}$ cytoplasmic fusion, supporting the idea that, in this case, CyDisCo

enabled correct DSB formation. A test comparing export in WT and Tat-null strains would conclude on Tat contribution to this export.

In conclusion, FABP4 was shown to export without a leader sequence similar to hGH (section 7.2.1.) and sfGFP (section 6.4.). FABP4 and sfGFP share a β-barrel structure but it is not known whether these proteins are exported by the same unknown mechanism. sfGFP was demonstrably active as indicated by its fluorescence (section 6.4.) so it would be interesting to determine if exported FABP4 is also correctly folded. However, this is not possible in this present work due to the complicated nature of these binding assays and the necessity for purified protein ((González and Fisher, 2015)). Furthermore, due to this 'auto-export' nature, it is difficult to attribute a pathway(s) responsible for export when fused to hGH. There is strong correlation that exported proteins are premature in that they still retain their signal peptides upon exit into the periplasm, particularly with $OmpA_{SP}$. Whether this is a phenotype of the unknown export mechanism is unclear.

Conversely, with PhoA at the N-terminus, export of the fusion is signal peptide dependent. PhoA being a Sec substrate, is not exported without a Sec signal peptide. However, although export was demonstrated with a signal peptide, the fusion was also deemed unstable or poorly exported with the respective OmpA and TorA signal peptides. However, given that FABP4 auto-exports as an intact protein into the periplasm and providing that it is correctly folded, the auto-export also raises the question of redundancy with respect to the fusion strategy.

7.4. Fusions with single variable domain of camelid heavy chain antibodies

As well as the conventional immunoglobulin (IgG1 subclass) antibody shared with humans, camelids and sharks possess another unique type of antibody which is characterised by only a heavy chain and no light chain (Bever et al., 2016) (Figure 7-8A). These heavy-chain only antibodies (HcAb) are homodimers of the heavy chain and consist of a variable heavy domain linked directly to the Fc region. The fragment crystallizable (Fc) region itself is classified as either a gamma 2 or gamma 3

subclass and therefore these HcAbs are known as IgG2 or 3 (IgG2/3), as illustrated in Figure 7-8A. Hence, they do not contain the CH1 constant region present in a human classical IgG antibody. The complementary domain region 3 (CDR3) of the variable heavy is also typically longer than the corresponding region of a conventional IgG antibody (Bever et al., 2016). The variable domain itself can be expressed independently and, in this case, is known as a VHH or Nanobody® (Steeland et al., 2016). Since their discovery in 1991, interest in VHHs has increased due to several features: their obvious small size enabling access to hidden viral epitopes, their stability, efficient production in *E. coli* or yeast, tissue penetration competency, rapid elimination by the kidneys and the possibility to create bivalent antibodies or add an effector domain by fusion (Harmsen and De Haard, 2007, Wesolowski et al., 2009, de Marco, 2011, Muyldermans, 2013). VHHs possess one canonical DSB linking framework (FW) 1 to FW3 which stabilizes the structure and is represented as S-S_1 in Figure 7-8B. But some llama germlines possess additional cysteines in the CDRs that are involved in non-canonical DSB formation. Indeed, a DSB between CDR1 and CDR3 (S-S_2) or between two cysteine residues in CDR3 (S-S_3, Figure 7-8B) have been identified (Nguyen et al., 2011, Iezzi et al., 2018). These non-canonical DSBs are known to confer a higher thermal stability (Akazawa-Ogawa et al., 2016). VHHs are often expressed in the periplasm of *E. coli* through Sec since the canonical DSB is necessary for the stability and activity of the antibody fragment (Yau et al., 2003, Akazawa-Ogawa et al., 2016, de Marco, 2015). Moreover, another type of antibody fragment, scFv (Figure 7-8A), has been reported to be exported by Tat when co-expressed with CyDisCo (Alanen et al., 2015). Here, a variable heavy and a variable light domain derived from an IgG antibody are connected via a linker. Since a VHH is a smaller but a similar class of antibody to a scFv, it raises the question whether this molecule could also be exported via the Tat pathway. Furthermore, Tat export of VHH antibody fragments has never been reported. Therefore, testing Tat export of VHHs is attractive due to the interest of these molecules in the biopharmaceuticals industry (Maffey et al., 2016, Wang et al., 2018, Iezzi et al., 2018).

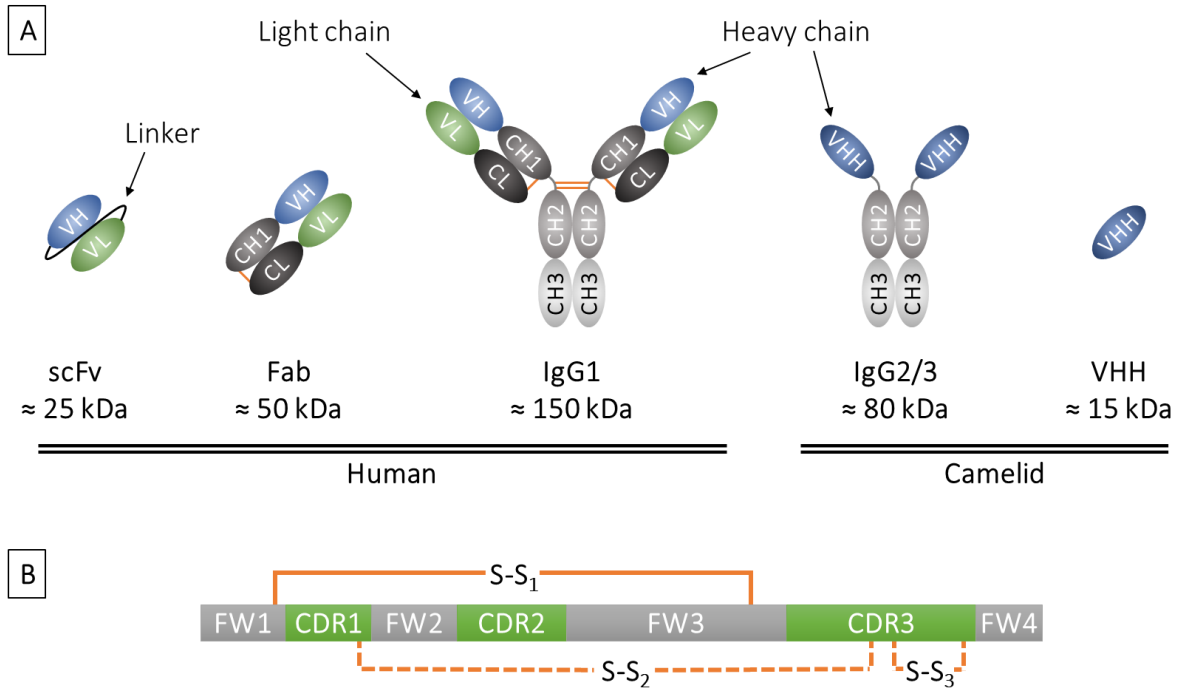


Figure 7-8: Origin and details of VHH antibody fragment

A. A typical representation of an immunoglobulin G (IgG) antibody (IgG1 subclass) and a camelid heavy chain only antibody (HcAb) of the subclass IgG2 or IgG3 (IgG2/3) is presented. The IgG1 is composed of two light and two heavy chains. The light chain consists of a variable light (VL) region and a constant light (CL) region. The heavy chain is composed of a variable heavy (VH), constant heavy (CH1) domains and a fragment crystallizable (Fc) region. The Fc consists of a CH2 and CH3 domain. The camelid IgG2/3 is composed of 2 heavy chains consisting of a VHH domain and an Fc region of subclass 2 or 3. The following antibody fragments can be engineered from their respective parental IgGs: Fab antibody fragment (Fab), scFv and VHH. The variable domain of the camelid antibody is called VHH single domain also known as Nanobody®. Disulphide bonds are represented in orange. The approximate molecular weights are indicated underneath the corresponding molecules in kilodalton (kDa). B. Schematic of a "conventional" VHH from llamas is indicated with frameworks (FW), complementary domain regions (CDR) and disulphide bonds (S-S). The canonical and non-canonical DSBs are represented by the respective plain and dashed lines. The Figure was adapted from Muyldermans (2001) and Iezzi et al. (2018).

Two VHHs were selected for assessment based on the specific properties of the molecule. These antibody fragments were originally selected by phage display panning where the VHHs are displayed on the surface of the phage as N-terminal fusions to the coat protein pIII (Ledsgaard et al., 2018). The Sec pathway of *E. coli* is utilised to enable DSB formation of the VHH by fusing the post-translational Sec signal peptide of *Erwinia carotovora* pectate lyase B ($PelB_{SP}$) (Solforosi et al., 2012). Thus, this ensures that only those VHHs that are correctly folded with DSB(s) and displayed in an active conformation albeit in a fusion context, can be selected for binding to an antigen via panning.

Table 7-1: Biochemical characteristics of the two VHHs

VHH	Length	Size (kDa)	DSB	Charge at pH 7.0	Library of origin
VHH1	116	12.7	1	1.1	Naïve
VHH2	119	12.9	2	-0.1	Immune

Two VHHs were selected from UCB Celltech's in-house screening platforms (A. Scott-Tucker, personal communication), based on specific features integral to each molecule as summarized in Table 7-1. Although both VHHs have very similar molecular weight size, VHH1 possesses one DSB whilst VHH2 possesses two DSBs. The additional DSB of VHH2 is represented as S-S_3 in Figure 7-8B and is known to increase the rigidity and thus the stability of the protein (Akazawa-Ogawa et al., 2016). The alignment of primary sequences of these VHHs reveals a 75 % similarity between the two antibody fragments (Appendix 11). Regarding the net charge, decreasing the surface charge to a negative value was demonstrated to improve protein solubility (Kramer et al., 2012, Chan et al., 2013). Therefore, VHH2 was predicted to be more stable than VHH1. Other discrepancies beyond this feature were investigated in terms of hydropathy and secondary structures to predict any potential differences with regards to export. Figure 7-9A revealed a globally similar hydropathy with the highest discrepancies located around positions 25 and 55. Furthermore, the secondary structure predictions indicate potential variances notably in the number and length of β-sheets which may result in different spatial conformations (Figure 7-9B). Overall, the two VHHs would be ideal test cases to investigate Tat export and the effect of CyDisCo.

To evaluate Tat export of these proteins and whether translocation can be improved by fusion, expression vectors were designed to produce the His-tagged VHHs alone as well as C-terminal fusions to PhoA and hGH as depicted in Figure 7-1. Constructs were cloned with either no signal peptide or fused to $OmpA_{SP}$ or $TorA_{SP}$, and with or without CyDisCo (Table 2-2). These plasmids were transformed into MC4100 WT, cultured at 30°C and expression was induced for 2 h at 30°C (Table 2-3, condition 1). The harvested cells were treated according to the PureFrac method to isolate the cytoplasmic insoluble and periplasmic subcellular compartments. Subsequently, the proteins were

separated by reducing SDS-PAGE and detected by Western-blot. The localisation results of both VHHs as well as the activity assays of PhoA are described in sections 7.4.1. and 7.4.2.

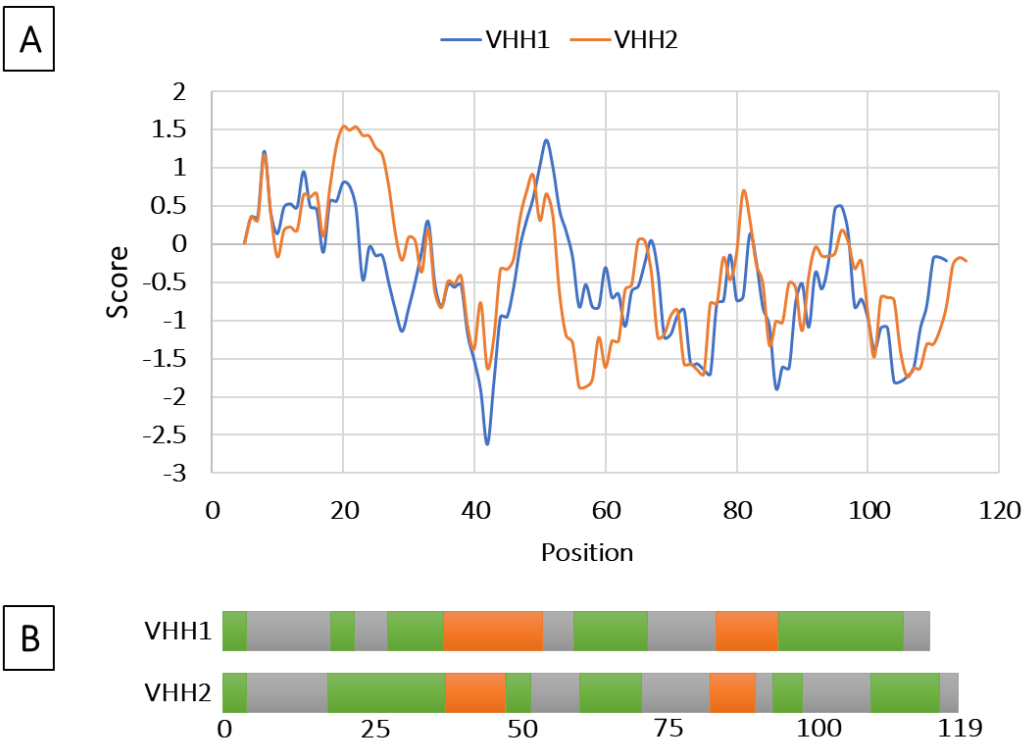


Figure 7-9: Hydropathy scale and predicted secondary structures for both VHH amino acid sequences

A. The graph represented the hydropathy scale of the VHH1 (blue) and VHH2 (orange) sequences using ProtScale online analysis tool with the Kyte and Doolittle package (Kyte and Doolittle, 1982). B. Alignment of the secondary structure of the VHH1 and VHH2 predicted using the online tool CFSSP (Ashok Kuma, 2013). The α-helixes are represented in orange, the β-sheets in green and the coil in grey.

7.4.1. Expression of fusions with VHH1

When VHH1-His was expressed without a signal peptide, no detectable protein was observed as represented in Figure 7-10 (panel A, His, 2). This demonstrated that the protein was either not expressed, expressed then degraded or expressed but the His tag was cleaved. Fusion to $OmpA_{SP}$, enabled comparable Sec export with respect to CyDisCo, where most of the preprotein was presumably translocated and processed to the mature form (Appendix 7). However, a protein corresponding to the predicted molecular weight of the mature protein was also observed in the cytoplasm, irrespective of CyDisCo. Cross-contamination was ruled out, as the periplasmic marker MBP was not detected in any other fraction, although an insignificant amount of LacI was found in

the periplasmic fraction of the sample without CyDisCo (Appendix 12). However, this does not account for level of VHH1 shown in the cytoplasmic fraction and is presumably proteolytic cleavage. As VHH1 was not detected when expressed without a leader, the nature of this protein is unclear. It is possible that through the recruitment of Sec specific chaperones, the protein was able to fold and stabilise (Figure 7-10, panel A, His, 3). It is apparent that CyDisCo has no effect on this localisation.

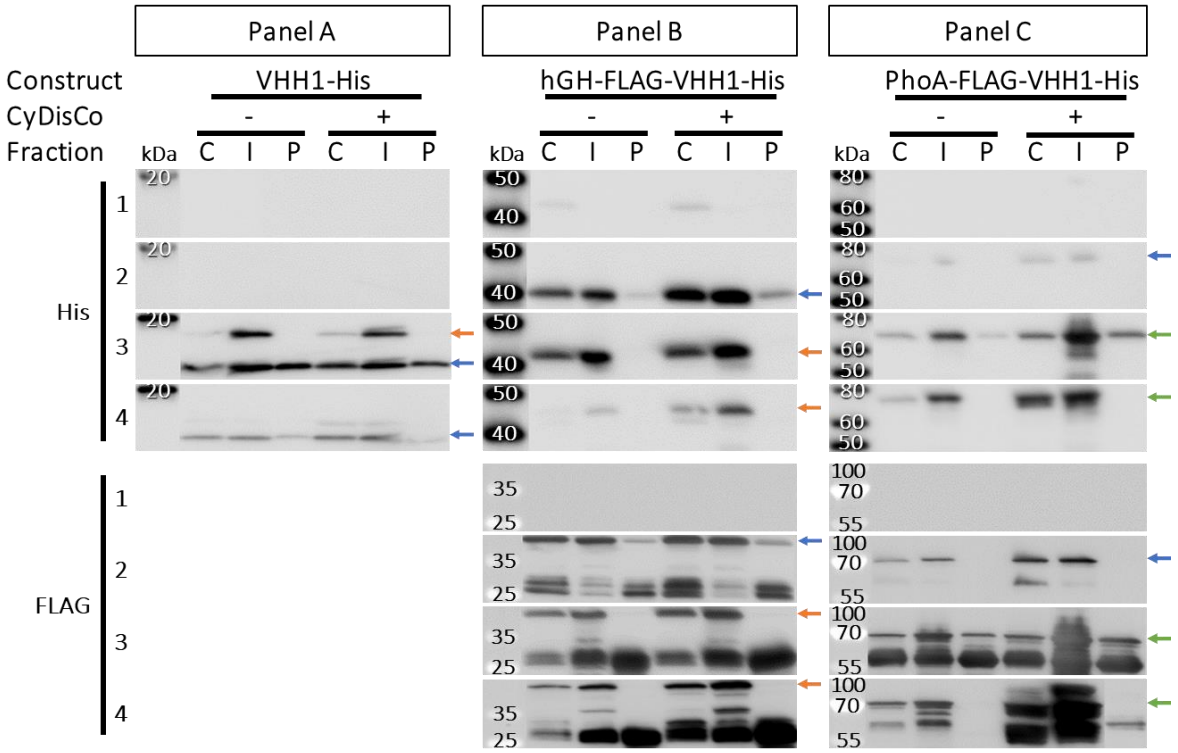


Figure 7-10: Localisation of fusions with VHH1 as protein of interest

His-tagged VHH1, the fusion hGH-VHH-His or the fusion PhoA-VHH1-His with no signal peptide (2) or fused with $OmpA_{SP}$ (3) or $TorA_{SP}$ (4) as well as an empty vector (1) were expressed with (+) and without (-) CyDisCo in the MC4100 WT strain. The cells were harvested after 2h induction at 30°C and fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. The Western-blots represent the detection of His-tagged and FLAG-tagged proteins using anti-His and anti-FLAG antibodies respectively. Arrows indicate the bands of premature (orange) and mature proteins (blue) when corresponding to the theoretical molecular weight size (Appendix 7). Green arrows indicate full-length protein where the presence of the absence of the signal peptide is not discernible. PageRuler Plus Prestained Protein Ladder (Life Technologies) and MagicMark™ (Life Technologies) molecular weight markers are indicated in kilodalton (kDa). The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

With $TorA_{SP}$, poor expression and export of VHH1 was observed irrespective of CyDisCo (Figure 7-10, panel A, 4). This suggests that the protein cannot fold into a native conformation in the cytoplasm

despite CyDisCo. (Figure 7-10, panel A, His, 4). Indeed, VHH1 instability is apparent as demonstrated by the overall reduced level of expressed protein but also by the presence of a similarly proteolytically cleaved species in the cytoplasm when compared to $OmpA_{SP}$. No cross-contamination was detected to indicate artefactual data, although interestingly, both CyDisCo proteins Erv1p-cMyc and PDI-cMyc, were detected in the periplasm with $OmpA_{SP}$ and $TorA_{SP}$ fusions (Appendix 12) These unexpected sub-cellular localisations will be further discussed in section 9.4.

With regards to the hGH fusion (Appendix 7) with no signal peptide, VHH1 was stable in the cytoplasm and presented minimal export with some improvement when co-expressed with CyDisCo (Figure 7-10, panel B, 2). This indicates that hGH, by way of its 'auto export' mechanism, can not only lead PhoA and FABP4 but also, minimally, VHH1 into the periplasm. It appears as though CyDisCo enhances expression of VHH1 including minimal export improvement. However, the exported protein status, folded or unfolded, is still unknown. The localisation data was supported by the absence of cross-contamination (Appendix 12). However, neither $OmpA_{SP}$ nor $TorA_{SP}$ allowed export of the full-length fusion protein irrespective of CyDisCo, although the fusion with $TorA_{SP}$ appeared to be more unstable compared to $OmpA_{SP}$ (Figure 7-10, panel B, 3 and 4). Proteolytically cleaved species were detected in all relevant fractions where the fusion was expressed with anti-FLAG but not with the anti-His antibody. This indicates that the cleavage point(s) is within the C-terminal region downstream from the FLAG tag (Figure 7-1) and that the resulting products will likely bear an intact hGH and signal peptide. The cytoplasmic profiles correlate with the corresponding periplasmic profiles, and further support that the truncated fusion proteins that still bear a signal peptide, are exported thereafter. This highlights a weakness in the fusion design which is further discussed in section 9.7.1.

On the other hand, when using PhoA as the soluble partner (Appendix 7), the fusion was highly unstable when expressed without a signal peptide and degraded or proteolytically cleaved as observed with anti-FLAG (Figure 7-10, panel C, 2). However, the protein was stabilised with $OmpA_{SP}$ and poor export was observed with slight improvement with CyDisCo (Figure 7-10, panel C, 3). Similar to the hGH fusion, anti-FLAG detection revealed export of truncated species irrespective of CyDisCo

which suggest that this is Sec export as well as highlighting fusion instability (Figure 7-10, panel C, FLAG 2). Significant PhoA activity (> 3kU/L) was detected in the periplasmic fractions (Figure 7-11, 3). However, due to the poor export of the full-length fusion and the presence of proteolytically cleaved species which likely contains an intact PhoA (Figure 7-10, panel C, 3 and Figure 7-11, 3), it is not clear if activity comes from the exported full-length fusion.

With regards to $TorA_{Sp}$, no export of the full-length fusion was noted but CyDisCo significantly increased the expression of the full-length and clipped products in the cytoplasm (Figure 7-10, panel C, 4). Poor export of a proteolytically cleaved species was detected only with anti-FLAG and with CyDisCo, suggesting that this likely contains an intact PhoA. PhoA was previously shown to export via Tat (section 7.2.2.). Lack of PhoA activity in any fraction indicates that targeting the fusion to the Tat pathway does not confer a native folding (Figure 7-11, 4).

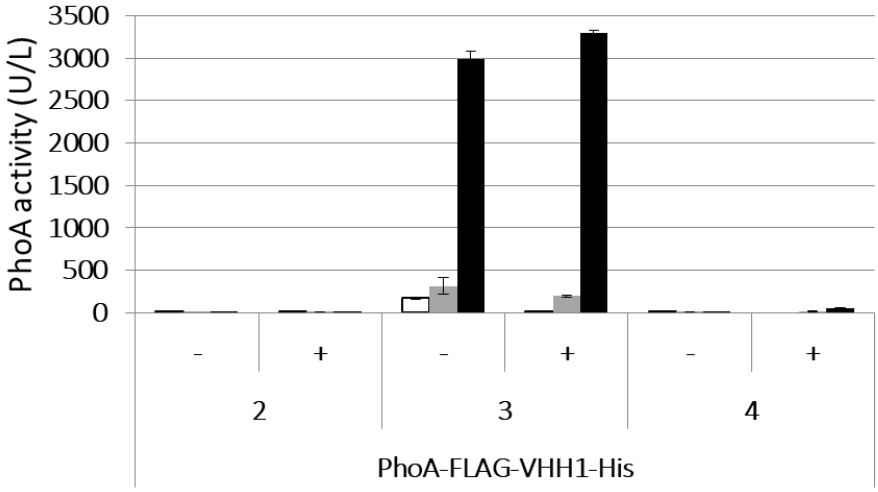


Figure 7-11: PhoA activity of fusions with VHH1 as protein of interest

The PhoA activity assays were performed in triplicate and measured against a titration curve using a commercial enzyme. The background signals were eliminated by subtracting the activity value from the empty vector's fractions to each corresponding fraction. PhoA-VHH1 fusions are represented as no signal peptide (2), fused with $OmpA_{SP}$ (3) or with $TorA_{SP}$ (4), and with (+) and without (-) CyDisCo. The bars indicate the cytoplasmic (white), insoluble (grey) and periplasmic (black) fractions.

To sum up, expression of VHH1 without a signal peptide was unstable in the cytoplasm of *E. coli* but fusing a signal peptide or a carrier protein to the N-terminus of the VHH improved stability. Export of the full-length protein was only observed with the signal peptide of OmpA, but this was poor. Due

to the export of proteolytically cleaved species which likely contain an intact PhoA, it is unclear if the exported full-length fusion is active. The presence of extensive truncated species indicates that the fusion design may need further optimisation.

7.4.2. Expression of fusions with VHH2

The export profile of the second antibody fragment, VHH2 (section 7.4.), revealed similarities but also discrepancies compared to VHH1 (Table 7-1 and Figure 7-9). Indeed, the lack of detected proteins for VHH2-His with no signal peptide irrespective of CyDisCo indicates that this VHH has expression and/or degradation issue(s) in *E. coli* just like VHH1 (Figure 7-12, panel A, His, 2). With the post-translational Sec signal peptide of OmpA and without CyDisCo, VHH2 was mostly located in the insoluble fraction in a mature form, although proteolytic cleavage is also detected (Figure 7-12, panel A, His, 3; Appendix 7). This suggests that the preprotein was exported by Sec and processed by LepB, but the mature form folded incorrectly in the periplasm resulting in the formation of periplasmic inclusion bodies. This indicates that this VHH is unstable whilst VHH1 was stable in similar settings. With CyDisCo however, the localisation and cleavage status of VHH2 appeared different from the no CyDisCo profile, implying that these chaperones improved VHH2 expression (Figure 7-12, panel A, His, 3). Very poor export was observed, though this was mostly premature whereas most of the mature form was found in the insoluble fraction. As the CyDisCo chaperones enable folding which is incompatible with Sec export, the insoluble mature proteins are likely to originate from Sec exported protein which aggregate in the periplasm due to incorrect folding. However, the mature form was also discovered in the periplasm along with the cytoplasmic markers, PDI, Erv1p and LacI, suggesting cytoplasmic contamination (Appendix 13). But, inner membrane permeability was ruled out since the periplasmic marker, MBP was not detected in the cytoplasm. This is, therefore, unclear if VHH2 exhibited Sec targeted export or employed a similar export mechanism to PDI, Erv1p and LacI. Periplasmic presence of PDI and Erv1p when expressing a protein fused to $OmpA_{SP}$ was observed in other cases and is further discussed in section 9.4 and 9.7.1.

The data with $TorA_{SP}$ and without CyDisCo further confirm the instability of VHH2 but also indicates that the stability conferred by $OmpA_{SP}$ is signal peptide dependent. CyDisCo marginally improved the expression of the protein as premature and a protein corresponding to the predicted molecular weight size of the mature form were observed in both the cytoplasm and insoluble fractions (Figure 7-12, panel A, 4). Cross contamination was not detected (Appendix 13) indicating that proteolysis is responsible for the cytoplasmic cleaved product. Mature protein was present in the periplasmic fraction though this indicates very poor export. Thus, VHH2 does not appear to fold correctly in any of the different cellular environments presented in this study.

However, fusion to hGH as a soluble partner appears to improve expression of VHH2 as demonstrated by improved detection of full-length proteins by both anti-His and anti-FLAG antibodies (Figure 7-12, panel B). Full-length hGH-VHH2 was detected in the periplasm irrespective of CyDisCo and without the need of a signal peptide (Figure 7-12, panel B, 2), in agreement with previous data that shows hGH's "auto export" ability (sections 7.2.2., 7.3 and 7.4.1.). This pathway is not understood, and it is still unknown whether the exported proteins are folded or unfolded. This translocation of VHH2 does not require DSB formation prior to export although, with CyDisCo, improved protein expression was evident. When VHH2 alone was targeted to the Sec or Tat pathways, it showed no expression despite CyDisCo. This therefore suggests that expression of VHH2 was improved, possibly through its enhanced solubility, and further improved by CyDisCo.

When the fusion was expressed with either $OmpA_{SP}$ and $TorA_{SP}$, and without CyDisCo, a mature protein was present in the periplasm, suggesting Sec and Tat export respectively (Figure 7-12, panel B, 3 and 4). But, with CyDisCo, premature protein was also observed in the periplasm regardless of the signal peptide. PDI and Erv1p were also detected in the periplasm, correlating with previous data where these proteins are exported by an unspecified mechanism with an $OmpA_{SP}$-substrate (Appendix 13). Cross contamination was not a factor as demonstrated by the localisation profiles of LacI and MBP. With $TorA_{SP}$, CyDisCo appeared to significantly improve export though proteolytic cleavage was also increased (Figure 7-12, panel B, 4).

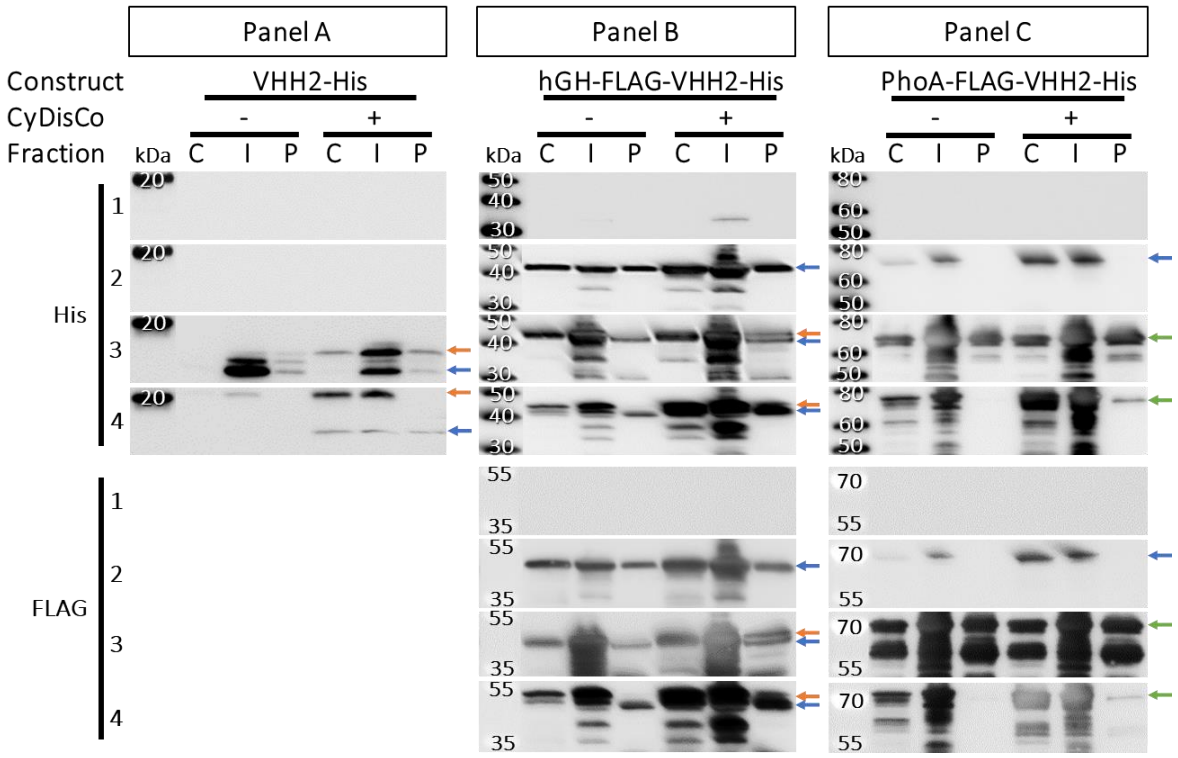


Figure 7-12: Localisation of the fusions with VHH2 as protein of interest

VHH2-His (His-tagged VHH2), the fusion hGH-VHH2-His or the fusion PhoA-VHH2-His with no signal peptide (2) or fused with $OmpA_{SP}$ (3) or $TorA_{SP}$ (4) as well as an empty vector (1) were expressed with (+) and without (-) CyDisCo in the MC4100 WT strain. The cells were harvested after 2h induction at 30°C and fractionated using the PureFrac method into the cytoplasmic (C), insoluble (I) and periplasmic (P) fractions. The Western-blots represent the detection of His-tagged and FLAG-tagged proteins using anti-His and anti-FLAG antibodies respectively. Arrows indicate the bands of premature (orange) and mature proteins (blue) when corresponding to the theoretical molecular weight size (Appendix 7). Green arrows indicate full-length protein where the presence of the absence of the signal peptide is not discernible. PageRuler Plus Prestained Protein Ladder (Life Technologies) and MagicMark™ (Life Technologies) molecular weight markers are indicated in kilodalton (kDa).

Using PhoA as a soluble partner, the expression of VHH2 was also improved as demonstrated by the presence of soluble protein in the cytoplasm and further enhanced by CyDisCo (Figure 7-12, panel C). With $OmpA_{SP}$, export was observed irrespective of CyDisCo indicating Sec export (Figure 7-12, panel C, 3) and the PhoA domain of the fusion was found to be active (Figure 7-13, 3). Multiple truncated products were also present, and these profiles correlated between fractions (Figure 7-12, panel C, 3). This suggests that either cleaved products with an intact PhoA and signal peptide were translocated, or the protein was cleaved following export, similar to the PhoA-VHH1 fusion (section 7.4.1.). Cross-contamination with the cytoplasmic marker LacI was evident and thereby needs to be considered during interpretation of this data (Appendix 13).

However, with $TorA_{SP}$, minimal export of the full-length protein was noted, only with CyDisCo (Figure 7-12, panel C, 4). This result is expected as PhoA was shown to require CyDisCo for Tat translocation (section 6.3.1.). Most of the soluble protein was detected in the cytoplasm suggesting that most of the fusions did not adopt a conformational status accepted by the proofreading ability of Tat, despite CyDisCo. Indeed, this protein and also the exported fusion was not active (Figure 7-13, 4), in agreement with previous data with respect to PhoA fusions (section 6.3., 7.2., 7.3. and 7.4.1.). However, with the latter, the activity could have been below detection level as a very low amount of protein was exported. No cross contamination was detected confirming the validity of this data (Appendix 13).

To conclude, VHH2 is naturally unstable but stability was slightly increased when fused to $OmpA_{SP}$ or $TorA_{SP}$ and considerably improved when fused to hGH or PhoA. CyDisCo also positively impacted stability resulting in improved protein yields. Though with hGH, the fusion adopts a more stable conformation where most of the protein is full-length. Without a signal peptide, this fusion is exported, thereby providing further support for an unspecified mechanism for hGH export. With the signal peptides, export of both premature and mature proteins complicates interpretation. However, with the mounting support that resident proteins exported by the unknown mechanism still retain the signal peptide, this suggests that the fusion is exported by both targeted and unknown pathways. The use of the Tat-null strain would at least lead to confirmation of Tat export. When VHH was fused to PhoA, the Sec and Tat specificity was confirmed with minimal export of the fusion. Moreover, Sec export, which handles unfolded polypeptide (section 1.3.1.), was unaffected by the presence of CyDisCo. VHH2 was overall shown to be more stable as a fusion, particularly when fused to hGH. However, it is not known if these proteins are folding correctly or can actively bind their ligands.

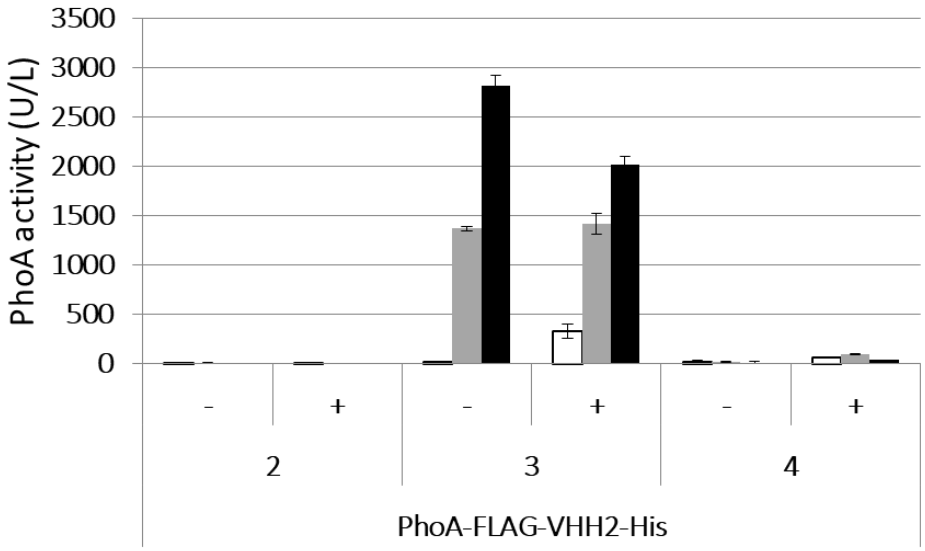


Figure 7-13: PhoA activity of fusions with VHH2 as protein of interest

The samples origin was described in the Figure 7-12 legend. The activity assays were performed in triplicate and measured against a titration curve using a commercial enzyme. The background signals were eliminated by subtracting the activity value from the empty vector's fractions to each corresponding fraction. The bars indicate the cytoplasmic (white), insoluble (grey) and periplasmic (black) fractions.

7.5. Proteolytic clipping of the fusion proteins

During experiments as outlined in this chapter, truncated species of the fusion proteins were a common observation detected by both anti-His and anti-FLAG antibodies. When hGH, FABP4 and VHH1 were expressed a fused soluble partner, they were observed as full-length, corresponding to their respective premature, and/or mature forms and thus, demonstrating their stability. The exceptions are PhoA and VHH2 which both showed proteolytic cleavage. Indeed, presence of cleaved species was observed for both proteins by anti-His, indicating that N-terminal degradation had occurred (Figure 7-2 and Figure 7-12, panel A, 3 and 4).

With regards to the fusions, truncated species were observed by anti-His for hGH-PhoA (Figure 7-2, panel B, His), hGH FABP4 (Figure 7-6, panel B, His), PhoA-FABP4 (Figure 7-6, panel C, His), hGH-VHH2 (Figure 7-12, panel B, His) and PhoA-VHH2 (Figure 7-12, panel C, His). This suggests that the fusion design is not ideal for all recombinant proteins. These proteolytically-clipped species were present in each fraction demonstrating that instability of the recombinant protein is not fraction dependent.

Moreover, no truncated species were detected in the periplasm by anti-His when PhoA was used as a soluble partner irrespective of CyDisCo, indicating that this protein is more stable than hGH as a carrier. For instance, only full-length proteins were observed in the periplasmic fractions when expressing PhoA-FABP4 whereas proteolytically-clipped products were present when hGH-FABP4 was expressed (Figure 7-6, panel B and C).

Furthermore, N-terminus degradation in the cytoplasm means loss of the signal peptide which would prevent export. Hence, if classical translocation occurred, periplasmic truncation species were originally exported as full-length and sustained degradation. For example, if the truncated species observed by anti-His in the periplasmic fraction when expressing $OmpA_{SP}$-hGH-PhoA (Figure 7-2, panel B, His, 3) have been exported by Sec, they must have been translocated prior to N-terminal degradation. However, the non-classical export mechanism(s), highlighted throughout this chapter, could have transported the truncated species from the cytoplasm across the inner membrane. Moreover, with anti-FLAG, the degradation patterns do not correlate with the patterns of truncated species revealed by anti-His. This suggests that a proportion of the truncation was C-terminally cleaved from the FLAG tag (Figure 7-1) especially for PhoA-hGH (Figure 7-4, panel B), hGH-VHH1 and PhoA-VHH1 (Figure 7-10, panels B and C). By subtracting the size of the proteins of interest to the full-length proteins (Appendix 7), the sizes of the truncated species revealed that degradation occurred within both the protein of interest and the linker. Since these proteins expressed alone were not cleaved, the fusion rendered them prone to degradation.

Finally, discounting the molecular weight size of hGH (22.3 kDa) or PhoA (47.3 kDa) from the full-length protein (Appendix 7) indicates that, in general terms, degradation is likely to have occurred in the soluble partner, with respect to anti-His detection, whereas, with anti-FLAG detection, the susceptible region is the C-terminal protein and the interconnecting linker region.

Overall, fusing these proteins together – except PhoA – created protease-sensitive sites present in both the soluble partners, proteins of interest and linker. Therefore, the design of the fusion needs to be optimised to reduce the truncation events and is further discussed in Chapter 9.

7.6. Conclusion on fusion proteins

To investigate whether expressing a recombinant protein as a fusion would improve its export by Tat, hGH and PhoA were selected as the soluble partner (Chapter 6). To test this hypothesis, five recombinant proteins were selected: the soluble partner PhoA and hGH themselves, FABP4 and two VHHs. Table 7-2 summarizes the export resulting from the different conditions tested. Data confirmed that hGH is exported without the help of a signal peptide by an unknown mechanism. Also, all proteins of interest when -fused to the C-terminus of hGH showed some unexpected degree of export without a signal peptide (Table 7-2). This is an interesting observation and questions the necessity for a classical export pathway. Although, questions regarding the understanding of this pathway and whether the exported proteins are folded into their native conformations remain unanswered. Hence, it makes it difficult to interpret data and attribute a role for the secretion pathways when using this soluble partner. However, PhoA as a soluble partner happens to be more specific with regards to export since no translocation was observed without a signal peptide. As expected with $OmpA_{SP}$, all fusions exported albeit mostly poorly, irrespective of CyDisCo. VHH1 is an exception, as it appeared to be particularly unstable without CyDisCo. With $TorA_{SP}$, only FABP4 and VHH2 showed export, though poor, suggesting that the other proteins were unfolded and were thus, rejected by Tat proofreading mechanism (Table 7-2). Overall, the fusion strategy did not improve Tat export and this work confirmed that targeting to Sec secretion pathway, the most prolifically used pathway for protein export in *E. coli* was generally more successful than Tat in terms of export. Nevertheless, fusing a soluble partner improved the stability of the proteins of interest especially for the VHH.

Protein stability was often improved by fusion to signal peptides and also by co-expressing CyDisCo. The chaperone and the DSB formation activities of PDI and Erv1p seemed to enhance the folding of the recombinant proteins. However, most of the proteins were not accepted by the Tat proofreading mechanism and the PhoA component was not active in the cytoplasm which suggests that CyDisCo conferred a stable but non-native conformation. With $OmpA_{SP}$, the CyDisCo proteins had the

tendency to co-localise in the periplasm with the periplasmically-targeted protein. Hence, an unknown connection between this signal peptide and the CyDisCo chaperones drove to their export via an uncharacterised mechanism (discussed in section 9.7.2.). It is not known if this particular signal peptide is responsible for export but highlights the need for further evaluation of leader sequences.

Table 7-2: Summary of protein export

The table represents the export of the proteins presented in this chapter. They were either unfused or fused to soluble partner, hGH or PhoA. The data is classified according to its signal peptide or lack of thereof: no signal peptide (No SP), the signal peptide of OmpA ($OmpA_{SP}$) or of TorA ($TorA_{SP}$) and co-expressed without (Cy-) or with (Cy+) CyDisCo. The relative export was represented as no export (-), non-significant export (+/-), poor export (+) and good export (++). Grey cells represent the conditions that were not tested.

Construct		No SP		$OmpA_{SP}$		$TorA_{SP}$	
Soluble partner	Carried protein	Cy-	Cy+	Cy-	Cy+	Cy-	Cy+
-	PhoA	-	-	++	++	-	+
	hGH	-	+/-	+	+	+	+
	FABP4	+	+	+	-	+	+
	VHH1	-	-	+	+	-	-
	VHH2	-	-	+/-	+/-	-	+/-
hGH	PhoA	+	++	++	++	+	+
	hGH						
	FABP4	+	+	+	++	+	+
	VHH1	-	+	-	-	-	-
	VHH2	+	+	+	+	+	+
PhoA	PhoA						
	hGH	-	-	++	++	-	-
	FABP4	-	-	+	+	-	+
	VHH1	-	-	-	+	-	-
	VHH2	-	-	+	+	-	+

The fusion proteins showed evidence of both N- and C-terminal degradation in the cytoplasmic and periplasmic fractions. These truncated products are thought to have arisen from proteolytic degradation within the soluble partner, the interconnecting linker and the proteins of interest regions (Figure 7-1). The exception for this is the hGH-VHH2 fusion which remarkably was present as mostly full-length in all relevant fractions (Figure 7-12, panel B, 2 and 4; panel C, 4). Nevertheless,

this does highlight that the design of the fusion protein is not optimum for most recombinant proteins as will be discussed in section 9.7.1.

On a more positive note, FABP4 was surprisingly transported across the inner membrane without a signal peptide similar to hGH and sfGFP (section 6.4. and 6.5.). The mechanism of transport is not understood nor is it known whether this is the same mechanism for all and therefore would be worth investigating.

8. GENETIC CHARACTERISATION OF THE TAT OPERON

8.1. Introduction to the genomic locus upstream of the Tat operon

The Tat operon contains the genes coding for the TatA, TatB, TatC and TatD proteins (77.8 min) while the additional *tatE* gene (14.2 min min) locates elsewhere on the genome (section 1.3.2.1.). The presence of a mRNA containing *tatA*, *tatB*, *tatC* and *tatD* was confirmed by Weiner et al. (1998) in the *E. coli* K12 HB101 strain. No biological evidence of the actual promoter was demonstrated but by using the prediction software BPROM (Solovyev, 2011), the estimated promoter region was established and is represented in Figure 8-1B. This bacterial promoter contains the consensus -10 and -35 regions upstream of the transcription start site (de Avila e Silva and Echeverrigaray, 2012). These regions are recognized by the housekeeping, sigma factor subunit of the RNA polymerase which after binding, initiate transcription (Borukhov and Nudler, 2003).

The *ubiEJB* operon is upstream of this operon and codes for proteins involved in the ubiquinone biosynthesis pathway which serve the respiratory electron transport chain (Aussel et al., 2014a, Wu et al., 2015) (Figure 8-1A). Indeed, a conserved -35 and -10 region was predicted by BPROM upstream of *ubiE* (Figure 8-1D). No other putative promoters were predicted within the operon, which supports that the *ubiE*, *ubiJ* and *ubiB* genes are part of the same operon (Solovyev, 2011). Having two operons next to one another is not unusual in bacteria but the interest of the following work came from a recently developed strain called TatExpress engineered from *E. coli* K12 W3110 (Browning et al., 2017). They inserted a cassette containing the *tac* promoter (de Boer et al., 1983) upstream of the predicted promoter site of the *tat* operon, and downstream of the *ubiB* stop codon as shown in Figure 8-1C. Accordingly, the *tac* promoter subsequently starts 196 bp downstream from the stop codon of the *ubiB* gene. Thus, natural levels of expression can be still accomplished during cell growth from the native promoter whilst overexpression can be achieved upon induction by IPTG. The non-coding region of 196 bp is a non-complementary scar region from the strain construction which contains restriction enzyme and Flp recombination sites (Browning et al., 2017).

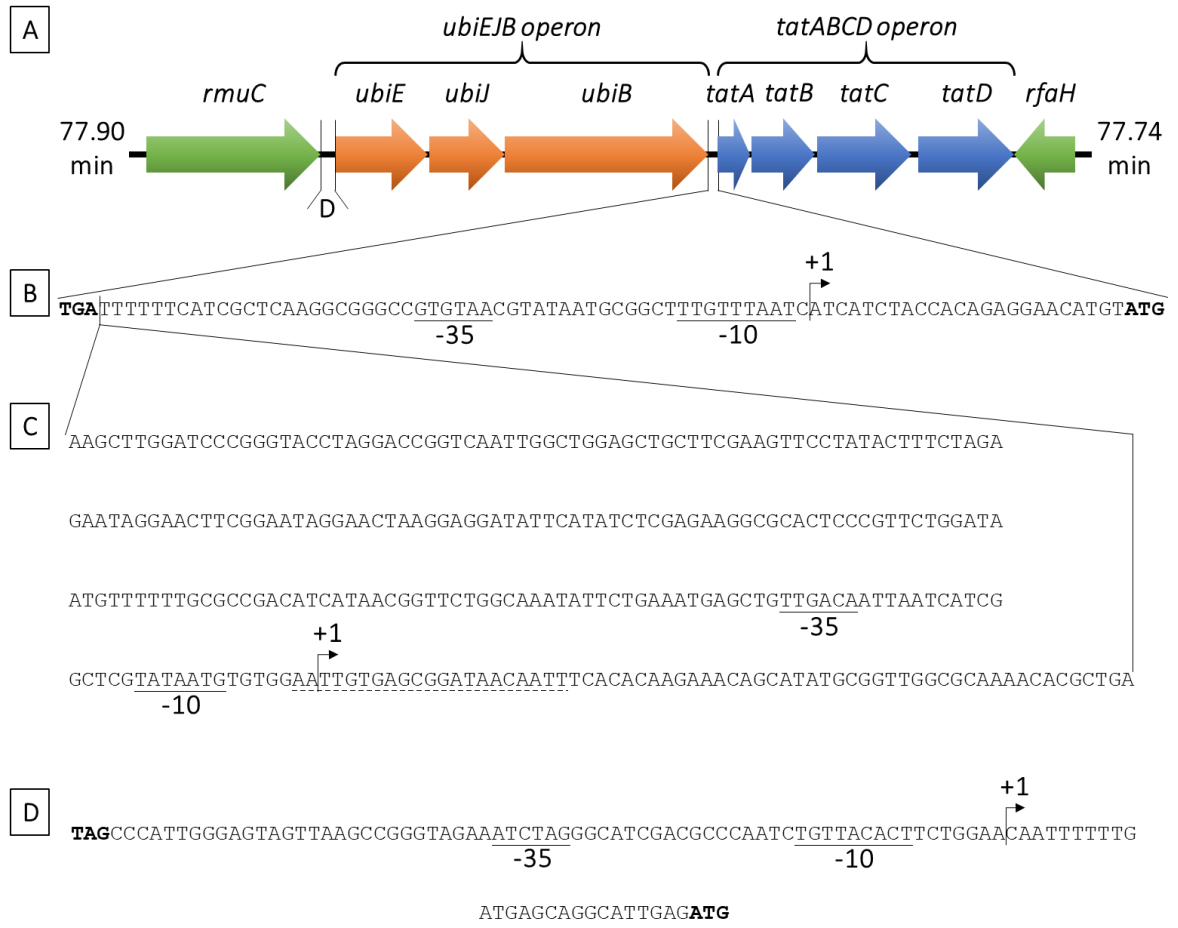


Figure 8-1: Genetic organisation of the Tat operon

A. Genetic representation of the *ubiEJB* and *tatABCD* operon loci flanked by the chromosomal minutes. The coloured arrows represent the genes on a relative scale where the *tatABCD* operon (blue), the *ubiEJB* operon (orange) and the *rmuC* and *rfaH* flanking genes (green) are indicated. **B.** The non-coding nucleotide sequence between the *ubiB* and *tatA* genes of WT W3110. The stop codon of *ubiB* (TGA) and the start codon of *tatA* (ATG) are represented in bold on the left and right side respectively. The promoter was predicted with strong conservation to the -35 and -10 motif specific for sigma transcription factors and is indicated as underlined. The transcription start site (+1) of the downstream *tatA* gene is indicated by an arrow (Solovyev, 2011). **C.** The non-coding nucleotide sequence inserted by Browning et al. (2017) strictly after the stop codon of the upstream *ubiB* gene to build the TatExpress strain. This sequence contains the *tac* promoter represented with the -35 and -10 motif (plain underlined), the *lac* operator (dashed underline) and the transcription start site (+1) as indicated by an arrow (de Boer et al., 1983). **D.** The non-coding nucleotide region between the *rmuC* and *ubiE* genes. The stop codon of *rmuC* (TGA) and the start codon of *ubiE* (ATG) are represented in bold on the left and right side respectively. The promoter was predicted with strong conservation to the -35 and -10 motif of sigma transcription factors, indicated as underlined. The transcription start site (+1) of downstream *ubiE* gene is indicated by an arrow (Solovyev, 2011). No Rho-independent terminator was predicted immediately downstream of the *rmuC* gene (Solovyev, 2011).

Transcription termination is operated by two mechanisms in bacteria: intrinsic or Rho-dependent terminations (Porrua et al., 2016). The former system corresponds to a GC-rich region, creating a hairpin loop in the nascent mRNA transcript downstream from the stop codon of the last gene being

transcribed. This can be detected by computational models (Gardner et al., 2011, Solovyev, 2011). Conversely, Rho is a bacterial factor which binds to a non-conserved 70-100 nucleotide binding site of the nascent transcript (Ciampi, 2006, Porrua et al., 2016, Mitra et al., 2017). Beyond being rich in cytosines and accounting for near half of the transcription terminations in *E. coli*, these sites do not have a consensus motif and so remain unpredictable (Ciampi, 2006).

If a terminator, whether Rho-dependent or-independent, was present downstream of the *ubiEJB* operon and halts with a distinct transcriptional terminator, it would be located in the region after its stop codon. Therefore, the transcription initiated from the *tac* promoter inserted strictly after *ubiB* stop codon would presumably, be terminated shortly after by the terminator region of the *ubiB* gene. However, Browning et al. (2017) demonstrated that the TatExpress strain enabled overexpression of TatA upon IPTG induction which suggests the absence of a transcriptional terminator downstream of the stop codon of *ubiB*. Indeed, using available online tools FindTerm and ARNold, no rho-independent hairpin terminator was identified downstream of the *ubiB* gene in WT strains (Solovyev, 2011, Naville et al., 2011). This suggests that if there is a terminator, it is likely to be Rho-dependent and cannot be predicted. All of this suggests that as one possible scenario, *ubiEJB-tatABCD* genes could be transcribed as one mRNA transcript with a 78 bp non-coding region between the two clusters. In this case, the internal promoter upstream of the *tat* operon would give a second transcript containing *tatABCD* mRNA. To validate this hypothesis, PCR on cDNA and total RNA sequencing was performed on the strain used by Browning et al. (2017).

8.2. Analysis of the non-coding region between the *ubiB* and the *tatA* genes

The hypothesis that both *ubiEJB* and *tatABCD* operons are transcribed as a single mRNA transcript was evaluated. Total RNA was extracted from W3110 WT cells harvested at mid-log phase (section 2.2.14.) and subsequently cDNA was prepared by reverse-transcription (section 2.2.15.). The latter was used as DNA template for PCR amplifications with various sets of primers shown in Figure 8-2.

They were designed to ensure specificity by respecting the following criteria: more than 20 nucleotides in length, guanine/cytosine content percentage near 50 %, fusion temperature above 55°C and a thymine or adenine nucleotide at the 3' terminus. The specificity of the primer sequences was established using the Vector NTI software (Life Technologies) and the BLAST online tool (Altschul et al., 1997). Then, the reaction products were analysed by agarose gel electrophoresis (section 2.2.6.) and the results are represented in Figure 8-3 and Figure 8-4. The theoretical amplicons expected in these figures are depicted in Figure 8-2.

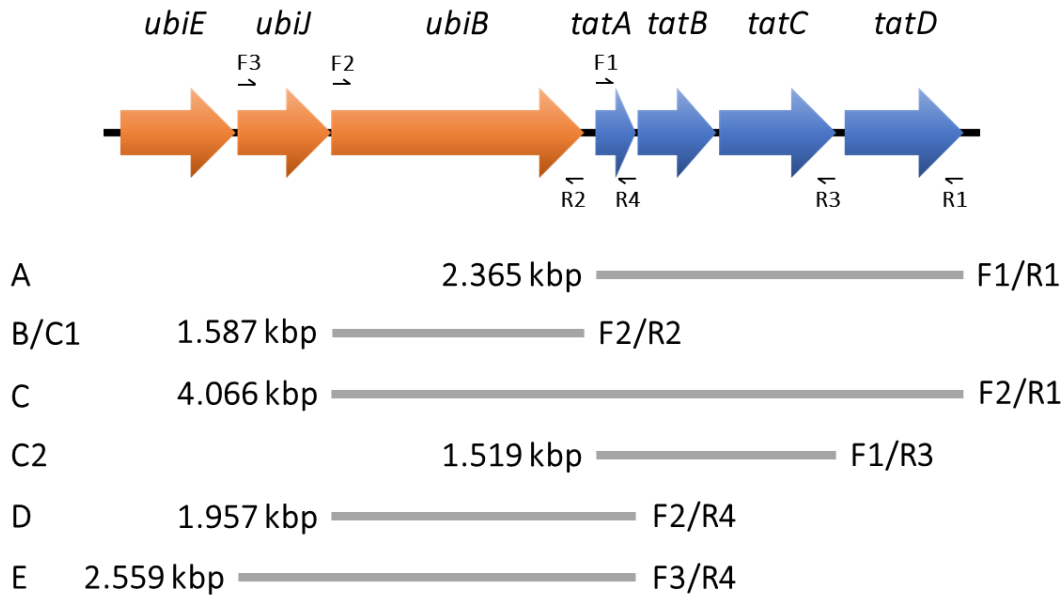


Figure 8-2: Theoretical PCR products

The genetic representation of the *ubiEJB* and *tatABCD* operons locus is represented on top. The coloured arrows represent the genes on a relative scale where the *tatABCD* operon and the *ubiEJB* operon are indicated in blue and orange respectively. The forward (F) and reverse (R) primers used in the PCR reactions are indicated by black arrows. The theoretical PCR amplicons designated as A to F are indicated below. They are framed by their respective size in kilobase pair (kbp) on the left and primers' name on the right.

In Figure 8-3A, control experiments were conducted to establish the size of the PCR products generated using the respective primer pairs F1/R1 and F2/R2 that correspond to the *tatABCD* operon (2.365 kbp; lane A) and the *ubiB* gene (1.587 kbp; lane B). In lane C, to test the presence of mRNA corresponding to the *ubiB-tatABCD* genes (4.066 kbp), F2/R1 primers were used. A ≈ 4.1 kbp band

was observed that corresponds to the predicted size of the PCR amplicon of this region suggesting that there is indeed no terminator between *ubiB* and *tatA*. To prove further this discovery and confirm the presence of *ubiB* and internal members of the tat operon, *tatABC*, the DNA from the band in lane C (Figure 8-3A) was extracted from the gel (section 2.2.7.) and the resulting purified DNA was used as the DNA template to perform Nested PCR. The respective primer pairs F2/R2 (Figure 8-3B, C1) and F1/R3 (Figure 8-3B, C2) resulted in the generation of a ≈ 1.6 kbp product for both. This is the expected size needed to confirm the presence of the *tatABC* genes as well as the *ubiB* gene. This data supports the supposition that these genes are transcribed as a single mRNA transcript. However, the presence of a 1.9 kbp additional band in the C2 lane suggested non-specific annealing (Figure 8-3B) despite specificity checks on the primers.

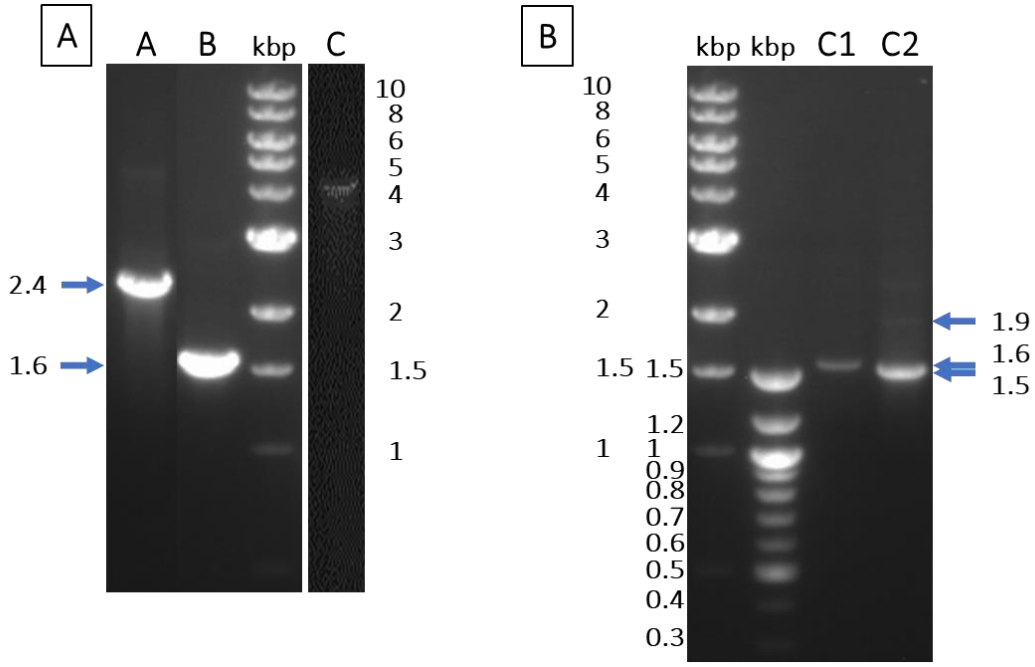


Figure 8-3: Experimental evidence of the absence of terminator between *ubiB* and *tatA*

Agarose gels at 1 % loaded with PCR products amplified from cDNA obtained from total RNA extracted from W3110 WT cells grew in LB to mid-log phase. **A.** The pairs of primers F1/R1, F2/R2 and F2/R1 were used in the PCR which results were loaded on the lanes A, B and C corresponding to the genes *tatABCD*, *ubiB* and *ubiB-tatABCD* respectively. Lane C was part of the same gel as lanes A and B but has been purposely over-contrasted for visual inspection to compensate the lower amplification yield than lanes A and B. **B.** The lanes C1 and C2 contained the PCR products amplified from the amplicon extracted from the lane C, using the pairs of primers F2/R2 and F1/R3 corresponding to the genes *ubiB* and *tatABC* respectively. The approximate sizes of the bands are indicated in kilobase pair (kbp) by blue arrows while the theoretical amplicon sizes are detailed in Figure 8-2. The molecular weight markers are indicated in kbp. The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

This outcome was reproduced in a more thorough experiment performed in triplicates. A control was also used for each condition to confirm that the PCR products originated from cDNA amplification and not from contaminating genomic DNA. This negative control was performed by replacing the reverse-transcriptase enzyme with water during the reverse transcription step with the extracted RNA. Using the pair primers F1/F3 (Figure 8-2), a 1.6 kbp amplicon (Figure 8-4, amplicon C2, a1-a3) corresponding to the *tatABC* genes was generated, while no bands were observed on the negative controls (Figure 8-4, amplicon C2, c1-c3). This confirmed that the *tatABC* genes amplified previously (Figure 8-3B, C2) originated from mRNA rather than contaminating genomic DNA. Further amplifications of a ≈ 1.9 kbp product with primers F2/R4 (Figure 8-4, amplicon D, a1-a3), and a ≈ 2.5 kbp product with F3/R4 (Figure 8-4, amplicon E, a1-a3), corresponding to the size of the respective *ubiB-tatA* and *ubiJB-tatA* regions (Figure 8-2) provided additional confirmation. As before, all controls (Figure 8-4, amplicons D and E, c1-c3) were clear demonstrating the absence of genomic DNA contamination. Unfortunately however, amplification of the *ubiEJB-tatA* (3.36 kbp), *ubiJB-tatABCD* (5.462 kbp) regions or even *ubiEJB* (3.012 kbp)) operon was not successful despite several optimisations. The latter included the application of several sets of primers, each pair annealing to different sites on the genes, a range of annealing temperatures for the annealing step during the PCR, and titration of DMSO concentration to relax the guanine/cytosine rich regions. Nevertheless, the data appears to support the absence of a transcriptional terminator between the *ubi* and *tat* operons.

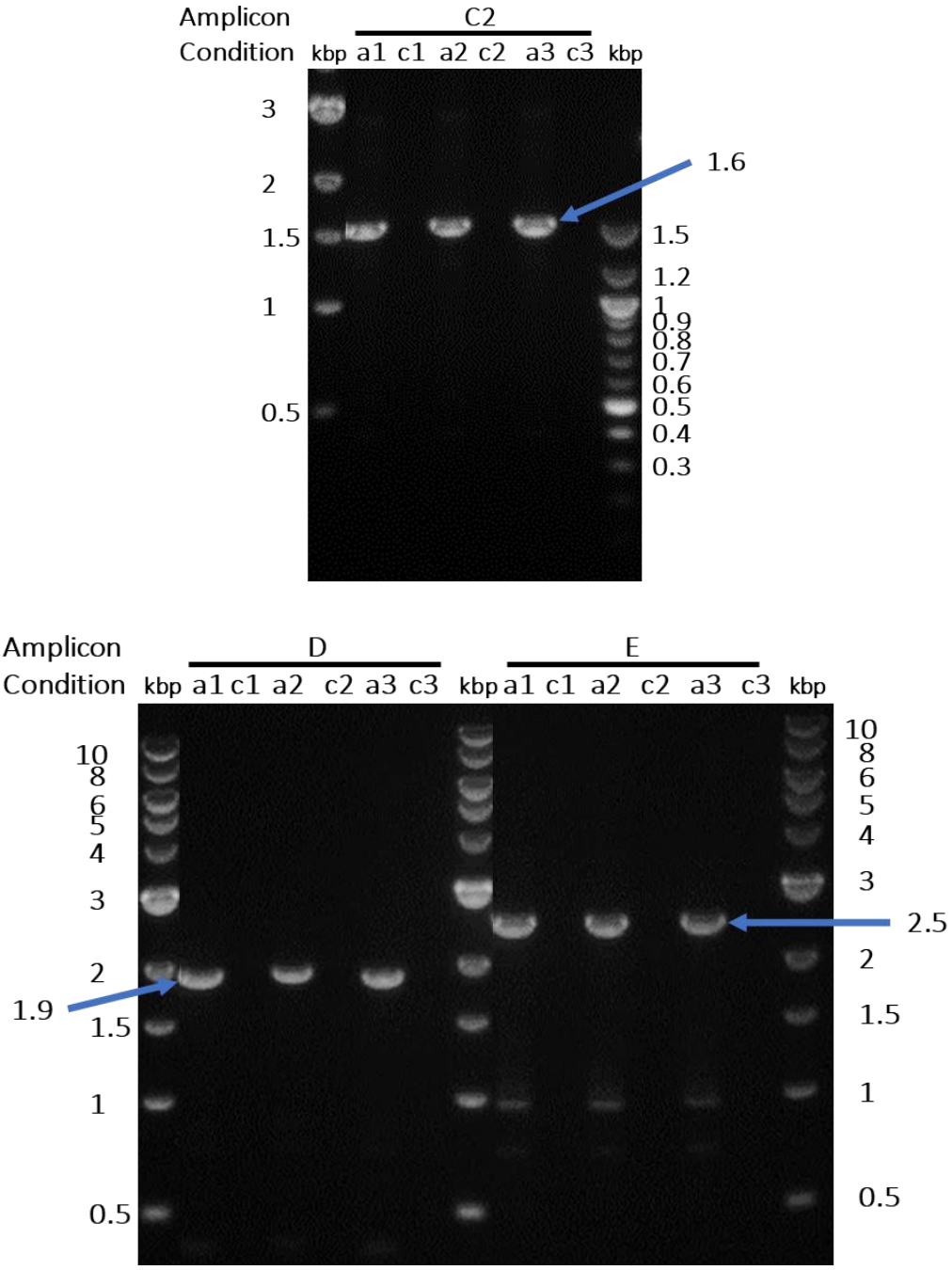


Figure 8-4: Further confirmation of the absence of terminator between *ubiB* and *tatA*

Agarose gels at 1 % loaded with PCR products amplified from cDNA obtained from total RNA extracted from W3110 WT cells grew in LB to mid-log phase. The biological triplicates (a1 to a3) were performed with negative controls (c1 to c3) where the reverse transcriptase enzyme was replaced by water. The results of the PCR using the respective pair of primers F1/R3, F2/R4 and F3/R4 to amplify the *tatABC* (Figure 8-3, C2), *ubiB-tatA* (Figure 8-3, D) and *ubiJB-tatA* (Figure 8-3, E) genes were loaded on the lanes D, E and F. The approximate sizes of the bands are indicated in kilobase pair (kbp) by blue arrows while the theoretical amplicon sizes are detailed in Figure 8-2. The molecular weight markers are indicated in kbp. The figure is a composite image where the marker lanes reflects the approximate position of the molecular weights.

To add support to *ubiEJB* and *tatABCD* forming one large operon, 5' RACE experiments were undertaken. Indeed, this technique allows sequencing of unknown 5' ends of mRNA requiring only knowledge of only a short sequence of the gene (Miller, 2016). 5' RACE allowed identification and characterisation of both eukaryotic and prokaryotic 5' sequences of RNA transcripts (Allen et al., 2007, Langley et al., 2017). However, this technique was never optimised for transcript longer than 1 kbp (Zismann and Nourbakhsh, 2014, Miller, 2016). The aim in this study is to identify an mRNA containing the *ubiEJB-tatABCD* genes (5.5 kbp) which is significantly longer and therefore a high risk of failure. 5' RACE was nonetheless selected to confirm the presence of *ubiB* and *tatA* on the same mRNA by an unbiased approach. Unfortunately, the amplification steps never reached reads long enough to encompass the whole *ubiB* and *tatA* genes on the same mRNA transcript, despite optimisation of the PCR parameters and primers.

8.3. Phylogenetic study of the *ubiEJB-tatABCD* locus

The data gathered in the section 8.2. strongly supports that *ubiEJB* and *tatABCD* are transcribed as a single mRNA presumably from the predicted promoter upstream of *ubiE*. This is in addition to the *tatABCD* genes being transcribed as a single transcript from the predicted promoter upstream of *tatA*. The cluster of these genes, belonging to ubiquinone biosynthesis and Tat export, may be conserved in other species. The *tat* region of several bacterial genomes of different bacterial lineage were downloaded from https://www.ncbi.nlm.nih.gov/genome/ website and analysed using alignment software tools (Geneious). The results represented in Figure 8-5 revealed that, out of the 10 bacteria, 7 have a chromosomal arrangement similar to *E. coli* K12 where the *ubi* operon is located strictly upstream from the *tat* operon: *E. coli* K12 W3110, *Salmonella enterica*, *Yersinia pestis*, *Vibrio cholerae*, *Aeromonas hydrophila*, *Shigella dysenteriae* and *Shewanella oneidensis*. These species belong to the same phylum called Gammaproteobacteria subclade III (Figure 8-6). In *Shigella dysenteriae*, *Yersinia pestis*, *Vibrio cholerae*, *Pseudomonas aeruginosa*, *Xanthomonas axonopodis* and *Neisseiria meningitidis A*, the *tatD* gene is missing from the *tat* operon but encoded elsewhere

in their genome. This gene is conserved but non-essential and codes for a DNase which was demonstrated to have no direct correlation to the Tat translocation machinery (section 1.3.2.3.).

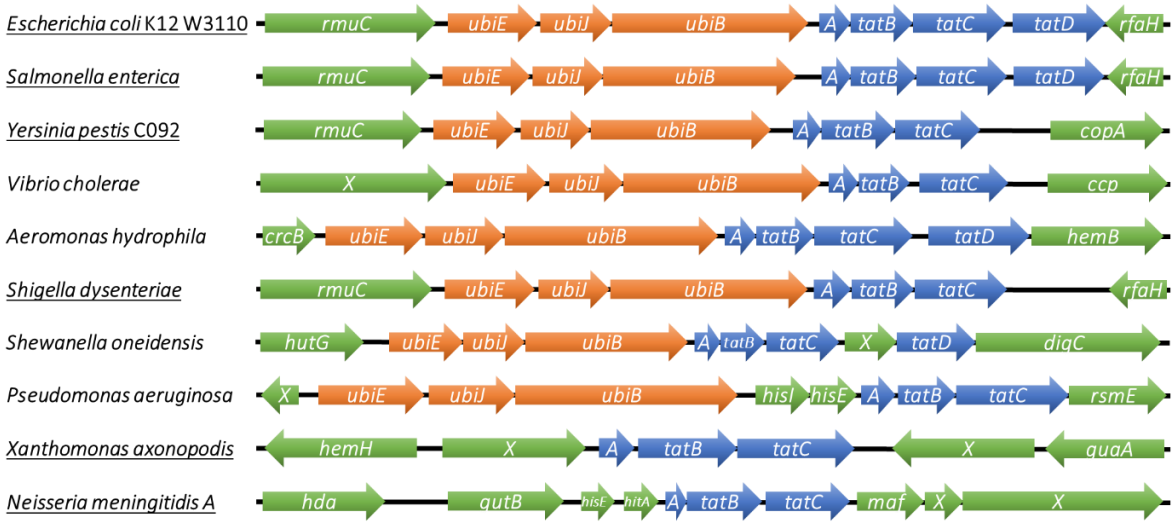


Figure 8-5: Comparison of tat regions across several bacterial species

The *tat* regions of the *Escherichia coli* K12 W3110 (NC_007779), *Salmonella enterica* (NC_003198), *Yersinia pestis* (NC_003143), *Vibrio cholerae* (NC_002505), *Aeromonas hydrophila* (NC_008570), *Shigella dysenteriae* (NC_007606), *Shewanella oneidensis* (NC_004347), *Pseudomonas aeruginosa* (NC_002516), *Xanthomonas axonopodis* (NC_020800) and *Neisseria meningitidis* (NC_003112) genomes were aligned using the Geneious software. The *ubi*, *tat* and other genes are indicated in orange, blue and green respectively. The underlined species indicate that their respective locus is on the complementary strand. The name of the genes is contained in the arrows with *A* standing for *tatA* and *X* for uncharacterised genes.

In Gammaproteobacteria subclade II, *Pseudomonas aeruginosa* possesses the *hisI* and *hisE* genes inserted between the *ubi* and *tat* operons (Figure 8-5 and Figure 8-6). These genes code for the metabolic proteins, phosphoribosyl-AMP cyclohydrolase and phosphoribosyl-ATP pyrophosphatase respectively. *Xanthomonas axonopodis* and *Neisseria meningitidis* were found to be distinct, as members of the Gammaproteobacteria subclade I and Betaproteobacteria respectively. Both do not have the *ubi* operon located upstream from the *tat* operon suggesting that the *ubi-tat* locus is conserved within the Gammaproteobacteria subclade III.

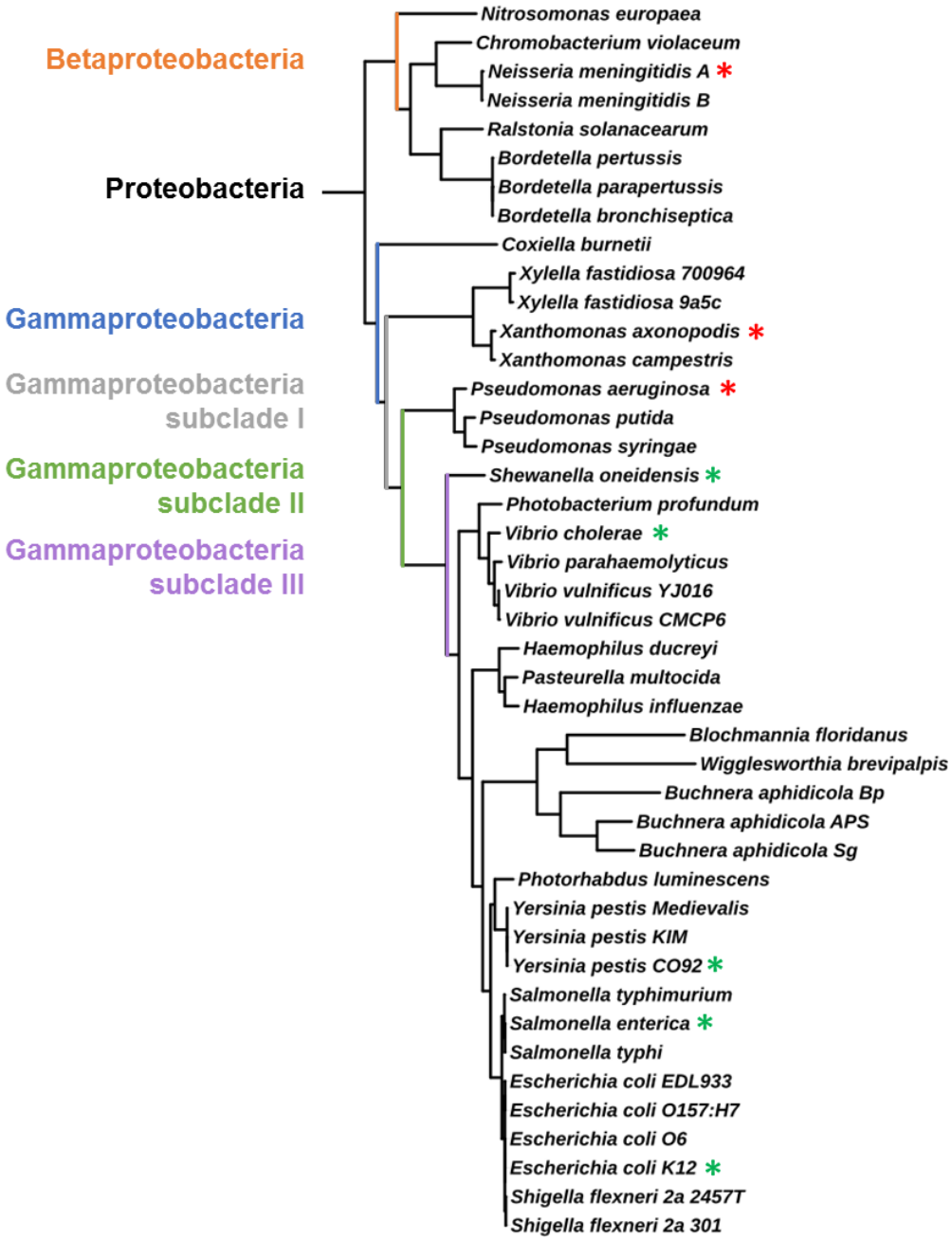


Figure 8-6: Phylogenetic tree of the proteobacteria

The tree was from the https://itol.embl.de/itol.cgi webpage with a focus on the Proteobacteria (Letunic and Bork, 2016). The stars indicated the species whose genomes were detailed in the Figure 8-5 with the *ubi* operon located strictly upstream of the *tat* operon (green) or not (red).

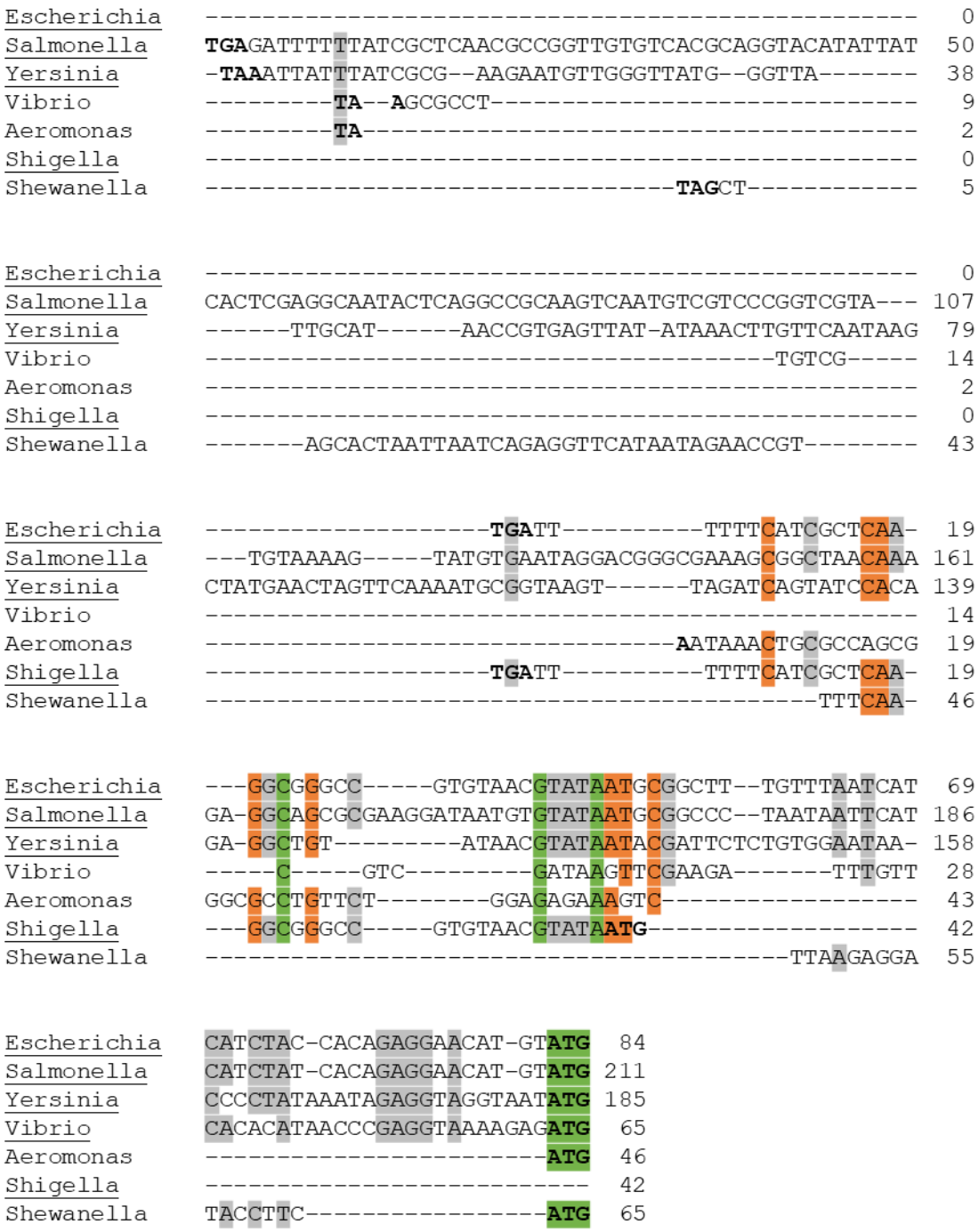


Figure 8-7: Alignment of the intergenic regions between *ubiB* and *tatA*

Intergenic regions between the *ubiB* and *tatA* genes were aligned using the online alignment tool Clustal Omega (Madeira et al., 2019). The regions were selected between the stop codon of *ubiB* and the start codon of *tatA*, both indicated in bold. The bacterial species represented are, in order, *Escherichia coli* K12 W3110 (NC_007779), *Salmonella enterica* (NC_003198), *Yersinia pestis* (NC_003143), *Vibrio cholerae* (NC_002505), *Aeromonas hydrophila* (NC_008570), *Shigella dysenteriae* (NC_007606) and *Shewanella oneidensis* (NC_004347). The underlined species indicate that their respective locus is on the complementary strand. The same bases aligned across 6, 5 and 4 species are highlighted in green, orange and grey respectively.

The species analysed in Figure 8-5 and present in the Gammaproteobacteria subclade III were then further investigated. The sequences of their intergenic region between *ubiB* and *tatA* were aligned (Figure 8-6). This figure reveals that the sequence lengths vary from 42 to 211 bp but that the sequences appeared relatively conserved in patches of homology around *tat* promoter. Moreover, using the BPROM and FindTerm online tools (Solovyev, 2011), a promoter but no Rho-independent terminator was found for each of the members of the Gammaproteobacteria subclade III analysed. Within this data set, all members of this phylum present similar features between *ubiB* and *tatA.* The conservation of this locus may have a biological significance.

8.4. Conclusion regarding the *tat* operon region

This work was initiated from uncertainties regarding the transcriptional control in the TatExpress strain, as developed by Browning et al. (2017). Indeed, the insertion sequence containing the *tac* inducible promoter strictly after *ubiB*'s stop codon suggested that there is no transcriptional terminator between the *ubiB* and *tatA* genes. The possibility of a transcript encompassing both the *ubiEJB* and *tatABCD* gene clusters was investigated by PCR analysis of cDNA derived from mRNA extracted from W3110 WT. Although putative promoters were predicted upstream of *ubiE* and *tatA*, data suggests that the operon organisation is much larger where two types of transcripts can potentially be transcribed due to the absence of a transcriptional terminator between the currently annotated *ubi* and *tat* operons. These are a ≈ 5.5 kbp and ≈ 2.4 kbp corresponding to *ubiEJB-tatABCD* and *tatABCD* respectively. Data supporting this theory was discovered from amplification of cDNA, corresponding to the *ubiJB-tatA* and *ubiB-tatABCD* genes. Moreover, the phylogenetic study revealed an evolutionary conservation of the two operons strictly located next to each other within the Gammaproteobacteria subclade III of which *E. coli* is a member. RNA sequencing was attempted but did not give meaningful data for interpretation. Therefore, this hypothesis requires further confirmation which can be brought by Northern-blot, cDNA sequencing and/or reporter gene assay.

Operons are an evolutionary result of clusters of genes with related functions but the connection between the ubiquinone synthesis and the Tat export pathway remains unclear (Jacob et al., 1960). None of the UbiEJB proteins possess a Tat_{SP} and UbiE and UbiJ are cytoplasmic proteins. The closer physical link between the operon is that the UbiB and TatABC proteins are located in the inner membrane due to α-helix(es) transmembrane domains (Palmer and Berks, 2012, Aussel et al., 2014a). In terms of role, the proteins may be connected through their shared involvement in the respiratory electron transport chain (Patel et al., 2014, Aussel et al., 2014b). Indeed, Tat exports several proteins which requires DSBs in order to be active. Such DSBs are formed by the DsbA protein in the periplasm (section 1.4.1.1). DsbA is then re-oxidized by the inner-membrane protein DsbB which donates the electron received from DsbA to ubiquinone in aerobic conditions (Berkmen, 2012). The electron is finally transferred to the respiratory chain (Cho and Collet, 2013). Through this, a connection might lay between the *ubi* and *tat* operons.

9. GENERAL DISCUSSION AND PERSPECTIVES

9.1. Using a robust fractionation method

The main objective of this project was to evaluate the Tat pathway to export recombinant proteins in *E. coli*. To demonstrate the protein localisation within the cells, the sub-cellular compartments were isolated by fractionation and detected by Western blot with specific antibodies. Hence, the selection of a method giving pure fractions was crucial. Cross-contamination was detected when using the main methods found in the literature and hence, optimisation was carried out to identify a robust technique. The PureFrac method (Malherbe et al., 2019) consists of cold osmotic shock with $MgCl_2$ for periplasmic extraction followed by subsequent sonication to generate the cytoplasmic fraction (Chapter 4). The robustness of these fractions' purity and high recovery of soluble cytoplasmic proteins was demonstrated by tracking the localisation of host control proteins such as the cytoplasmic LacI and periplasmic MBP. Throughout this project, 93.6 % of the expressed cells treated by PureFrac were successfully fractionated without cross-contamination demonstrating high level of robustness and consistency.

9.2. Limitations of this study

The PureFrac method is the most robust *E. coli* fractionation method published to date and is continuously validated by tracking host protein localisation (section 9.1.). However, there are improvements to be made that would positively improve consistency of the data generated. For instance, the addition of a wash step to the insoluble pellet arising after the removal of the cytoplasmic fraction could improve the purity of the insoluble fraction by eliminating potential cytoplasmic contamination. Another limitation in these experiments was to consciously over-expose the Western-Blot membranes to avoid missing any data arising from low intensity bands due to low protein amount. The exposure times were kept strictly identical between each experiment to maintain consistency. The downside was that this resulted in the saturation of the signals which prevented running densitometric analysis (Bass et al., 2017). Densitometry could provide

quantitative or semi-quantitative protein data from Western-Blot images (Gassmann et al., 2009). Finally, this pre-screen methodology was evaluated with only few biological replicates due to the number of samples brought by the high number of conditions tested. Biological replicates are necessary to demonstrate the consistency of a result. However, in the case of PhoA and hGH expressed without a fusion partner, these were repeated as controls in other experiments and show data consistency (Figure 6-2, Figure 7-2 for PhoA and Figure 6-9, Figure 7-4 for hGH).

Regarding the activity assays, these were run with crude fractionated samples and therefore an accurate specific activity based on level of protein expressed could not be determined. Specific activities of β-galactosidase, PhoA and sfGFP would require the purification of the relevant protein prior to the assay. However, in hindsight a semi-quantitative analysis where the specific activity as a measure of the total protein per fractionated sample could have been used which could have enabled a measurable fraction comparison between experiments. Moreover, regarding PhoA, His and Flag purification would be required on the fusion samples since C-terminal degradation was observed. However, as the main focus of this study concentrated on preliminary experiments to evaluate the Tat pathway for export of recombinant proteins, these assays were not attempted due to the size of the sample pools and the small sample volumes. Furthermore, for the other recombinant proteins expressed, correctly folded and therefore active proteins can only be determined by complex cell-based binding assays in the case of hGH and FABP4, or ligand binding assays with respect to the VHHs. Due to the strategic focus of this work, these assays were not attempted. However, the importance of this data with respect to the correct folding of proteins is appreciated and should be an integral part of further studies to understand the mechanism of export of these proteins.

Other culturing methods could also be considered that may impact the quality of expression and export. Indeed initially, the Enpresso® medium (BioSilta) was utilised in an attempt to mimic a fed-batch system by providing a constant and slow release of the carbon source from a polysaccharide (Ukkonen et al., 2017). However, this medium was commercially discontinued during the course of this project. Auto-induction media were also considered to screen a high number of conditions (Fox

and Blommel, 2009, Studier, 2014). But as these media do not enable precise control over the induction period, a necessary factor for reproducibility and to limit inclusion body formation, these media were not used.

Furthermore, the process described in this study from transformation of the expression strains to detection by Western-blot of the proteins of interest in the fractions was time-intensive. Ways to apply a high-throughput process were investigated including the usage of 24 deep-wells plates instead of shake-flasks for cell culture but were unsuccessful. Indeed, in the case of the latter example, the plates presented slower growth presumably due to inefficient gas exchange and the sample volumes were too low to run the required analysis. However, a miniaturisation of the process could still be investigated to pursue this project such as the use of ambr® 15 (Sartorius) or BioLector® (m2p-labs) bioreactors. Another method to have a more high-throughput methodology would be to remove the fractionation step by having a direct assay to detect the protein of interest into the periplasm.

9.3. Exporting recombinant proteins via the Tat pathway

9.3.1. Direct fusion to a signal peptide

Several aspects make the periplasm an attractive location to produce recombinant proteins in *E. coli* including DSB formation and a lower amount of host cell proteins (Chapter 3). To target a protein of interest to this cellular compartment, the protein is fused to a signal peptide at its N-terminus which then targets export via a specific route. The Sec pathway is mainly used but, more recently, interests have been raised towards the Tat pathway because of its so-called proofreading mechanism and ability to export folded proteins (section 1.3.2.5.). However, successful Tat export has only been shown for a few proteins (section 3.1.) which suggests that many proteins fail to export due to the proofreading activity. To alleviate this bottleneck, the strategy used in the present work, involved the fusion of the protein of interest to a soluble partner. The aim was to improve the solubility and

folding of the protein of interest in the cytoplasm, enabling a more successful screening by the Tat proofreading.

NTS were initially considered as the soluble partner for the fusions and five candidates were selected: AmiA, HyaA, PaoA, TorA and YcbK (Chapter 5). Only TorA was successfully exported via the Tat pathway and the reasons why the other candidates failed to do so remains unidentified. The unknown characteristics of the NTS such as their constitutive expression, regulatory elements or requirement for chaperones may critically impact their exportability. Therefore, the remaining NTS were not considered (Appendix 1). Furthermore, recombinant proteins β-galactosidase, PhoA, sfGFP and hGH were considered as reporter protein to play the role of the protein of interest in the fusion design (Chapter 6). Unexpectedly, sfGFP and hGH presented export capabilities without the requirement of a signal peptide as discussed in section 9.7.2. Despite this, hGH and also PhoA were proven to be translocated specifically via the Tat pathway when fused to $TorA_{SP}$. Consequently, the fusion strategy was modified, and the successful reporter proteins were used as the soluble partner in fusion to proteins of interest (section 9.3.2.).

9.3.2. Fusion to a soluble partner

The Table 9-1 summarizes the export results obtained during the present work, with the proteins of interest expressed alone and in fusion with hGH and PhoA (detailed in Chapters 5, 6 and 7). The present work aimed to evaluate whether the fusion of a $TorA_{SP}$-fused soluble partner would improve export via Tat of the protein of interest compared to fusion of solely the TorA signal peptide. Export of the fusions using hGH as a soluble partner and without a signal peptide challenged interpretation of export observations of the $TorA_{SP}$-hGH fusions proteins (section 9.7.2.). Hence, only the fusions using PhoA as a soluble partner were considered with which, no improvement in terms of export were observed. Indeed, out of the four proteins of interest fused to $TorA_{SP}$-PhoA (hGH, FABP4, VHH1 and VHH2), only two proteins (FABP4 and VHH2) have shown signs of export when co-expressed with CyDisCo (Table 9-1). Not only did these two proteins export when fused directly to $TorA_{SP}$, but also one additional protein, VHH1, was exported. Although, Tat specificity needs to be investigated by

using the Tat-null strain and the activity of the proteins of interest needs to be evaluated to assess their native conformation. Moreover, the amount of periplasmic proteins appeared relatively low compared to the cytoplasmic proteins which may be the result of Tat saturation. Equally, the presence of a significant level of protein in the cytoplasm suggests misfolding which might be reduced by co-expressing chaperones (Shriver-Lake et al., 2017).

Overall, over the set of proteins tested in this study, the fusion strategy used to improve export via Tat was unsuccessful. Moreover, export appears to be protein dependent. However, before definitively ruling out the idea of fusing to a carrier protein, more soluble protein candidates and proteins of interest need to be evaluated. Based on the limitations identified in this study, it would be prudent to select proteins that have a simple activity assay as this measure is indicative of a correct folded protein. Furthermore, it would be interesting to understand why FABP4 and VHH2 were translocated when fused to $TorA_{SP}$-PhoA whereas hGH and VHH2 were not (Table 9-1). Since the cytoplasmic and periplasmic PhoA were not active, the fusion proteins were not correctly folded even if the data suggests Tat export. Therefore, $TorA_{SP}$-PhoA-FABP4 and $TorA_{SP}$-PhoA-VHH2 gained a stable conformation in the cytoplasm but not native. Hence, the question remains as per why $TorA_{SP}$-PhoA-hGH and $TorA_{SP}$-PhoA-VHH1 are not in a similar conformation allowing export.

9.3.3. General discussion over the suggested use of the Tat pathway in biotechnology

Three proteins were observed to be transported across the inner membrane without a signal peptide: sfGFP, hGH and FABP4 (Table 9-1), and are discussed in section 9.7.2. Most of the remaining proteins when expressed with $OmpA_{SP}$ were observed in the periplasmic fraction and data suggests they were exported via the Sec pathway (Table 9-1). But most of the proteins expressed with a Tat_{SP} were not exported (YcbK, AmiA, HyaA, PaoA, β-galactosidase, VHH1, hGH-VHH1, PhoA-hGH and PhoA-VHH1) or exported but not specifically ($PoaA_{SP}$-PhoA and $PaoA_{SP}$-sfGFP) via the Tat pathway (Table 9-1) indicating that export via Tat is protein dependent. Tat specific export was justified by the presence of the protein of interest in the periplasm in the WT strain but not in the Tat-null strain. Rejection from the Tat translocon is attributed to the proofreading mechanism (section 1.3.2.5.)

selecting substrates over criteria not yet fully understood. Moreover, the only cases where no export was reported when the recombinant protein was fused to $OmpA_{SP}$ were with β-galactosidase and hGH-VHH1 fusion. In both cases, targeting to the Tat pathway did not offer an alternative to unsuccessful Sec targeting.

The data presented in this study, albeit the number of proteins tested is low, indicates the unpredictability of the Tat export mechanism. For instance, two VHHs with similar biophysical properties (section 7.4.) presented distinct export profiles (Table 9-1). Indeed, VHH1 presented no sign of export with $TorA_{SP}$ whereas VHH2 were detected in the periplasm when fused to this signal peptide whether it was expressed alone or fused to PhoA or hGH. Although these proteins share some similar structural and biophysical motifs, their protein sequences are divergent (Figure 7-9, Table 7-1, Appendix 11). Another example is the NTS PaoA which was anticipated to be exported via the Tat pathway (Lee et al., 2014). However, despite thorough optimisation, no specific export via Tat was obtained to fulfil the expectation and corroborate published results (section 5.2 Furthermore, the specificity of the export mechanism needs to be established via expression in the Tat-null strain. Also, the activity of the exported protein has to be assessed as well as the ratio of active periplasmic proteins over the periplasmic proteins. This would determine whether Tat exports proteins in a native state or at least, in a confirmation allowing native folding in the periplasm.

The present data, added to the short list of successfully exported proteins published, suggest a limited use of the Tat pathway to produce recombinant proteins. Tat has seemingly many advantages over Sec, but the proofreading mechanism which appears to reject many recombinant proteins, needs to be further understood in order to better anticipate export of a protein of interest. For instance, Ulfig and Freudl (2018) presented the importance of the N-terminal residues of the recombinant protein for TatBC binding and thus, translocation (section 1.3.2.4.). Therefore, a comparability study of the N-terminal sequence of the successful and unsuccessful candidates may help understand the acceptance mechanism of the Tat translocon. Moreover, recombinant proteins may require helper factor such as chaperones to be translocated by the Tat pathway which was not the focus of this study.

Table 9-1: Summary of protein export results

The table includes the export profile of all the proteins presented in this work: uncharacterised protein YcbK, N-acetylmuramoyl-L-alanine amidase (AmiA), small chain of hydrogenase-1 (HyaA), aldehyde oxidoreductase iron-sulphur-binding subunit A (PaoA), trimethylamine-N-oxide reductase 1 (TorA), β-galactosidase (β-gal), alkaline phosphatase (PhoA), superfolder green fluorescent protein (sfGFP), human growth hormone (hGH), fatty acid-binding protein 4 (FABP4), VHH1 and VHH2. Some exports were studied in co-expression without (Cy-) or with (Cy+) CyDisCo. The relative exports are represented as no export (-), non-significant export (+/-), low export (+) and high export (++). The green boxes represent Tat specific export confirmed by the lack of export on a Tat-null strain (section 3.6.), the orange boxes are not Tat-specific export, the blue boxes indicate that Tat specificity of the export was not confirmed, and the grey cells represent the conditions that were not tested.

Construct		No SP		$OmpA_{SP}$		$YcbK_{SP}$		$AmiA_{SP}$		$HyaA_{SP}$		$PaoA_{SP}$		$TorA_{SP}$	
Soluble partner	Protein of interest	Cy-	Cy+	Cy-	Cy+	Cy-	Cy+	Cy-	Cy+	Cy-	Cy+	Cy-	Cy+	Cy-	Cy+
-	YcbK					-									
	AmiA							-							
	HyaA									-					
	PaoA											-	-		
	TorA													+	
	β-gal	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	PhoA	-	-	++	++							-	+	-	+
	sfGFP	+		+								+		+	
	hGH	-	+	+	+									+	+
	FABP4	+	+	+	-									+	+
	VHH1	-	-	+	+									-	-
	VHH2	-	-	+/-	+/-									-	+/-
hGH	PhoA	+	++	++	++									+	+
	hGH														
	FABP4	+	+	+	++									+	+
	VHH1	-	+	-	-									-	-
	VHH2	+	+	+	+									+	+
PhoA	PhoA														
	hGH	-	-	++	++									-	-
	FABP4	-	-	+	+									-	+
	VHH1	-	-	-	+									-	-
	VHH2	-	-	+	+									-	+/-

9.4. The evaluation of CyDisCo

This technology was used to form cytoplasmic DSB without disrupting the reducing pathways by co-expressing the sulfhydryl oxidase Erv1p and the isomerase PDI. During the present work, the co-expression of CyDisCo was confirmed to be essential for the Tat export of PhoA in which case this protein was observed active in the periplasm (Table 9-2). However, CyDisCo is suspected to not form the native DSB on PhoA because of the inactivity of this enzyme in the cytoplasm (Table 9-2). Indeed, if CyDisCo formed the native DSB on PhoA prior to Tat export, the cytoplasmic proteins would have

been active. Hence, the periplasmic activity of PhoA is likely the result of the Dsb family periplasmic chaperones activity conferring native conformation to PhoA (section 1.4.1.1.). Furthermore, since cytoplasmic PhoA was not active in the presence of CyDisCo, PhoA was exported via Tat in a non-native conformation. However, Matos et al. (2014) claimed that PhoA was correctly folded when exported by the Tat pathway. In their work, evidence of active cytoplasmic protein was lacking to support this claim.

However, this technology increased PhoA's expression levels and so, was hypothesized to enable DSB formation thanks to the homologous sulfhydryl oxidase but the heterologous isomerase shuffled the DSB into a non-native status. This way PhoA adopts a conformation accepted by the proofreading system of Tat but remains inactive and then shuffles into its native form in the periplasm. Following this hypothesis, PDI could be replaced by an endogenous isomerase like DsbC without its secretion peptide to test this theory. Indeed, when $TorA_{SP}$-PhoA was expressed in the SHuffle strain which expresses cytoplasmic DsbC, PhoA was found active in the cytoplasm. However, PhoA was inactive in the cytoplasm of this strain, when expressed without a signal. Hence, it would be interesting to investigate the folding state of these proteins to understand the impact of DsbC on PhoA folding. For instance, the structural configuration of cytoplasmic PhoA and $TorA_{SP}$-PhoA could be examined using circular dichroism and mass spectrometry.

Another hypothesis was that the inactivity of the cytoplasmic PhoA could have been the result of the sonication explaining also the inactivity of the cytoplasmic sfGFP. Indeed, the cells appeared green upon direct observation (Figure 6-8) suggesting active cytoplasmic protein but sfGFP was found inactive in the cytoplasmic fraction after fractionation (Figure 6-7). This hypothesis could be assessed by lysing the cells using an alternative way like freeze/thaw cycling but with due diligence to maintaining the fractions purity. However, CyDisCo improved the solubility of most of the recombinant proteins used in this project as seen by the higher intensity of the bands with CyDisCo than without. The benefit of co-expressing chaperones on recombinant protein expression was not surprising (Ellis et al., 2017, Kang et al., 2012, de Marco, 2007) but these chaperones have proved their efficacy in improving the expression of soluble proteins.

Table 9-2: Impact of CyDisCo on the activity of PhoA

The table summarizes the activity of PhoA reported in the present work. PhoA activity was analysed in the cytoplasmic (C) and periplasmic (P) fractions from cells expressing protein of interests, potentially fused to a soluble partner, fused to no signal peptide (No SP), the signal peptide of OmpA ($OmpA_{SP}$) or TorA ($TorA_{SP}$) and co-expressed with CyDisCo (Cy+) or not (Cy-). Active (+) and inactive (-) PhoA was defined via a 250 U/µl threshold.

Soluble partner	Protein of interest	No SP				$OmpA_{SP}$				$TorA_{SP}$				Reference
		Cy-		Cy+		Cy-		Cy+		Cy-		Cy+		
		C	P	C	P	C	P	C	P	C	P	C	P	
-	PhoA	-	-	-	-	-	+	-	+	-	-	-	+	Figure 6-2B
-	PhoA	-	-	-	-	-	+	-	+	-	-	-	+	Figure 6-4B
-	PhoA	-	-	-	-	-	+	-	+	-	-	-	+	Figure 7-2
hGH	PhoA	-	-	-	-	-	+	-	+	-	-	-	-	
PhoA	hGH	-	-	-	-	-	+	-	+	-	-	-	-	Figure 7-4
PhoA	FABP4	-	-	-	-	-	+	-	+	-	-	-	-	Figure 7-6
PhoA	VHH1	-	-	-	-	-	+	-	+	-	-	-	-	Figure 7-10
PhoA	VHH2	-	-	-	-	-	+	-	+	-	-	-	-	Figure 7-12

During this project, PDI-cMyc and sometimes Erv1p-cMyc located in the periplasm unexpectedly since they do not possess a known signal peptide (Table 9-3). In all cases, the cytoplasmic marker LacI was not reported in the periplasm so the periplasmic presence of CyDisCo was not due to cell lysis during fractionation but rather due to a specific translocation mechanism during growth. VHH2 is an exception but in this case, LacI was suspected to contaminate the periplasm due to cell lysis during fractionation. No prior observation of the periplasmic localisation of these chaperones has been reported. This export was not observed when CyDisCo was expressed alone (Table 9-3) indicating that the co-expression of a recombinant protein was somehow involved in the translocation of the CyDisCo proteins. The common feature between the recombinant proteins triggering CyDisCo export was the presence of OmpASP and/or N-terminal hGH with the exception of TorASP-αIL17 (Figure 6-9; Appendix 6; Appendix 8; Appendix 9; Appendix 12 and Appendix 13).

The signal peptide of OmpA is known to bind to SecB which then recruits chaperones such as GroEL, DnaK and TF (section 7.2.2.). Using a similar interaction, the PDI chaperone and, to a lower extent, Erv1p might be recruited by $OmpA_{SP}$ which would explain the higher translocation rate of the chaperones in presence of overexpressed $OmpA_{SP}$-fused proteins. However, this hypothesis does not yet explain how the protein crosses the inner membrane. With this signal peptide, only a small

proportion of these proteins were found in the periplasm. Hence, the question remains as per what it different about these proteins which explains why they were translocated whereas the remaining proteins remained in the cytoplasm. PDI and Erv1p were not exported via the Tat pathway since $OmpA_{SP}$ was demonstrated to be Tat-incompatible (Huang and Palmer, 2017) and a putative signal peptide sequence for export in Gram-negative bacteria was not detected on either PDI nor Erv1p using signal peptide prediction software SignalP (Nielsen, 2017b) (Table 9-4). The involvement of the N-terminus sequence of the CyDisCo chaperones could be evaluated by expression of N-terminal truncated mutant. Subsequently, other Sec signal peptides known to interact with cytoplasmic chaperones could be tested in place of $OmpA_{SP}$ to investigate whether it is the chaperone-recruiting ability of a Sec signal peptide which allows export of PDI and Erv1p. Moreover, this uncharacterised export might not be limited to the CyDisCo chaperones and hence, looking at the localisation of other chaperones such DnaK, GroEL and TF would allow to better grasp this mechanism.

9.5. The conserved large operon

Evidences supporting the idea that no terminator was present between the *ubiEJB* and the *tatABCD* operons in *E. coli* were presented in Chapter 8. Indeed, PCR performed on cDNA demonstrated the presence of mRNA containing genes from both operons (section 8.2.). Also, a phylogenetic study revealed that the arrangement of these genes on the chromosome is conserved amongst the Gammaproteobacteria subclade III phylum which includes *E. coli* (section 8.3.). Further analysis of this locus suggests that the absence of a terminator between these operons is conserved within this phylum.

Table 9-3: Periplasmic observations of the CyDisCo chaperones

The table represents the detection of PDI and Erv1p in the periplasmic fractions when expressed from an empty vector (no construct) or co-expressed with the detailed constructs. The latter contained a protein of interest: alkaline phosphatase (PhoA), human growth hormone (hGH), fatty acid-binding protein 4 (FABP4), VHH1 or VHH2. Some of the constructs also contained a fusion partner: hGH, PhoA or none (Ø). The constructs were either fused to no signal peptide (No SP), the signal peptide of OmpA ($OmpA_{SP}$) or of TorA ($TorA_{SP}$). The relative exports are represented as no export (-), non-significant export (+/-), low export (+) and high export (++).

Reference	Periplasmic localisation when co-expressed with									
	No Construct		Construct							
	PDI	Erv1p	Soluble partner	Protein of interest	No SP		$OmpA_{SP}$		$TorA_{SP}$	
					PDI	Erv1p	PDI	Erv1p	PDI	Erv1p
Figure 6-2	-	-	Ø	PhoA	-	-	-	-	-	-
Figure 6-4	-	-	Ø	PhoA	-	-	-	-	-	-
Figure 6-9	-	-	Ø	hGH	-	-	+	+/-	-	-
Appendix 6	-	-	Ø	PhoA	-	-	-	-	-	-
	-	-	hGH	PhoA	-	-	+	+/-	+/-	-
Appendix 8	-	-	Ø	hGH	-	-	++	+/-	-	-
	-	-	PhoA	hGH	-	-	+	-	-	-
Appendix 9	-	-	Ø	FABP4	-	-	-	-	-	-
	-	-	hGH	FABP4	-	-	+/-	-	-	-
	-	-	PhoA	FABP4	-	-	-	-	-	-
Appendix 12	-	-	Ø	VHH1	-	-	+	+/-	+	-
	-	-	hGH	VHH1	+/-	-	+/-	-	+/-	-
	-	-	PhoA	VHH1	-	-	-	-	-	-
Appendix 13	-	-	Ø	VHH2	-	-	+	+	-	-
	-	-	hGH	VHH2	-	-	+	+/-	-	-
	-	-	PhoA	VHH2	-	-	-	-	-	-

Table 9-4: Comparison of N-terminal sequences

The sequence in blue corresponds to the signal peptide of OmpA.

Protein	30 N-terminal residues
OmpA	**MKKTAIAIAVALAGFATVAQAAPKDNTWYT**
PDI	**MDAPEEEDHVLVLRKSNFAEALAAHKYLLV**
Erv1p	**MKAIDKMTDNPPQEGLSGRKIIYDEDGKPC**
hGH	**MFPTIPLSRLFDNAMLRAHRLHQLAFDTYQ**

However, more evidence should be brought before drawing a final conclusion with regards to this hypothesis. For instance, the cDNA obtained could be sequenced to confirm the presence of the genes within the same mRNA. Also, Northern-blot analysis could further confirm that the transcription of the *ubiEJB* operon does not stop between the *ubiB* and the *tatA* genes. Most importantly, reporter gene assays should be performed to experimentally validate the absence of a terminator but also to confirm the presence of an active promoter upstream of both *ubiB* and *tatA*. To do so, the regions upstream of the *ubiE* and *tatA* genes would be cloned separately upstream of a reporter gene such as PhoA or β-galactosidase to assess the presence of a functional promoter. Similarly, the region downstream of the *ubiB* gene would be cloned between two reporter genes in a polycistronic cassette to evaluate its transcription termination capacity. Moreover, assessing whether both predicted promoters are constitutive or induced would shed light on their respective function. The promoter upstream of the *tat* operon would be expected to be constitutive since NTS are relevant at multiple cell life cycle stages (section 1.3.2.1.).

Finally, if both operons can be expressed under the same promoter and this is a conserved phenomenon, a common role is likely to exist. To evaluate a potential interaction between the proteins encoded by these genes, immunofluorescence, co-immunoprecipitation, pull down assay or FRET can be used. Also, RT-qPCR could help determine the transcription levels of these genes when the bacteria are in different conditions such as anaerobiosis, oxidising stress or carbon source depletion.

9.6. Conclusion

This project brought a better understanding over recombinant protein production in *E. coli* regarding fractionation methods, the CyDisCo technology and export pathways. Thus, a new method for fractionation was developed, with enough robustness to examine protein export. The CyDisCo technology did not fulfil its claimed ability to form correct DSB in the cytoplasm. Indeed, it was demonstrated to improve protein expression but the recombinant cytoplasmic proteins whose

activity was assessed were inactive indicating incorrect folding. Regarding the Tat export pathway, it was proven to be difficult to translocate recombinant protein, even its own substrates. The fusion strategy did not present any improvement in terms of export but highlights the need for additional optimisation. Tat proofreading mechanism appeared very selective and so, recombinant protein production using this pathway would first require a greater comprehension of this ability.

Indeed, the investigation in this study to determine the transcriptional elements controlling expression of the Tat machinery may spark some interest. Data suggests the absence of terminator between the *tat* operon and the upstream *ubi* operon. This feature appears to be conserved amongst a certain phylum suggesting of a related function between these genes which remains to be determined.

This work has also revealed that the proteins sfGFP, FABP4 and hGH are able to cross the inner membrane without the need for a signal peptide. This discovery might lead to the characterisation of a new export mechanism(s). Furthermore, the recombinant proteins fused to the C-terminus of hGH were shown to be exported without a signal peptide which may offer an alternative way to target proteins of interest the periplasm. Overall, the present study demonstrated that expression and export is protein dependent and hence, Tat translation needs to be optimised for each protein of interest. Also, the published repertoire of successful Tat export was not increase.

9.7. Future perspectives

From the work presented in this thesis, two major perspectives appear promising. The first one would be the pursuit of the fusion strategy to export recombinant proteins into the periplasm via Tat and is discussed in section 9.7.1. The second one would be the identification of the unknown mechanism(s) responsible for export of the sfGFP, hGH fusions, and FABP4 proteins without a signal peptide and is presented in section 9.7.2.

9.7.1. Optimize the fusion design

Using fusions to enhance protein production and quality is not new and has been used for cytoplasmic expressions as well as periplasmic via Sec (Kosobokova et al., 2016, Malik, 2016). Although the results shown in section 9.3. may caution the use of a fusion strategy to improve Tat export, this work is still a pilot study and hence, there is room for improvements. A major drawback observed was the proteolytic clipping (section 7.5.). Indeed, the proteins expressed in fusion showed no sign of degradation when expressed alone except PhoA (Chapter 7). Since truncated products appeared when these proteins were expressed as fusions, the linker which brings the independently stable proteins together could be optimised. This linker was designed to be flexible and long to allow space for folding of both proteins. However, proteins folding and hence activity require the linker to be either long or short linker and either flexible or rigid (Chen et al., 2013). The current linker contains 34 residues which is considered relatively long (Chen et al., 2013), so it could be truncated to test a shorter length. A specific cleavage site is mandatory and TEV cleavage site is one of the shortest and most specific one and hence, should not be altered. However, initial studies should include a control protein bearing a linker without a TEV site to investigate if this motif impacts proteolytic cleavage. The FLAG detection tag (section 3.2.), however, could be replaced for the shorter more traditional version (DYKDDDDK) or by V5 or HA to keep a high detection sensitivity (Kosobokova et al., 2016). Furthermore, a variety of linkers using a GS rich sequence for flexibility (Argos, 1990) or a (EAAAK)n α-helix type for rigidity could be tested (Bai and Shen, 2006, Amet et al., 2009). Another reason for using a more rigid linker is that the proofreading mechanism of Tat was considered to “sense” the rigidity of the substrate and dismiss them when too flexible (Jones et al., 2016, Sutherland et al., 2018). Also, variants without the FLAG tag or the TEV cleavage site could bring a better understanding of the impact from these sections compared to the full-length linker.

To evaluate a new range of linkers, the fusion proteins chosen should be, in theory, capable of being exported by Tat. Hence, following the present results (Chapter 6 and 7), a fusion protein containing candidates successfully exported via Tat is recommended. For instance, the fusions PhoA-hGH or hGH-PhoA could be considered. Indeed, these proteins have been shown to be exported Tat

specifically and PhoA was active in the periplasm. Furthermore, the usage of signal peptide of TorA should be continued as it exhibited the highest success rate with regards to Tat specific export.

Another parameter which could reduce the fusions truncation is the culture condition. Indeed, low cultivation temperatures can reduce or impair protein degradation due to a poor activity of heat shock proteases that are usually induced during protein overproduction in E. coli (Chesshyre and Hipkiss, 1989, Costa et al., 2014). Moreover, a lowered induction level would reduce the cellular stress which might improve the folding and therefore, the export yield of the protein of interest. On a related note, over-expressing the Tat proteins from either a plasmid (Matos et al., 2012) or an engineered strain (Browning et al., 2017) might relieve the membrane stress due to jamming the export translocases. To this aim, induction at lower temperature for a longer time was tested in sections 4.4., 5.2.4. and 5.2.5. However, this condition did not impact positively expression nor export and even had a negative impact on PaoA as less proteins was detected after a 16h induction at 22°C than 2 or 4 h at 30°C (Figure 5-5).

9.7.2. Characterise the unknown mechanism(s) of export

Table 9-1 highlights an unexpected phenomenon: the export of proteins with no signal peptide. Through this project, three proteins were identified with “auto-export” properties: sfGFP (section 6.4.), hGH (section 6.5.) and FABP4 (section 7.3.). Furthermore, hGH-fused proteins were also shown to auto export (Chapter 7). The online tool SignalP (Nielsen, 2017b) predicted no putative signal peptide sequence in their N-termini. The case of sfGFP was investigated by Zhang et al. (2017) who discovered that this protein did not require its 20 N-terminal residues to export. Therefore, the “auto-export” is allowed by another means than a N-terminal motif. A similar evaluation could be conducted on the remaining “auto-exporters” to assess the presence of an unpredicted N-terminal signal peptide.

Also, Zhang’s team found sfGFP to be located in the extra-cellular medium after a 24 h induction period. They demonstrated that the secretion of sfGFP was not the result of cell lysis by the absence of detectable cytoplasmic GroEL and periplasmic β-lactamase markers in the extracellular medium

(Zhang et al., 2017). The amount of sfGFP observed in the extracellular medium was substantial and so, could be the result of cellular stress due to overexpression. To test this possibility, the expression level of sfGFP should be lowered, e.g. by using an origin of replication with a very low copy number such as SC101 (Friehs, 2004). Also, it would be interesting to extend the induction times of the other "auto-exporting" proteins listed here to assess their "auto-secretion". Moreover, since the sfGFP was detected to cross both membranes, identifying all the compartments in which these "auto-exporters" locate (cytoplasm, inner membrane, periplasm, outer membrane and extra-cellular medium) would bring a better understanding on the phenomenon.

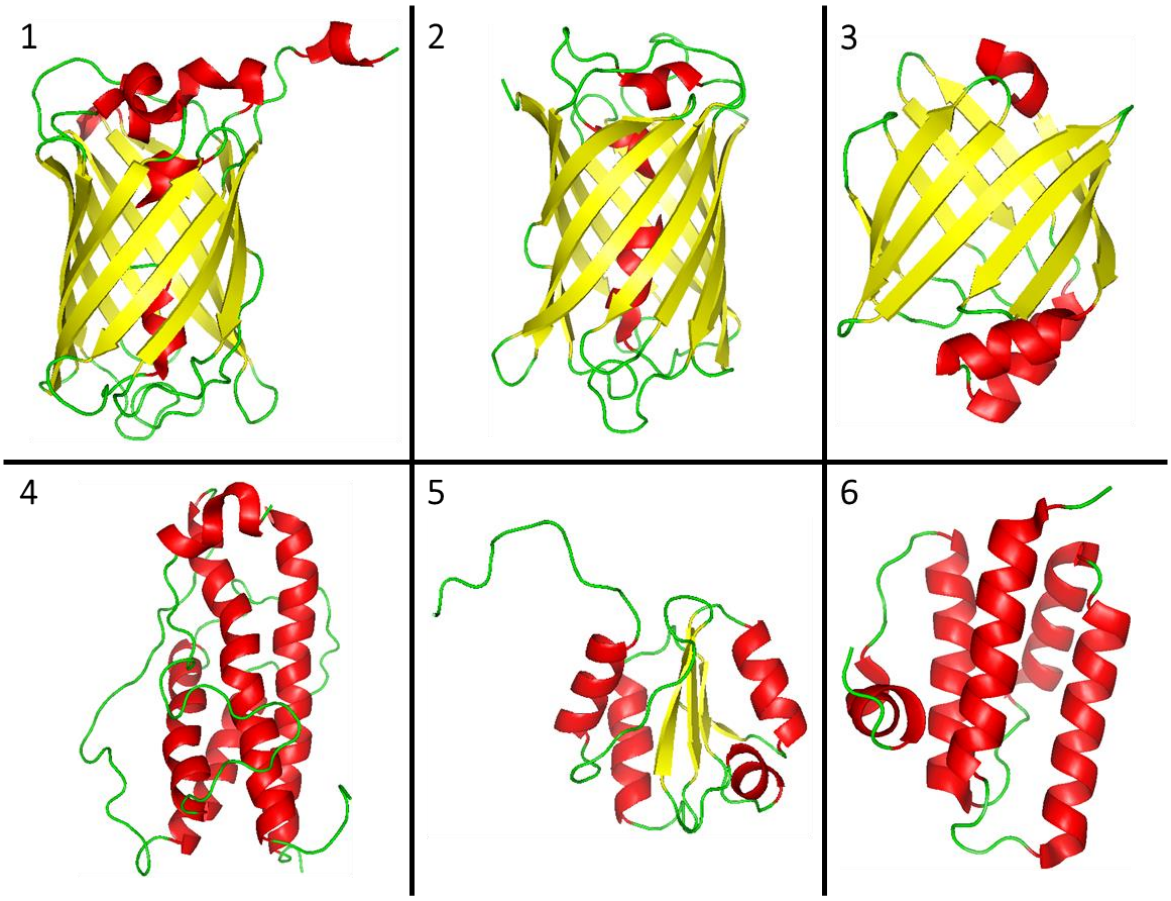


Figure 9-1: Structures of non-conventional exporting proteins

The structures of superfolder green fluorescent protein, mCherry, fatty acid binding protein 4, human growth hormone, human protein disulphide isomerase and sulfhydryl oxidase Erv1p from *Saccharomyces cerevisiae* represented from 1 to 6 were solved by X-ray crystallography at 1.85, 1.6, 1.08, 2.5 and 2Å respectively except Erv1p which was solved by NMR (Dippel et al., 2016, Shu et al., 2006, González and Fisher, 2015, Chantalat et al., 1995, Kemmink et al., 1999, Guo et al., 2012). The structures represent the following residues: 1-238, 1-236, 1-132, 27-217, 136-345, 86-189 respectively. The secondary structures are coloured with the β-sheets in yellow, the α-helixes in red and the coil in green.

Zhang et al. (2017) did not identify how sfGFP was secreted but they discovered that the β-barrel structure (Figure 9-1) was essential for secretion. Hence, they expressed mCherry, a protein with a similar β-barrel structure (Figure 9-1), and identified it in the extra-cellular medium (Shu et al., 2006). They did not express a protein that does not possess a β-barrel structure, but they discovered that sfGFP was no longer secreted when its structure was impaired. This structure is shared with FABP4 (Figure 9-1) and thus, the involvement of this particular structure could be investigated on FABP4. Similar to the strategy from Zhang et al. (2017), the impact of specific mutations impairing the β-barrel structure on the "auto-export" ability could be assessed.

Zhang et al. (2017) suggested that sfGFP and mCherry are secreted thanks to their β-barrel structure. This theory could apply to FABP4 as well but not to hGH which showed "auto-exported" properties but does not share the same 3D structure (Figure 9-1). Therefore, either the theory from Zhang et al. (2017) is incomplete or these proteins are exported via different means. The unknown translocation mechanism by which hGH can be exported appeared to be used only if this protein is on the N-terminal of the fusion but not on the C-terminal (Chapter 7). However, no signal peptide was predicted at the N-terminus of hGH (section 6.5.). Furthermore, the data suggests that hGH is exported via the Tat pathway when expressed in fusion with $TorA_{SP}$. This was demonstrated by the presence of mature protein in the periplasm suggesting that $TorA_{SP}$ was processed after Tat translocation while data suggests that the unknown pathway does not process the signal peptide. A similar profile was observed for sfGFP and FABP4 suggesting a common mechanism of export is used by both proteins. This also means that the signal peptide of TorA prevents export via the unknown pathway(s). This observation would potentially help characterizing this mechanism. Indeed, mutant versions of $TorA_{SP}$ can be assessed as well as other Tat_{SP} to determine the characteristics of the N-terminal residues that lead to export via the unknown mechanism. Such mutations could include N-terminal truncated versions, replacement of the twin arginine by lysine residues, increasing or decreasing the net charge and the hydrophobicity of the signal peptide. Moreover, data suggests that $OmpA_{SP}$ did not prevent the unknown translocation events. Hence, this signal peptide could be altered to allow Tat translocation to further investigate the impact of the N-terminal region. This can

be achieved by decreasing the hydrophobicity of the h-region (section 1.3.1.1.) and creating the consensus twin-arginine motif as inspired by (Cristóbal et al., 1999).

Also, hGH presented the ability to carry all the recombinant proteins fused to its C-terminus during this project (Table 7-2). However, in the case of hGH-PhoA, the reporter protein PhoA was not active in the periplasm. This fusion protein was also found in the periplasm when expressed with $OmpA_{SP}$ and $TorA_{SP}$ but the PhoA component was only active in the periplasm from expression of $OmpA_{SP}$-hGH-PhoA (Table 9-1 and Table 9-2). However, it was not established if the activity came from the full-length protein or a truncated species containing PhoA (section 7.2.2.). Since PhoA is naturally a dimer, it is possible that fusion to hGH may have hindered dimer formation (section 6.3.), thus preventing activity. In order to test this hypothesis, the biological functionality of the other piggy-backed proteins, FABP4 and the VHHs which are monomeric, could be determined. Nevertheless, the present work demonstrated the possibility of exporting recombinant proteins in the periplasm and perhaps beyond in a piggy-back fashion by hGH and potentially FABP4 which could be a promising alternative for recombinant protein production.

Another mechanism by which proteins were reported to export without a signal peptide was considered. The protein localization (*prl*) mutations are located on the Sec pathway proteins such as SecA, SecE, SecG and SecY (section 1.3.1.). SecY mutations named *prlA* mutations allow export of proteins with defective or even no signal peptide without altering export of natural substrates (Derman et al., 1993, Flower et al., 1994, Bost and Belin, 1997, van der Wolk et al., 1998, Smith et al., 2005). A patent was granted for particular *prl* mutations on SecY for exporting antibodies, antibody fragments, or antibody-related polypeptides from prokaryotes without a signal peptide (Yarranton and Bebbington 2005). However, genome sequencing of the strains used in the present project (section 3.6.) revealed that the *secY* genes contained none of the published *prlA* mutations (Appendix 14) (Smith et al., 2005).

BIBLIOGRAPHY

ABERG, A., HAHNE, S., KARLSSON, M., LARSSON, A., ORMO, M., AHGREN, A. & SJOBERG, B. M. 1989. Evidence for two different classes of redox-active cysteines in ribonucleotide reductase of Escherichia coli. *J Biol Chem,* 264**,** 12249-52.

ADAMS, H., SCOTTI, P. A., DE COCK, H., LUIRINK, J. & TOMMASSEN, J. 2002. The presence of a helix breaker in the hydrophobic core of signal sequences of secretory proteins prevents recognition by the signal-recognition particle in Escherichia coli. *Eur J Biochem,* 269**,** 5564-71.

AKAZAWA-OGAWA, Y., UEGAKI, K. & HAGIHARA, Y. 2016. The role of intra-domain disulfide bonds in heat-induced irreversible denaturation of camelid single domain VHH antibodies. *J Biochem,* 159**,** 111-21.

ALAMI, M., LUKE, I., DEITERMANN, S., EISNER, G., KOCH, H. G., BRUNNER, J. & MULLER, M. 2003. Differential interactions between a twin-arginine signal peptide and its translocase in Escherichia coli. *Mol Cell,* 12**,** 937-46.

ALANEN, H. I., WALKER, K. L., LOURDES VELEZ SUBERBIE, M., MATOS, C. F., BONISCH, S., FREEDMAN, R. B., KESHAVARZ-MOORE, E., RUDDOCK, L. W. & ROBINSON, C. 2015. Efficient export of human growth hormone, interferon alpha2b and antibody fragments to the periplasm by the Escherichia coli Tat pathway in the absence of prior disulfide bond formation. *Biochim Biophys Acta,* 1853**,** 756-63.

ALBINIAK, A. M., MATOS, C. F., BRANSTON, S. D., FREEDMAN, R. B., KESHAVARZ-MOORE, E. & ROBINSON, C. 2013. High-level secretion of a recombinant protein to the culture medium with a Bacillus subtilis twin-arginine translocation system in Escherichia coli. *Febs j,* 280**,** 3810-21.

ALCOCK, F., BAKER, M. A., GREENE, N. P., PALMER, T., WALLACE, M. I. & BERKS, B. C. 2013. Live cell imaging shows reversible assembly of the TatA component of the twin-arginine protein transport system. *Proc Natl Acad Sci U S A,* 110**,** E3650-9.

ALLEN, R. S., LI, J., STAHLE, M. I., DUBROUE, A., GUBLER, F. & MILLAR, A. A. 2007. Genetic analysis reveals functional redundancy and the major target genes of the Arabidopsis miR159 family. *Proc Natl Acad Sci U S A,* 104**,** 16371-6.

ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res,* 25**,** 3389-402.

AMET, N., LEE, H. F. & SHEN, W. C. 2009. Insertion of the designed helical linker led to increased expression of tf-based fusion proteins. *Pharm Res,* 26**,** 523-8.

ARGOS, P. 1990. An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. *J Mol Biol,* 211**,** 943-58.

ARIGONI, F., POGLIANO, K., WEBB, C. D., STRAGIER, P. & LOSICK, R. 1995. Localization of protein implicated in establishment of cell type to sites of asymmetric division. *Science,* 270**,** 637-40.

ARONSON, D. E., COSTANTINI, L. M. & SNAPP, E. L. 2011. Superfolder GFP is fluorescent in oxidizing environments when targeted via the Sec translocon. *Traffic,* 12**,** 543-8.

ARYA, R., SABIR, J. S., BORA, R. S. & SAINI, K. S. 2015. Optimization of culture parameters and novel strategies to improve protein solubility. *Methods Mol Biol,* 1258**,** 45-63.

ASHOK KUMA, T. 2013. CFSSP: Chou and Fasman Secondary Structure Prediction server. *WIDE SPECTRUM: Research Journal,* 1**,** 15-19.

AUCLAIR, S. M., BHANU, M. K. & KENDALL, D. A. 2012. Signal peptidase I: cleaving the way to mature proteins. *Protein Sci,* 21**,** 13-25.

AUSSEL, L., LOISEAU, L., HAJJ CHEHADE, M., POCACHARD, B., FONTECAVE, M., PIERREL, F. & BARRAS, F. 2014a. ubiJ, a new gene required for aerobic growth and proliferation in macrophage, is involved in coenzyme Q biosynthesis in Escherichia coli and Salmonella enterica serovar Typhimurium. *J Bacteriol,* 196**,** 70-9.

AUSSEL, L., PIERREL, F., LOISEAU, L., LOMBARD, M., FONTECAVE, M. & BARRAS, F. 2014b. Biosynthesis and physiology of coenzyme Q in bacteria. *Biochim Biophys Acta,* 1837**,** 1004-11.

BAESHEN, M. N., AL-HEJIN, A. M., BORA, R. S., AHMED, M. M., RAMADAN, H. A., SAINI, K. S., BAESHEN, N. A. & REDWAN, E. M. 2015. Production of Biopharmaceuticals in E. coli: Current Scenario and Future Perspectives. *J Microbiol Biotechnol,* 25**,** 953-62.

BAI, Y. & SHEN, W. C. 2006. Improving the oral efficacy of recombinant granulocyte colony-stimulating factor and transferrin fusion protein by spacer optimization. *Pharm Res,* 23**,** 2116-21.

BALASUNDARAM, B., HARRISON, S. & BRACEWELL, D. G. 2009. Advances in product release strategies and impact on bioprocess design. *Trends Biotechnol,* 27**,** 477-85.

BARNETT, J. P., EIJLANDER, R. T., KUIPERS, O. P. & ROBINSON, C. 2008. A minimal Tat system from a gram-positive organism: a bifunctional TatA subunit participates in discrete TatAC and TatA complexes. *J Biol Chem,* 283**,** 2534-42.

BASS, J. J., WILKINSON, D. J., RANKIN, D., PHILLIPS, B. E., SZEWCZYK, N. J., SMITH, K. & ATHERTON, P. J. 2017. An overview of technical considerations for Western blotting applications to physiological research. *Scandinavian Journal of Medicine & Science in Sports,* 27**,** 4-25.

BASSFORD, P. J. J. 1990. Export of the Periplasmic Maltose-Binding Protein of *Escherichia coli*. *J. Bionenerg. Biomembr,* 22**,** 401-39.

BECKER, G. W. & HSIUNG, H. M. 1986. Expression, secretion and folding of human growth hormone in Escherichia coli. Purification and characterization. *FEBS Lett,* 204**,** 145-50.

BERKMEN, M. 2012. Production of disulfide-bonded proteins in Escherichia coli. *Protein Expr Purif,* 82**,** 240-51.

BERKS, B. C. 1996. A common export pathway for proteins binding complex redox cofactors? *Mol Microbiol,* 22**,** 393-404.

BERKS, B. C. 2015. The twin-arginine protein translocation pathway. *Annu Rev Biochem,* 84**,** 843-64.

BERKS, B. C., SARGENT, F. & PALMER, T. 2000. The Tat protein export pathway. *Mol Microbiol,* 35**,** 260-74.

BERNHARDT, T. G. & DE BOER, P. A. J. 2003. The Escherichia coli amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway. *Molecular Microbiology,* 48**,** 1171-1182.

BERTHELMANN, F. & BRUSER, T. 2004. Localization of the Tat translocon components in Escherichia coli. *FEBS Lett,* 569**,** 82-8.

BESSETTE, P. H., ASLUND, F., BECKWITH, J. & GEORGIOU, G. 1999. Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. *Proc Natl Acad Sci U S A,* 96**,** 13703-8.

BEVER, C. S., DONG, J. X., VASYLIEVA, N., BARNYCH, B., CUI, Y., XU, Z. L., HAMMOCK, B. D. & GEE, S. J. 2016. VHH antibodies: emerging reagents for the analysis of environmental chemicals. *Anal Bioanal Chem,* 408**,** 5985-6002.

BEWLEY, T. A., BROVETTO-CRUZ, J. & LI, C. H. 1969. Human pituitary growth hormone. XXI. Physicochemical investigations of the native and reduced-alkylated protein. *Biochemistry,* 8**,** 4701-4708.

BLUMMEL, A. S., DREPPER, F., KNAPP, B., EIMER, E., WARSCHEID, B., MULLER, M. & FROBEL, J. 2017. Structural features of the TatC membrane protein that determine docking and insertion of a twin-arginine signal peptide. *J Biol Chem,* 292**,** 21320-21329.

BLUMMEL, A. S., HAAG, L. A., EIMER, E., MULLER, M. & FROBEL, J. 2015. Initial assembly steps of a translocase for folded proteins. *Nat Commun,* 6**,** 7234.

BORUKHOV, S. & NUDLER, E. 2003. RNA polymerase holoenzyme: structure, function and biological implications. *Curr Opin Microbiol,* 6**,** 93-100.

BOST, S. & BELIN, D. 1997. prl mutations in the Escherichia coli secG gene. *J Biol Chem,* 272**,** 4087-93.

BROWN, N. L., BARRETT, S. R., CAMAKARIS, J., LEE, B. T. & ROUCH, D. A. 1995. Molecular genetics and transport analysis of the copper-resistance determinant (pco) from Escherichia coli plasmid pRJ1004. *Mol Microbiol,* 17**,** 1153-66.

BROWNING, D. F., RICHARDS, K. L., PESWANI, A. R., ROOBOL, J., BUSBY, S. J. W. & ROBINSON, C. 2017. Escherichia coli "TatExpress" strains super-secrete human growth hormone into the bacterial periplasm by the Tat pathway. *Biotechnol Bioeng,* 114**,** 2828-2836.

BURAK, M. F., INOUYE, K. E., WHITE, A., LEE, A., TUNCMAN, G., CALAY, E. S., SEKIYA, M., TIROSH, A., EGUCHI, K., BIRRANE, G., LIGHTWOOD, D., HOWELLS, L., ODEDE, G., HAILU, H., WEST, S., GARLISH, R., NEALE, H., DOYLE, C., MOORE, A. & HOTAMISLIGIL, G. S. 2015. Development of a therapeutic monoclonal antibody that targets secreted fatty acid-binding protein aP2 to treat type 2 diabetes. *Sci Transl Med,* 7**,** 319ra205.

BURGER, G., FORGET, L., ZHU, Y., GRAY, M. W. & LANG, B. F. 2003. Unique mitochondrial genome architecture in unicellular relatives of animals. *Proceedings of the National Academy of Sciences,* 100**,** 892.

BUTLER, M. & SPEARMAN, M. 2014. The choice of mammalian cell host and possibilities for glycosylation engineering. *Current Opinion in Biotechnology,* 30**,** 107-112.

BUTTNER, D. 2012. Protein export according to schedule: architecture, assembly, and regulation of type III secretion systems from plant- and animal-pathogenic bacteria. *Microbiol Mol Biol Rev,* 76**,** 262-310.

CALDELARI, I., PALMER, T. & SARGENT, F. 2008. Escherichia coli tat mutant strains are able to transport maltose in the absence of an active malE gene. *Arch Microbiol,* 189**,** 597-604.

CAO, H., SEKIYA, M., ERTUNC, M. E., BURAK, M. F., MAYERS, J. R., WHITE, A., INOUYE, K., RICKEY, L. M., ERCAL, B. C., FURUHASHI, M., TUNCMAN, G. & HOTAMISLIGIL, G. S. 2013. Adipocyte lipid chaperone AP2 is a secreted adipokine regulating hepatic glucose production. *Cell Metab,* 17**,** 768-78.

CARRIE, C., WEISSENBERGER, S. & SOLL, J. 2016. Plant mitochondria contain the protein translocase subunits TatB and TatC. *J Cell Sci,* 129**,** 3935-3947.

CARRIO, M. M. & VILLAVERDE, A. 2003. Role of molecular chaperones in inclusion body formation. *FEBS Lett,* 537**,** 215-21.

CASCALES, E. & CHRISTIE, P. J. 2003. The versatile bacterial type IV secretion systems. *Nat Rev Microbiol,* 1**,** 137-49.

CASPERS, M., BROCKMEIER, U., DEGERING, C., EGGERT, T. & FREUDL, R. 2010. Improvement of Sec-dependent secretion of a heterologous model protein in Bacillus subtilis by saturation mutagenesis of the N-domain of the AmyE signal peptide. *Appl Microbiol Biotechnol,* 86**,** 1877-85.

CASTANIE-CORNET, M. P., BRUEL, N. & GENEVAUX, P. 2014. Chaperone networking facilitates protein targeting to the bacterial cytoplasmic membrane. *Biochim Biophys Acta,* 1843**,** 1442-56.

CESARATTO, F., BURRONE, O. R. & PETRIS, G. 2016. Tobacco Etch Virus protease: A shortcut across biotechnologies. *J Biotechnol,* 231**,** 239-249.

CHALFIE, M., TU, Y., EUSKIRCHEN, G., WARD, W. W. & PRASHER, D. C. 1994. Green fluorescent protein as a marker for gene expression. *Science,* 263**,** 802.

CHAN, C. S., WINSTONE, T. M. L., CHANG, L., STEVENS, C. M., WORKENTINE, M. L., LI, H., WEI, Y., ONDRECHEN, M. J., PAETZEL, M. & TURNER, R. J. 2008. Identification of Residues in DmsD for Twin-Arginine Leader Peptide Binding, Defined through Random and Bioinformatics-Directed Mutagenesis. *Biochemistry,* 47**,** 2749-2759.

CHAN, P., CURTIS, R. A. & WARWICKER, J. 2013. Soluble expression of proteins correlates with a lack of positively-charged surface. *Scientific Reports,* 3**,** 3333.

CHANAL, A., SANTINI, C.-L. & WU, L.-F. 1998. Potential receptor function of three homologous components, TatA, TatB and TatE, of the twin-arginine signal sequence-dependent metalloenzyme translocation pathway in Escherichia coli. *Molecular Microbiology,* 30**,** 674-676.

CHANG, A. C. & COHEN, S. N. 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol,* 134**,** 1141-56.

CHANTALAT, L., JONES, N. D., KORBER, F., NAVAZA, J. & PAVLOZSKY, A. G. 1995. The crystal structure of wild-type growth hormone at 2.5 Å resolution. *Protein Pept. Lett.,* 2**,** 333-340.

CHEN, C., SNEDECOR, B., NISHIHARA, J. C., JOLY, J. C., MCFARLAND, N., ANDERSEN, D. C., BATTERSBY, J. E. & CHAMPION, K. M. 2004. High-level accumulation of a recombinant antibody fragment in the periplasm of Escherichia coli requires a triple-mutant (degP prc spr) host strain. *Biotechnol Bioeng,* 85**,** 463-74.

CHEN, C. C., WALIA, R., MUKHERJEE, K. J., MAHALIK, S. & SUMMERS, D. K. 2015. Indole generates quiescent and metabolically active Escherichia coli cultures. *Biotechnol J,* 10**,** 636-46.

CHEN, H., KIM, J. & KENDALL, D. A. 1996. Competition between functional signal peptides demonstrates variation in affinity for the secretion pathway. *J Bacteriol,* 178**,** 6658-64.

CHEN, J., SONG, J. L., ZHANG, S., WANG, Y., CUI, D. F. & WANG, C. C. 1999. Chaperone activity of DsbC. *J Biol Chem,* 274**,** 19601-5.

CHEN, R. 2012. Bacterial expression systems for recombinant protein production: E. coli and beyond. *Biotechnol Adv,* 30**,** 1102-7.

CHEN, X., ZARO, J. & SHEN, W.-C. 2013. Fusion Protein Linkers: Property, Design and Functionality. *Advanced drug delivery reviews,* 65**,** 1357-1369.

CHEN, Y. C., LI, C. L., HSIAO, Y. Y., DUH, Y. & YUAN, H. S. 2014. Structure and function of TatD exonuclease in DNA repair. *Nucleic Acids Res,* 42**,** 10776-85.

CHENG, C., WU, S., CUI, L., WU, Y., JIANG, T. & HE, B. 2017. A novel Ffu fusion system for secretory expression of heterologous proteins in Escherichia coli. *Microb Cell Fact,* 16**,** 231.

CHESSHYRE, J. A. & HIPKISS, A. R. 1989. Low temperatures stabilize interferon α-2 against proteolysis in Methylophilus methylotrophus and Escherichia coli. *Applied Microbiology and Biotechnology,* 31**,** 158-162.

CHO, S.-H. & COLLET, J.-F. 2013. Many roles of the bacterial envelope reducing pathways. *Antioxidants & redox signaling,* 18**,** 1690-1698.

CHUNG, C. T., NIEMELA, S. L. & 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 U S A,* 86**,** 2172-2175.

CIAMPI, M. S. 2006. Rho-dependent terminators and transcription termination. *Microbiology,* 152**,** 2515-28.

CIANCIOTTO, N. P. & WHITE, R. C. 2017. Expanding Role of Type II Secretion in Bacterial Pathogenesis and Beyond. *Infect Immun,* 85.

CLIFTON, L. A., SKODA, M. W., LE BRUN, A. P., CIESIELSKI, F., KUZMENKO, I., HOLT, S. A. & LAKEY, J. H. 2015. Effect of divalent cation removal on the structure of gram-negative bacterial outer membrane models. *Langmuir,* 31**,** 404-12.

CLINE, K. 2015. Mechanistic Aspects of Folded Protein Transport by the Twin Arginine Translocase (Tat). *J Biol Chem,* 290**,** 16530-8.

CLINE, K., ETTINGER, W. F. & THEG, S. M. 1992. Protein-specific energy requirements for protein transport across or into thylakoid membranes. Two lumenal proteins are transported in the absence of ATP. *Journal of Biological Chemistry,* 267**,** 2688-2696.

COLLET, J. F. & BARDWELL, J. C. 2002. Oxidative protein folding in bacteria. *Mol Microbiol,* 44**,** 1-8.

CORREIA, M. A., OTRELO-CARDOSO, A. R., SCHWUCHOW, V., SIGFRIDSSON CLAUSS, K. G., HAUMANN, M., ROMAO, M. J., LEIMKUHLER, S. & SANTOS-SILVA, T. 2016. The Escherichia coli Periplasmic Aldehyde Oxidoreductase Is an Exceptional Member of the Xanthine Oxidase Family of Molybdoenzymes. *ACS Chem Biol,* 11**,** 2923-2935.

COSTA, S., ALMEIDA, A., CASTRO, A. & DOMINGUES, L. 2014. Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system. *Frontiers in microbiology,* 5**,** 63-63.

COSTA, S. J., ALMEIDA, A., CASTRO, A., DOMINGUES, L. & BESIR, H. 2013a. The novel Fh8 and H fusion partners for soluble protein expression in Escherichia coli: a comparison with the traditional gene fusion technology. *Appl Microbiol Biotechnol,* 97**,** 6779-91.

COSTA, S. J., COELHO, E., FRANCO, L., ALMEIDA, A., CASTRO, A. & DOMINGUES, L. 2013b. The Fh8 tag: a fusion partner for simple and cost-effective protein purification in Escherichia coli. *Protein Expr Purif,* 92**,** 163-70.

COSTERTON, J. W., INGRAM, J. M. & CHENG, K. J. 1974. Structure and function of the cell envelope of gram-negative bacteria. *Bacteriological Reviews,* 38**,** 87-110.

CRISTOBAL, S., DE GIER, J., NIELSEN, H. & VON HEIJNE, G. 1999. Competition between Sec- and TAT-dependent protein translocation in *Escherichia coli*. *EMBO J,* 18**,** 2982-2990.

CRISTÓBAL, S., DE GIER, J. W., NIELSEN, H. & VON HEIJNE, G. 1999. Competition between Sec- and TAT-dependent protein translocation in Escherichia coli. *The EMBO journal,* 18**,** 2982-2990.

CROMMELIN, D. J. A., SINDELAR, R. D. & MEIBOHM, B. 2008. *Pharmaceutical Biotechnology : fundamentals and application*, New York : Informa Healthcare.

CROXEN, M. A., LAW, R. J., SCHOLZ, R., KEENEY, K. M., WLODARSKA, M. & FINLAY, B. B. 2013. Recent Advances in Understanding Enteric Pathogenic <span class="named-content genus-species" id="named-content-1">Escherichia coli</span&gt. *Clinical Microbiology Reviews,* 26**,** 822.

DAEGELEN, P., STUDIER, F. W., LENSKI, R. E., CURE, S. & KIM, J. F. 2009. Tracing Ancestors and Relatives of Escherichia coli B, and the Derivation of B Strains REL606 and BL21(DE3). *Journal of Molecular Biology,* 394**,** 634-643.

DALBEY, R. E. & VON HEIJNE, G. 2002. *In:* ELSEVIER (ed.) *Protein Targeting, Transport, and Translocation.* Academic Press.

DATE, T. 1983. Demonstration by a novel genetic technique that leader peptidase is an essential enzyme of Escherichia coli. *J Bacteriol,* 154**,** 76-83.

DAY, R. N. & DAVIDSON, M. W. 2009. The fluorescent protein palette: tools for cellular imaging. *Chemical Society reviews,* 38**,** 2887-2921.

DE AVILA E SILVA, S. & ECHEVERRIGARAY, S. 2012. Bacterial Promoter Features Description and Their Application on E. coli in silico Prediction and Recognition Approaches. *Bioinformatics.*

DE BOER, H. A., COMSTOCK, L. J. & VASSER, M. 1983. The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proceedings of the National Academy of Sciences of the United States of America,* 80**,** 21-25.

DE GROOT, N. S. & VENTURA, S. 2006. Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates. *Journal of Biotechnology,* 125**,** 110-113.

DE MARCO, A. 2007. Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in Escherichia coli. *Nature Protocols,* 2**,** 2632.

DE MARCO, A. 2009. Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. *Microb Cell Fact,* 8**,** 26.

DE MARCO, A. 2011. Biotechnological applications of recombinant single-domain antibody fragments. *Microbial Cell Factories,* 10**,** 44.

DE MARCO, A. 2015. Recombinant antibody production evolves into multiple options aimed at yielding reagents suitable for application-specific needs. *Microb Cell Fact,* 14**,** 125.

DE OLIVEIRA, J. E., SOARES, C. R. J., PERONI, C. N., GIMBO, E., CAMARGO, I. M. C., MORGANTI, L., BELLINI, M. H., AFFONSO, R., ARKATEN, R. R., BARTOLINI, P. & RIBELA, M. T. C. P. 1999. High-yield purification of biosynthetic human growth hormone secreted in Escherichia coli periplasmic space. *Journal of Chromatography A,* 852**,** 441-450.

DELISA, M. P., SAMUELSON, P., PALMER, T. & GEORGIOU, G. 2002. Genetic analysis of the twin arginine translocator secretion pathway in bacteria. *J Biol Chem,* 277**,** 29825-31.

DELISA, M. P., TULLMAN, D. & GEORGIOU, G. 2003. Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *Proc Natl Acad Sci U S A,* 100**,** 6115-20.

DENONCIN, K., VERTOMMEN, D., PAEK, E. & COLLET, J. F. 2010. The protein-disulfide isomerase DsbC cooperates with SurA and DsbA in the assembly of the essential beta-barrel protein LptD. *J Biol Chem,* 285**,** 29425-33.

DEPUYDT, M., LEONARD, S. E., VERTOMMEN, D., DENONCIN, K., MORSOMME, P., WAHNI, K., MESSENS, J., CARROLL, K. S. & COLLET, J. F. 2009. A periplasmic reducing system protects single cysteine residues from oxidation. *Science,* 326**,** 1109-11.

DERMAN, A. I. & BECKWITH, J. 1991. Escherichia coli alkaline phosphatase fails to acquire disulfide bonds when retained in the cytoplasm. *J Bacteriol,* 173**,** 7719-22.

DERMAN, A. I., PUZISS, J. W., BASSFORD, P. J., JR. & BECKWITH, J. 1993. A signal sequence is not required for protein export in prlA mutants of Escherichia coli. *Embo j,* 12**,** 879-88.

DINH, T. & BERNHARDT, T. G. 2011. Using superfolder green fluorescent protein for periplasmic protein localization studies. *J Bacteriol,* 193**,** 4984-7.

DIPPEL, A. B., OLENGINSKI, G. M., MAURICI, N., LISKOV, M. T., BREWER, S. H. & PHILLIPS-PIRO, C. M. 2016. Probing the effectiveness of spectroscopic reporter unnatural amino acids: a structural study. *Acta crystallographica. Section D, Structural biology,* 72**,** 121-130.

DJOKO, K. Y., CHONG, L. X., WEDD, A. G. & XIAO, Z. 2010. Reaction mechanisms of the multicopper oxidase CueO from Escherichia coli support its functional role as a cuprous oxidase. *J Am Chem Soc,* 132**,** 2005-15.

DUBINI, A. & SARGENT, F. 2003. Assembly of Tat-dependent [NiFe] hydrogenases: identification of precursor-binding accessory proteins. *FEBS Lett,* 549**,** 141-6.

EIMER, E., FROBEL, J., BLUMMEL, A. S. & MULLER, M. 2015. TatE as a Regular Constituent of Bacterial Twin-arginine Protein Translocases. *J Biol Chem,* 290**,** 29281-9.

EL YAAGOUBI, A., KOHIYAMA, M. & RICHARME, G. 1994. Localization of DnaK (chaperone 70) from Escherichia coli in an osmotic-shock-sensitive compartment of the cytoplasm. *J Bacteriol,* 176**,** 7074-8.

ELENA, V. K., MARIYA, V. S. & RIMA, P. E. 1998. Some peculiarities of synthesis of cysteine-containing peptides. *Russian Chemical Reviews,* 67**,** 545.

ELLIS, M., PATEL, P., EDON, M., RAMAGE, W., DICKINSON, R. & HUMPHREYS, D. P. 2017. Development of a high yielding E. coli periplasmic expression system for the production of humanized Fab' fragments. *Biotechnol Prog,* 33**,** 212-220.

ENGLER, C., KANDZIA, R. & MARILLONNET, S. 2008. A one pot, one step, precision cloning method with high throughput capability. *PLoS One,* 3**,** e3647.

ESPOSITO, D. & CHATTERJEE, D. K. 2006. Enhancement of soluble protein expression through the use of fusion tags. *Curr Opin Biotechnol,* 17**,** 353-8.

FAHNERT, B., LILIE, H. & NEUBAUER, P. 2004. Inclusion bodies: formation and utilisation. *Adv Biochem Eng Biotechnol,* 89**,** 93-142.

FEILMEIER, B. J., ISEMINGER, G., SCHROEDER, D., WEBBER, H. & PHILLIPS, G. J. 2000. Green fluorescent protein functions as a reporter for protein localization in Escherichia coli. *J Bacteriol,* 182**,** 4068-76.

FERNANDEZ, L. A. & DE LORENZO, V. 2001. Formation of disulphide bonds during secretion of proteins through the periplasmic-independent type I pathway. *Mol Microbiol,* 40**,** 332-46.

FERRER-MIRALLES, N., DOMINGO-ESPÍN, J., CORCHERO, J. L., VÁZQUEZ, E. & VILLAVERDE, A. 2009. Microbial factories for recombinant pharmaceuticals. *Microbial Cell Factories,* 8**,** 17.

FISHER, A. C., KIM, J. Y., PEREZ-RODRIGUEZ, R., TULLMAN-ERCEK, D., FISH, W. R., HENDERSON, L. A. & DELISA, M. P. 2008. Exploration of twin-arginine translocation for expression and purification of correctly folded proteins in Escherichia coli. *Microb Biotechnol,* 1**,** 403-15.

FLOWER, A. M., DOEBELE, R. C. & SILHAVY, T. J. 1994. PrlA and PrlG suppressors reduce the requirement for signal sequence recognition. *J Bacteriol,* 176**,** 5607-14.

FORZI, L. & SAWERS, R. G. 2007. Maturation of [NiFe]-hydrogenases in Escherichia coli. *Biometals,* 20**,** 565-78.

FOX, B. G. & BLOMMEL, P. G. 2009. Autoinduction of protein expression. *Current protocols in protein science,* Chapter 5**,** Unit-5.23.

FRAIN, K. M., ROBINSON, C. & VAN DIJL, J. M. 2019. Transport of Folded Proteins by the Tat System. *The Protein Journal*.

FREUDL, R. 2018. Signal peptides for recombinant protein secretion in bacterial expression systems. *Microb Cell Fact,* 17**,** 52.

FRIEHS, K. 2004. Plasmid copy number and plasmid stability. *Adv Biochem Eng Biotechnol,* 86**,** 47-82.

FROBEL, J., BLUMMEL, A. S., DREPPER, F., WARSCHEID, B. & MULLER, M. 2019. Surface-exposed domains of TatB involved in the structural and functional assembly of the Tat translocase in Escherichia coli. *J Biol Chem*.

FROBEL, J., ROSE, P., LAUSBERG, F., BLUMMEL, A. S., FREUDL, R. & MULLER, M. 2012. Transmembrane insertion of twin-arginine signal peptides is driven by TatC and regulated by TatB. *Nat Commun,* 3**,** 1311.

GACIARZ, A., KHATRI, N. K., VELEZ-SUBERBIE, M. L., SAARANEN, M. J., UCHIDA, Y., KESHAVARZ-MOORE, E. & RUDDOCK, L. W. 2017. Efficient soluble expression of disulfide bonded proteins in the cytoplasm of Escherichia coli in fed-batch fermentations on chemically defined minimal media. *Microb Cell Fact,* 16**,** 108.

GACIARZ, A. & RUDDOCK, L. W. 2017. Complementarity determining regions and frameworks contribute to the disulfide bond independent folding of intrinsically stable scFv. *PLoS One,* 12**,** e0189964.

GACIARZ, A., VEIJOLA, J., UCHIDA, Y., SAARANEN, M. J., WANG, C., HORKKO, S. & RUDDOCK, L. W. 2016. Systematic screening of soluble expression of antibody fragments in the cytoplasm of E. coli. *Microb Cell Fact,* 15**,** 22.

GARCÍA-FRUITÓS, E. 2010. Inclusion bodies: a new concept. *Microbial cell factories,* 9**,** 80-80.

GARCÍA-FRUITÓS, E., GONZÁLEZ-MONTALBÁN, N., MORELL, M., VERA, A., FERRAZ, R. M., ARÍS, A., VENTURA, S. & VILLAVERDE, A. 2005. Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microbial Cell Factories,* 4**,** 27.

GARDNER, P. P., BARQUIST, L., BATEMAN, A., NAWROCKI, E. P. & WEINBERG, Z. 2011. RNIE: genome-wide prediction of bacterial intrinsic terminators. *Nucleic Acids Research,* 39**,** 5845-5852.

GASSER, B., PRIELHOFER, R., MARX, H., MAURER, M., NOCON, J., STEIGER, M., PUXBAUM, V., SAUER, M. & MATTANOVICH, D. 2013. Pichia pastoris: protein production host and model organism for biomedical research. *Future Microbiology,* 8**,** 191-208.

GASSMANN, M., GRENACHER, B., ROHDE, B. & VOGEL, J. 2009. Quantifying Western blots: pitfalls of densitometry. *Electrophoresis,* 30**,** 1845-55.

GERARD, F. & CLINE, K. 2006. Efficient twin arginine translocation (Tat) pathway transport of a precursor protein covalently anchored to its initial cpTatC binding site. *J Biol Chem,* 281**,** 6130-5.

GLASBRENNER, K. 1986. Child medicine grows up. *JAMA,* 255**,** 443-447.

GOEL, N. & STEPHENS, S. 2010. Certolizumab pegol. *mAbs,* 2**,** 137-147.

GOH, S. & GOOD, L. 2008. Plasmid selection in Escherichia coli using an endogenous essential gene marker. *BMC Biotechnol,* 8**,** 61.

GON, S., GIUDICI-ORTICONI, M.-T., MÉJEAN, V. & IOBBI-NIVOL, C. 2001. Electron Transfer and Binding of the c-Type Cytochrome TorC to the Trimethylamine N-Oxide Reductase in Escherichia coli. *Journal of Biological Chemistry,* 276**,** 11545-11551.

GONZÁLEZ, J. M. & FISHER, S. Z. 2015. Structural analysis of ibuprofen binding to human adipocyte fatty-acid binding protein (FABP4). *Acta crystallographica. Section F, Structural biology communications,* 71**,** 163-170.

GOODMAN, D. B., CHURCH, G. M. & KOSURI, S. 2013. Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. *Science,* 342**,** 475.

GOTTESMAN, S. 1996. Proteases and their targets in Escherichia coli. *Annu Rev Genet,* 30**,** 465-506.

GOURIDIS, G., KARAMANOU, S., GELIS, I., KALODIMOS, C. G. & ECONOMOU, A. 2009. Signal peptides are allosteric activators of the protein translocase. *Nature,* 462**,** 363-7.

GRÄSLUND, S., STRUCTURAL GENOMICS, C., NORDLUND, P., WEIGELT, J., HALLBERG, B. M., BRAY, J., GILEADI, O., KNAPP, S., OPPERMANN, U., ARROWSMITH, C., HUI, R., MING, J., DHE-PAGANON, S., PARK, H.-W., SAVCHENKO, A., YEE, A., EDWARDS, A., ARCHITECTURE ET FONCTION DES MACROMOLÉCULES, B., VINCENTELLI, R., CAMBILLAU, C., BERKELEY STRUCTURAL GENOMICS, C., KIM, R., KIM, S.-H., CHINA STRUCTURAL GENOMICS, C., RAO,

Z., SHI, Y., INTEGRATED CENTER FOR, S., FUNCTION, I., TERWILLIGER, T. C., KIM, C.-Y., HUNG, L.-W., WALDO, G. S., ISRAEL STRUCTURAL PROTEOMICS, C., PELEG, Y., ALBECK, S., UNGER, T., DYM, O., PRILUSKY, J., SUSSMAN, J. L., JOINT CENTER FOR STRUCTURAL, G., STEVENS, R. C., LESLEY, S. A., WILSON, I. A., MIDWEST CENTER FOR STRUCTURAL, G., JOACHIMIAK, A., COLLART, F., DEMENTIEVA, I., DONNELLY, M. I., ESCHENFELDT, W. H., KIM, Y., STOLS, L., WU, R., ZHOU, M., NEW YORK STRUCTURAL GENOMI, X. R. C. F. S. G., BURLEY, S. K., EMTAGE, J. S., SAUDER, J. M., THOMPSON, D., BAIN, K., LUZ, J., GHEYI, T., ZHANG, F., ATWELL, S., ALMO, S. C., BONANNO, J. B., FISER, A., SWAMINATHAN, S., STUDIER, F. W., CHANCE, M. R., SALI, A., NORTHEAST STRUCTURAL GENOMICS, C., ACTON, T. B., XIAO, R., ZHAO, L., MA, L. C., HUNT, J. F., TONG, L., CUNNINGHAM, K., INOUYE, M., ANDERSON, S., JANJUA, H., SHASTRY, R., HO, C. K., WANG, D., WANG, H., JIANG, M., MONTELIONE, G. T., OXFORD PROTEIN PRODUCTION, F., STUART, D. I., OWENS, R. J., DAENKE, S., PROTEIN SAMPLE PRODUCTION FACILITY, M. D. C. F. M. M., SCHÜTZ, A., HEINEMANN, U., INITIATIVE, R. S. G. P., YOKOYAMA, S., COMPLEXES, S., BÜSSOW, K. & GUNSALUS, K. C. 2008. Protein production and purification. *Nature Methods,* 5**,** 135.

GRAY, G. L., BALDRIDGE, J. S., MCKEOWN, K. S., HEYNEKER, H. L. & CHANG, C. N. 1985. Periplasmic production of correctly processed human growth hormone in Escherichia coli: natural and bacterial signal sequences are interchangeable. *Gene,* 39**,** 247-54.

GREEN, E. R. & MECSAS, J. 2016. Bacterial Secretion Systems: An Overview. *Microbiol Spectr,* 4.

GUO, P. C., MA, J. D., JIANG, Y. L., WANG, S. J., BAO, Z. Z., YU, X. J., CHEN, Y. & ZHOU, C. Z. 2012. Structure of yeast sulfhydryl oxidase erv1 reveals electron transfer of the disulfide relay system in the mitochondrial intermembrane space. *J Biol Chem,* 287**,** 34961-9.

GUSTAFSSON, C., GOVINDARAJAN, S. & MINSHULL, J. 2004. Codon bias and heterologous protein expression. *Trends Biotechnol,* 22**,** 346-53.

HAN, L., ENFORS, S.-O. & HÄGGSTRÖM, L. 2003. Escherichia coli high-cell-density culture: carbon mass balances and release of outer membrane components. *Bioprocess and Biosystems Engineering,* 25**,** 205-212.

HARMSEN, M. M. & DE HAARD, H. J. 2007. Properties, production, and applications of camelid single-domain antibody fragments. *Applied Microbiology and Biotechnology,* 77**,** 13-22.

HARRISON, S. T. L. 1991. Bacterial cell disruption: A key unit operation in the recovery of intracellular products. *Biotechnology Advances,* 9**,** 217-240.

HATAHET, F., NGUYEN, V. D., SALO, K. E. & RUDDOCK, L. W. 2010. Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of E. coli. *Microb Cell Fact,* 9**,** 67.

HAYASHI, K., MOROOKA, N., YAMAMOTO, Y., FUJITA, K., ISONO, K., CHOI, S., OHTSUBO, E., BABA, T., WANNER, B. L., MORI, H. & HORIUCHI, T. 2006. Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110. *Mol Syst Biol,* 2**,** 2006 0007.

HE, C. & OHNISHI, K. 2017. Efficient renaturation of inclusion body proteins denatured by SDS. *Biochem Biophys Res Commun,* 490**,** 1250-1253.

HEIDRICH, C., TEMPLIN, M. F., URSINUS, A., MERDANOVIC, M., BERGER, J., SCHWARZ, H., DE PEDRO, M. A. & HOLTJE, J. V. 2001. Involvement of N-acetylmuramyl-L-alanine amidases in cell separation and antibiotic-induced autolysis of Escherichia coli. *Mol Microbiol,* 41**,** 167-78.

HENRY, R., KAPAZOGLOU, A., MCCAFFERY, M. & CLINE, K. 1994. Differences between lumen targeting domains of chloroplast transit peptides determine pathway specificity for thylakoid transport. *Journal of Biological Chemistry,* 269**,** 10189-10192.

HERCUS, T. R., BARRY, E. F., DOTTORE, M., MCCLURE, B. J., WEBB, A. I., LOPEZ, A. F., YOUNG, I. G. & MURPHY, J. M. 2013. High yield production of a soluble human interleukin-3 variant from E. coli with wild-type bioactivity and improved radiolabeling properties. *PLoS One,* 8**,** e74376.

HOFFMAN, C. S. & WRIGHT, A. 1985. Fusions of secreted proteins to alkaline phosphatase: an approach for studying protein secretion. *Proc Natl Acad Sci U S A,* 82**,** 5107-11.

HOU, B., HEIDRICH, E. S., MEHNER-BREITFELD, D. & BRUSER, T. 2018. The TatA component of the twin-arginine translocation system locally weakens the cytoplasmic membrane of Escherichia coli upon protein substrate binding. *J Biol Chem,* 293**,** 7592-7605.

HUA, Y., CUI, S. W., WANG, Q., MINE, Y. & POYSA, V. 2005. Heat induced gelling properties of soy protein isolates prepared from different defatted soybean flours. *Food Research International,* 38**,** 377-385.

HUANG, C., ROSSI, P., SAIO, T. & KALODIMOS, C. G. 2016. Structural basis for the antifolding activity of a molecular chaperone. *Nature,* 537**,** 202-206.

HUANG, C. J., LIN, H. & YANG, X. 2012. Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements. *J Ind Microbiol Biotechnol,* 39**,** 383-99.

HUANG, K. X., BADGER, M., HANEY, K. & EVANS, S. L. 2007. Large scale production of Bacillus thuringiensis PS149B1 insecticidal proteins Cry34Ab1 and Cry35Ab1 from Pseudomonas fluorescens. *Protein Expr Purif,* 53**,** 325-30.

HUANG, Q., ALCOCK, F., KNEUPER, H., DEME, J. C., ROLLAUER, S. E., LEA, S. M., BERKS, B. C. & PALMER, T. 2017. A signal sequence suppressor mutant that stabilizes an assembled state of the twin arginine translocase. *Proc Natl Acad Sci U S A,* 114**,** E1958-E1967.

HUANG, Q. & PALMER, T. 2017. Signal Peptide Hydrophobicity Modulates Interaction with the Twin-Arginine Translocase. *mBio,* 8**,** e00909-17.

HULFORD, A., HAZELL, L., MOULD, R. M. & ROBINSON, C. 1994. Two distinct mechanisms for the translocation of proteins across the thylakoid membrane, one requiring the presence of a stromal protein factor and nucleotide triphosphates. *Journal of Biological Chemistry,* 269**,** 3251-3256.

IEZZI, M. E., POLICASTRO, L., WERBAJH, S., PODHAJCER, O. & CANZIANI, G. A. 2018. Single-Domain Antibodies and the Promise of Modular Targeting in Cancer Imaging and Treatment. *Front Immunol,* 9**,** 273.

ILBERT, M., MEJEAN, V., GIUDICI-ORTICONI, M. T., SAMAMA, J. P. & IOBBI-NIVOL, C. 2003. Involvement of a mate chaperone (TorD) in the maturation pathway of molybdoenzyme TorA. *J Biol Chem,* 278**,** 28787-92.

ILBERT, M., MEJEAN, V. & IOBBI-NIVOL, C. 2004. Functional and structural analysis of members of the TorD family, a large chaperone family dedicated to molybdoproteins. *Microbiology,* 150**,** 935-43.

INOUYE, S., SOBERON, X., FRANCESCHINI, T., NAKAMURA, K., ITAKURA, K. & INOUYE, M. 1982. Role of positive charge on the amino-terminal region of the signal peptide in protein secretion across the membrane. *Proc Natl Acad Sci U S A,* 79**,** 3438-41.

ISMAIL, N. F., HAMDAN, S., MAHADI, N. M., MURAD, A. M., RABU, A., BAKAR, F. D., KLAPPA, P. & ILLIAS, R. M. 2011. A mutant L-asparaginase II signal peptide improves the secretion of recombinant cyclodextrin glucanotransferase and the viability of Escherichia coli. *Biotechnol Lett,* 33**,** 999-1005.

ITAKURA, K., HIROSE, T., CREA, R., RIGGS, A. D., HEYNEKER, H. L., BOLIVAR, F. & BOYER, H. W. 1977. Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin. *Science,* 198**,** 1056-63.

ITO, K., HIROTA Y FAU - AKIYAMA, Y. & AKIYAMA, Y. 1989. Temperature-sensitive sec mutants of Escherichia coli: inhibition of protein export at the permissive temperature. *J. Bacteriol.,* 171**,** 1742-1743.

IZARD, J. W. & KENDALL, D. A. 1994. Signal peptides: exquisitely designed transport promoters. *Mol Microbiol,* 13**,** 765-73.

IZE, B., GERARD, F. & WU, L. F. 2002. In vivo assessment of the Tat signal peptide specificity in Escherichia coli. *Arch Microbiol,* 178**,** 548-53.

IZE, B., PORCELLI, I., LUCCHINI, S., HINTON, J. C., BERKS, B. C. & PALMER, T. 2004. Novel phenotypes of Escherichia coli tat mutants revealed by global gene expression and phenotypic analysis. *J Biol Chem,* 279**,** 47543-54.

IZE, B., STANLEY, N. R., BUCHANAN, G. & PALMER, T. 2003. Role of the Escherichia coli Tat pathway in outer membrane integrity. *Molecular Microbiology,* 48**,** 1183-1193.

JACK, R. L., BUCHANAN, G., DUBINI, A., HATZIXANTHIS, K., PALMER, T. & SARGENT, F. 2004. Coordinating assembly and export of complex bacterial proteins. *Embo j,* 23**,** 3962-72.

JACK, R. L., SARGENT, F., BERKS, B. C., SAWERS, G. & PALMER, T. 2001. Constitutive expression of Escherichia coli tat genes indicates an important role for the twin-arginine translocase during aerobic and anaerobic growth. *J Bacteriol,* 183**,** 1801-4.

JACOB, F., PERRIN, D., SANCHEZ, C. & MONOD, J. 1960. [Operon: a group of genes with the expression coordinated by an operator]. *C R Hebd Seances Acad Sci,* 250**,** 1727-9.

JOHNSON, I. S. 1983. Human insulin from recombinant DNA technology. *Science,* 219**,** 632-7.

JONDA, S., HUBER-WUNDERLICH, M., GLOCKSHUBER, R. & MÖSSNER, E. 1999. Complementation of DsbA deficiency with secreted thioredoxin variants reveals the crucial role of an efficient dithiol oxidant for catalyzed protein folding in the bacterial periplasm. *The EMBO Journal,* 18**,** 3271-3281.

JONES, A. S., AUSTERBERRY, J. I., DAJANI, R., WARWICKER, J., CURTIS, R., DERRICK, J. P. & ROBINSON, C. 2016. Proofreading of substrate structure by the Twin-Arginine Translocase is highly dependent on substrate conformational flexibility but surprisingly tolerant of surface charge and hydrophobicity changes. *Biochim Biophys Acta,* 1863**,** 3116-3124.

JONGBLOED, J. D., MARTIN, U., ANTELMANN, H., HECKER, M., TJALSMA, H., VENEMA, G., BRON, S., VAN DIJL, J. M. & MULLER, J. 2000. TatC is a specificity determinant for protein secretion via the twin-arginine translocation pathway. *J Biol Chem,* 275**,** 41350-7.

JORMAKKA, M., TORNROTH, S., ABRAMSON, J., BYRNE, B. & IWATA, S. 2002. Purification and crystallization of the respiratory complex formate dehydrogenase-N from Escherichia coli. *Acta Crystallographica Section D,* 58**,** 160-162.

JOURLIN, C., SIMON, G., POMMIER, J., CHIPPAUX, M. & MEJEAN, V. 1996. The periplasmic TorT protein is required for trimethylamine N-oxide reductase gene induction in Escherichia coli. *J Bacteriol,* 178**,** 1219-23.

JOZALA, A. F., GERALDES, D. C., TUNDISI, L. L., FEITOSA, V. A., BREYER, C. A., CARDOSO, S. L., MAZZOLA, P. G., OLIVEIRA-NASCIMENTO, L., RANGEL-YAGUI, C. O., MAGALHAES, P. O., OLIVEIRA, M. A. & PESSOA, A., JR. 2016. Biopharmaceuticals from microorganisms: from production to purification. *Braz J Microbiol,* 47 Suppl 1**,** 51-63.

JUERS, D. H., MATTHEWS, B. W. & HUBER, R. E. 2012. LacZ beta-galactosidase: structure and function of an enzyme of historical and molecular biological importance. *Protein Sci,* 21**,** 1792-807.

KANE, J. F. 1995. Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. *Current Opinion in Biotechnology,* 6**,** 494-500.

KANG, D. G., KIM, C. S., SEO, J. H., KIM, I. G., CHOI, S. S., HA, J. H., NAM, S. W., LIM, G. & CHA, H. J. 2012. Coexpression of molecular chaperone enhances activity and export of organophosphorus hydrolase in Escherichia coli. *Biotechnol Prog,* 28**,** 925-30.

KAPOOR, G., SAIGAL, S. & ELONGAVAN, A. 2017. Action and resistance mechanisms of antibiotics: A guide for clinicians. *Journal of anaesthesiology, clinical pharmacology,* 33**,** 300-305.

KARAMYSHEV, A. L., KARAMYSHEVA, Z. N., KAJAVA, A. V., KSENZENKO, V. N. & NESMEYANOVA, M. A. 1998. Processing of Escherichia coli alkaline phosphatase: role of the primary structure of the signal peptide cleavage region. *J Mol Biol,* 277**,** 859-70.

KEMMINK, J., DIJKSTRA, K., MARIANI, M., SCHEEK, R. M., PENKA, E., NILGES, M. & DARBY, N. J. 1999. The structure in solution of the b domain of protein disulfide isomerase. *J Biomol NMR,* 13**,** 357-68.

KENNY, B., HAIGH, R. & HOLLAND, I. B. 1991. Analysis of the haemolysin transport process through the secretion from Escherichia coli of PCM, CAT or beta-galactosidase fused to the Hly C-terminal signal domain. *Mol Microbiol,* 5**,** 2557-68.

KHOW, O. & SUNTRARACHUN, S. 2012. Strategies for production of active eukaryotic proteins in bacterial expression system. *Asian Pacific journal of tropical biomedicine,* 2**,** 159-162.

KHUSHOO, A., PAL, Y., SINGH, B. N. & MUKHERJEE, K. J. 2004. Extracellular expression and single step purification of recombinant Escherichia coli L-asparaginase II. *Protein Expr Purif,* 38**,** 29-36.

KOLLI, R., SOLL, J. & CARRIE, C. 2018. Plant Mitochondrial Inner Membrane Protein Insertion. *Int J Mol Sci,* 19.

KOSOBOKOVA, E. N., SKRYPNIK, K. A. & KOSORUKOV, V. S. 2016. Overview of Fusion Tags for Recombinant Proteins. *Biochemistry (Mosc),* 81**,** 187-200.

KOSTER, W. 1991. Iron(III) hydroxamate transport across the cytoplasmic membrane of Escherichia coli. *Biol Met,* 4**,** 23-32.

KOUIDMI, I., ALVAREZ, L., COLLET, J. F., CAVA, F. & PARADIS-BLEAU, C. 2018. The Chaperone Activities of DsbG and Spy Restore Peptidoglycan Biosynthesis in the elyC Mutant by Preventing Envelope Protein Aggregation. *J Bacteriol,* 200.

KRAMER, R. M., SHENDE, V. R., MOTL, N., PACE, C. N. & SCHOLTZ, J. M. 2012. Toward a molecular understanding of protein solubility: increased negative surface charge correlates with increased solubility. *Biophysical journal,* 102**,** 1907-1915.

KRISHNAPPA, L., MONTEFERRANTE, C. G. & VAN DIJL, J. M. 2012. Degradation of the twin-arginine translocation substrate YwbN by extracytoplasmic proteases of Bacillus subtilis. *Appl Environ Microbiol,* 78**,** 7801-4.

KYTE, J. & DOOLITTLE, R. F. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol,* 157**,** 105-32.

LAMBERT, P. W., MEERS, J. L., BEST, D. J., HARTLEY BRIAN, S., ATKINSON, T. & LILLY MALCOLM, D. 1983. The production of industrial enzymes. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences,* 300**,** 263-282.

LANGE, C., MULLER, S. D., WALTHER, T. H., BURCK, J. & ULRICH, A. S. 2007. Structure analysis of the protein translocating channel TatA in membranes using a multi-construct approach. *Biochim Biophys Acta,* 1768**,** 2627-34.

LANGLEY, R. J., TING, Y. T., CLOW, F., YOUNG, P. G., RADCLIFF, F. J., CHOI, J. M., SEQUEIRA, R. P., HOLTFRETER, S., BAKER, H. & FRASER, J. D. 2017. Staphylococcal enterotoxin-like X (SElX) is a unique superantigen with functional features of two major families of staphylococcal virulence factors. *PLoS Pathog,* 13**,** e1006549.

LAUSBERG, F., FLECKENSTEIN, S., KREUTZENBECK, P., FROBEL, J., ROSE, P., MULLER, M. & FREUDL, R. 2012. Genetic evidence for a tight cooperation of TatB and TatC during productive recognition of twin-arginine (Tat) signal peptides in Escherichia coli. *PLoS One,* 7**,** e39867.

LEDSGAARD, L., KILSTRUP, M., KARATT-VELLATT, A., MCCAFFERTY, J. & LAUSTSEN, A. H. 2018. Basics of Antibody Phage Display Technology. *Toxins (Basel),* 10.

LEE, C., LI, P., INOUYE, H., BRICKMAN, E. R. & BECKWITH, J. 1989. Genetic studies on the inability of beta-galactosidase to be translocated across the Escherichia coli cytoplasmic membrane. *J Bacteriol,* 171**,** 4609-16.

LEE, H. C., PORTNOFF, A. D., ROCCO, M. A. & DELISA, M. P. 2014. An engineered genetic selection for ternary protein complexes inspired by a natural three-component hitchhiker mechanism. *Sci Rep,* 4**,** 7570.

LEE, J., HOFHAUS, G. & LISOWSKY, T. 2000. Erv1p from Saccharomyces cerevisiae is a FAD-linked sulfhydryl oxidase. *FEBS Lett,* 477**,** 62-6.

LEE, P. A., TULLMAN-ERCEK, D. & GEORGIOU, G. 2006. The bacterial twin-arginine translocation pathway. *Annu Rev Microbiol,* 60**,** 373-95.

LETUNIC, I. & BORK, P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res,* 44**,** W242-5.

LI, C. H. 1982. Human growth hormone: 1974–1981. *Molecular and Cellular Biochemistry,* 46**,** 31-41.

LI, J. & ZHANG, Y. 2014. Relationship between promoter sequence and its strength in gene expression. *Eur Phys J E Soft Matter,* 37**,** 44.

LILLY, M. D. & DUNNILL, P. 1969. *Isolation of intracellular enzymes from micro-organisms - the development of a continuous process. In Fermentation Advances,* London, Academic Press.

LINTON, E., WALSH, M. K., SIMS, R. C. & MILLER, C. D. 2012. Translocation of green fluorescent protein by comparative analysis with multiple signal peptides. *Biotechnol J,* 7**,** 667-76.

LOBSTEIN, J., EMRICH, C. A., JEANS, C., FAULKNER, M., RIGGS, P. & BERKMEN, M. 2012. SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories,* 11.

LOW, K. O., MUHAMMAD MAHADI, N. & MD ILLIAS, R. 2013. Optimisation of signal peptide for recombinant protein secretion in bacterial hosts. *Appl Microbiol Biotechnol,* 97**,** 3811-26.

MADEIRA, F., PARK, Y. M., LEE, J., BUSO, N., GUR, T., MADHUSOODANAN, N., BASUTKAR, P., TIVEY, A. R. N., POTTER, S. C., FINN, R. D. & LOPEZ, R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research*.

MAENG, B. H., NAM, D. H. & KIM, Y. H. 2011. Coexpression of molecular chaperones to enhance functional expression of anti-BNP scFv in the cytoplasm of Escherichia coli for the detection of B-type natriuretic peptide. *World J Microbiol Biotechnol,* 27**,** 1391-8.

MAFFEY, L., VEGA, C. G., MINO, S., GARAICOECHEA, L. & PARRENO, V. 2016. Anti-VP6 VHH: An Experimental Treatment for Rotavirus A-Associated Disease. *PLoS One,* 11**,** e0162351.

MAGNET, S., DUBOST, L., MARIE, A., ARTHUR, M. & GUTMANN, L. 2008. Identification of the L,D-transpeptidases for peptidoglycan cross-linking in Escherichia coli. *J Bacteriol,* 190**,** 4782-5.

MALHERBE, G., HUMPHREYS, D. P. & DAVE, E. 2019. A robust fractionation method for protein subcellular localization studies in Escherichia coli. *Biotechniques,* 66**,** 171-178.

MALIK, A. 2016. Protein fusion tags for efficient expression and purification of recombinant proteins in the periplasmic space of E. coli. *3 Biotech,* 6**,** 44-44.

MANOIL, C. & BECKWITH, J. 1985. TnphoA: a transposon probe for protein export signals. *Proc Natl Acad Sci U S A,* 82**,** 8129-33.

MANOIL, C. & BECKWITH, J. 1986. A genetic approach to analyzing membrane protein topology. *Science,* 233**,** 1403-8.

MARKS, J., KANNAN, K., RONCASE, E. J., KLEPACKI, D., KEFI, A., ORELLE, C., VAZQUEZ-LASLOP, N. & MANKIN, A. S. 2016. Context-specific inhibition of translation by ribosomal antibiotics targeting the peptidyl transferase center. *Proc Natl Acad Sci U S A,* 113**,** 12150-12155.

MATIAS, V. R. & BEVERIDGE, T. J. 2005. Cryo-electron microscopy reveals native polymeric cell wall structure in Bacillus subtilis 168 and the existence of a periplasmic space. *Mol Microbiol,* 56**,** 240-51.

MATOS, C. F., BRANSTON, S. D., ALBINIAK, A., DHANOYA, A., FREEDMAN, R. B., KESHAVARZ-MOORE, E. & ROBINSON, C. 2012. High-yield export of a native heterologous protein to the periplasm by the tat translocation pathway in Escherichia coli. *Biotechnol Bioeng,* 109**,** 2533-42.

MATOS, C. F., DI COLA, A. & ROBINSON, C. 2009. TatD is a central component of a Tat translocon-initiated quality control system for exported FeS proteins in Escherichia coli. *EMBO Rep,* 10**,** 474-9.

MATOS, C. F., ROBINSON, C., ALANEN, H. I., PRUS, P., UCHIDA, Y., RUDDOCK, L. W., FREEDMAN, R. B. & KESHAVARZ-MOORE, E. 2014. Efficient export of prefolded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in Escherichia coli CyDisCo strains. *Biotechnol Prog,* 30**,** 281-90.

MATOS, C. F. R. O., ROBINSON, C. & DI COLA, A. 2008. The Tat system proofreads FeS protein substrates and directly initiates the disposal of rejected molecules. *The EMBO journal,* 27**,** 2055-2063.

MCDEVITT, C. A., BUCHANAN, G., SARGENT, F., PALMER, T. & BERKS, B. C. 2006. Subunit composition and in vivo substrate-binding characteristics of Escherichia coli Tat protein complexes expressed at native levels. *Febs j,* 273**,** 5656-68.

MEJEAN, V., IOBBI-NIVOL, C., LEPELLETIER, M., GIORDANO, G., CHIPPAUX, M. & PASCAL, M. C. 1994. TMAO anaerobic respiration in Escherichia coli: involvement of the tor operon. *Mol Microbiol,* 11**,** 1169-79.

MENON, N. K., ROBBINS, J., PECK, H. D., JR., CHATELUS, C. Y., CHOI, E. S. & PRZYBYLA, A. E. 1990. Cloning and sequencing of a putative Escherichia coli [NiFe] hydrogenase-1 operon containing six open reading frames. *J Bacteriol,* 172**,** 1969-77.

MENON, N. K., ROBBINS, J., WENDT, J. C., SHANMUGAM, K. T. & PRZYBYLA, A. E. 1991. Mutational analysis and characterization of the Escherichia coli hya operon, which encodes [NiFe] hydrogenase 1. *J Bacteriol,* 173**,** 4851-61.

MERGULHAO, F. J., SUMMERS, D. K. & MONTEIRO, G. A. 2005. Recombinant protein secretion in Escherichia coli. *Biotechnol Adv,* 23**,** 177-202.

MESSENS, J. & COLLET, J. F. 2006. Pathways of disulfide bond formation in Escherichia coli. *Int J Biochem Cell Biol,* 38**,** 1050-62.

MIAO, X., WANG, Y., WANG, W., LV, X., WANG, M. & YIN, H. 2015. The mAb against adipocyte fatty acid-binding protein 2E4 attenuates the inflammation in the mouse model of high-fat diet-induced obesity via toll-like receptor 4 pathway. *Mol Cell Endocrinol,* 403**,** 1-9.

MICHAELIS, S., INOUYE, H., OLIVER, D. & BECKWITH, J. 1983. Mutations that alter the signal sequence of alkaline phosphatase in Escherichia coli. *J Bacteriol,* 154**,** 366-74.

MILLER, E. 2016. Rapid Amplification of cDNA Ends for RNA Transcript Sequencing in Staphylococcus. *Methods Mol Biol,* 1373**,** 169-83.

MIRAGLIA, L. J., KING, F. J. & DAMOISEAUX, R. 2011. Seeing the light: luminescent reporter gene assays. *Comb Chem High Throughput Screen,* 14**,** 648-57.

MITRA, P., GHOSH, G., HAFEEZUNNISA, M. & SEN, R. 2017. Rho Protein: Roles and Mechanisms. *Annu Rev Microbiol,* 71**,** 687-709.

MOLIK, S., KARNAUCHOV, I., WEIDLICH, C., HERRMANN, R. G. & KLOSGEN, R. B. 2001. The Rieske Fe/S protein of the cytochrome b6/f complex in chloroplasts: missing link in the evolution of protein transport pathways in chloroplasts? *J Biol Chem,* 276**,** 42761-6.

MONTEFERRANTE, C. G., MACKICHAN, C., MARCHADIER, E., PREJEAN, M. V., CARBALLIDO-LOPEZ, R. & VAN DIJL, J. M. 2013. Mapping the twin-arginine protein translocation network of Bacillus subtilis. *Proteomics,* 13**,** 800-11.

MOORKENS, E., MEUWISSEN, N., HUYS, I., DECLERCK, P., VULTO, A. G. & SIMOENS, S. 2017. The Market of Biopharmaceutical Medicines: A Snapshot of a Diverse Industrial Landscape. *Front Pharmacol,* 8**,** 314.

MORA, L., MONCOQ, K., ENGLAND, P., OBERTO, J. & DE ZAMAROCZY, M. 2015. The Stable Interaction Between Signal Peptidase LepB of Escherichia coli and Nuclease Bacteriocins Promotes Toxin Entry into the Cytoplasm. *The Journal of biological chemistry,* 290**,** 30783-30796.

MORI, H. & CLINE, K. 2001. Post-translational protein translocation into thylakoids by the Sec and DeltapH-dependent pathways. *Biochim Biophys Acta,* 1541**,** 80-90.

MORRA, R., DEL CARRATORE, F., MUHAMADALI, H., HORGA, L. G., HALLIWELL, S., GOODACRE, R., BREITLING, R. & DIXON, N. 2018. Translation Stress Positively Regulates MscL-Dependent Excretion of Cytoplasmic Proteins. *MBio,* 9.

MUNCH, R., HILLER, K., GROTE, A., SCHEER, M., KLEIN, J., SCHOBERT, M. & JAHN, D. 2005. Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes. *Bioinformatics,* 21**,** 4187-9.

MUTALIK, V. K., GUIMARAES, J. C., CAMBRAY, G., LAM, C., CHRISTOFFERSEN, M. J., MAI, Q. A., TRAN, A. B., PAULL, M., KEASLING, J. D., ARKIN, A. P. & ENDY, D. 2013. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods,* 10**,** 354-60.

MUYLDERMANS, S. 2001. Single domain camel antibodies: current status. *Reviews in Molecular Biotechnology,* 74**,** 277-302.

MUYLDERMANS, S. 2013. Nanobodies: Natural Single-Domain Antibodies. *Annual Review of Biochemistry,* 82**,** 775-797.

NARHI, L. O., ARAKAWA, T., AOKI, K., WEN, J., ELLIOTT, S., BOONE, T. & CHEETHAM, J. 2001. Asn to Lys mutations at three sites which are N-glycosylated in the mammalian protein decrease the aggregation of Escherichia coli-derived erythropoietin. *Protein Eng,* 14**,** 135-40.

NATALE, P., BRÜSER, T. & DRIESSEN, A. J. M. 2008. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—Distinct translocases and mechanisms. *Biochimica et Biophysica Acta (BBA) - Biomembranes,* 1778**,** 1735-1756.

NAVILLE, M., GHUILLOT-GAUDEFFROY, A., MARCHAIS, A. & GAUTHERET, D. 2011. ARNold: a web tool for the prediction of Rho-independent transcription terminators. *RNA Biol,* 8**,** 11-3.

NEIDHARDT, F. C., CURTISS, R., INGRAHAM, J. L., LIN, E. C. C., LOW, K. B., MAGASANIK, B., REZNIKOFF, W. S., RILEY, M., SCHAECHTER, M. & UMBARGER, H. E. 1996. *Escherichia coli and Salmonella : cellular and molecular biology,* Washington, D.C., ASM Press.

NESMEYANOVA, M. A., KARAMYSHEV, A. L., KARAMYSHEVA, Z. N., KALININ, A. E., KSENZENKO, V. N. & KAJAVA, A. V. 1997. Positively charged lysine at the N-terminus of the signal peptide of the Escherichia coli alkaline phosphatase provides the secretion efficiency and is involved in the interaction with anionic phospholipids. *FEBS Lett,* 403**,** 203-7.

NEUMANN, M., MITTELSTADT, G., IOBBI-NIVOL, C., SAGGU, M., LENDZIAN, F., HILDEBRANDT, P. & LEIMKUHLER, S. 2009. A periplasmic aldehyde oxidoreductase represents the first molybdopterin cytosine dinucleotide cofactor containing molybdo-flavoenzyme from Escherichia coli. *FEBS J,* 276**,** 2762-74.

NGUYEN, V. D., HATAHET, F., SALO, K. E., ENLUND, E., ZHANG, C. & RUDDOCK, L. W. 2011. Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of E.coli. *Microb Cell Fact,* 10**,** 1.

NI, Y. & CHEN, R. 2009. Extracellular recombinant protein production from Escherichia coli. *Biotechnol Lett,* 31**,** 1661-70.

NIELSEN, H. 2017a. Predicting Secretory Proteins with SignalP. *In:* KIHARA, D. (ed.) *Protein Function Prediction: Methods and Protocols.* New York, NY: Springer New York.

NIELSEN, H. 2017b. Predicting Secretory Proteins with SignalP. *Methods Mol Biol,* 1611**,** 59-73.

NIVIERE, V., WONG, S. L. & VOORDOUW, G. 1992. Site-directed mutagenesis of the hydrogenase signal peptide consensus box prevents export of a beta-lactamase fusion protein. *J Gen Microbiol,* 138**,** 2173-83.

NOTHAFT, H. & SZYMANSKI, C. M. 2013. Bacterial protein N-glycosylation: new perspectives and applications. *J Biol Chem,* 288**,** 6912-20.

NYGREN, P. A., STAHL, S. & UHLEN, M. 1994. Engineering proteins to facilitate bioprocessing. *Trends Biotechnol,* 12**,** 184-8.

ORESNIK, I. J., LADNER, C. L. & TURNER, R. J. 2001. Identification of a twin-arginine leader-binding protein. *Molecular Microbiology,* 40**,** 323-331.

ORFANOUDAKI, G. & ECONOMOU, A. 2014. Proteome-wide Subcellular Topologies of E. coli Polypeptides Database (STEPdb). *Molecular & Cellular Proteomics,* 13**,** 3674.

OTRELO-CARDOSO, A. R., DA SILVA CORREIA, M. A., SCHWUCHOW, V., SVERGUN, D. I., ROMÃO, M. J., LEIMKÜHLER, S. & SANTOS-SILVA, T. 2014. Structural data on the periplasmic aldehyde oxidoreductase PaoABC from Escherichia coli: SAXS and preliminary X-ray crystallography analysis. *International journal of molecular sciences,* 15**,** 2223-2236.

OVERTON, T. W. 2014. Recombinant protein production in bacterial hosts. *Drug Discov Today,* 19**,** 590-601.

PALMER, T. & BERKS, B. C. 2012. The twin-arginine translocation (Tat) protein export pathway. *Nat Rev Microbiol,* 10**,** 483-96.

PANAHANDEH, S., MAURER, C., MOSER, M., DELISA, M. P. & MULLER, M. 2008. Following the path of a twin-arginine precursor along the TatABC translocase of Escherichia coli. *J Biol Chem,* 283**,** 33267-75.

PARK, S. J., GEORGIOU, G. & LEE, S. Y. 1999. Secretory production of recombinant protein by a high cell density culture of a protease negative mutant Escherichia coli strain. *Biotechnol Prog,* 15**,** 164-7.

PATEL, R., SMITH, S. M. & ROBINSON, C. 2014. Protein transport by the bacterial Tat pathway. *Biochim Biophys Acta,* 1843**,** 1620-8.

PECHSRICHUANG, P., SONGSIRIRITTHIGUL, C., HALTRICH, D., ROYTRAKUL, S., NAMVIJTR, P., BONAPARTE, N. & YAMABHAI, M. 2016. OmpA signal peptide leads to heterogenous secretion of B. subtilis chitosanase enzyme from E. coli expression system. *Springerplus,* 5**,** 1200.

PECHSRICHUANG, P., YOOHAT, K. & YAMABHAI, M. 2013. Production of recombinant Bacillus subtilis chitosanase, suitable for biosynthesis of chitosan-oligosaccharides. *Bioresour Technol,* 127**,** 407-14.

PECIAK, K., TOMMASI, R., CHOI, J. W., BROCCHINI, S. & LAURINE, E. 2014. Expression of soluble and active interferon consensus in SUMO fusion expression system in E. coli. *Protein Expr Purif,* 99**,** 18-26.

PEDELACQ, J. D., CABANTOUS, S., TRAN, T., TERWILLIGER, T. C. & WALDO, G. S. 2006. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol,* 24**,** 79-88.

PETERS, J. E., THATE, T. E. & CRAIG, N. L. 2003. Definition of the Escherichia coli MC4100 Genome by Use of a DNA Array. *Journal of Bacteriology,* 185**,** 2017-2021.

PETT, W. & LAVROV, D. V. 2013. The twin-arginine subunit C in Oscarella: origin, evolution, and potential functional significance. *Integr Comp Biol,* 53**,** 495-502.

PIERCE, J. J., TURNER, C., KESHAVARZ-MOORE, E. & DUNNILL, P. 1997. Factors determining more efficient large-scale release of a periplasmic enzyme from E. coli using lysozyme. *Journal of Biotechnology,* 58**,** 1-11.

PLOTKIN, J. B. & KUDLA, G. 2011. Synonymous but not the same: the causes and consequences of codon bias. *Nature reviews. Genetics,* 12**,** 32-42.

POMMIER, J., MEJEAN, V., GIORDANO, G. & IOBBI-NIVOL, C. 1998. TorD, a cytoplasmic chaperone that interacts with the unfolded trimethylamine N-oxide reductase enzyme (TorA) in Escherichia coli. *J Biol Chem,* 273**,** 16615-20.

PORRUA, O., BOUDVILLAIN, M. & LIBRI, D. 2016. Transcription Termination: Variations on Common Themes. *Trends Genet,* 32**,** 508-522.

PRASHER, D. C., ECKENRODE, V. K., WARD, W. W., PRENDERGAST, F. G. & CORMIER, M. J. 1992. Primary structure of the Aequorea victoria green-fluorescent protein. *Gene,* 111**,** 229-33.

PRINZ, W. A., ASLUND, F., HOLMGREN, A. & BECKWITH, J. 1997. The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the Escherichia coli cytoplasm. *J Biol Chem,* 272**,** 15661-7.

PUGSLEY, A. 1993. The complete general secretory pathway in Gram-negative bacteria. *Microbio Rev,* 57**,** 50-108.

PUNGINELLI, C., MALDONADO, B., GRAHL, S., JACK, R., ALAMI, M., SCHRÖDER, J., BERKS, B. C. & PALMER, T. 2007. Cysteine Scanning Mutagenesis and Topological Mapping of the *Escherichia coli* Twin-Arginine Translocase TatC Component. *Journal of Bacteriology,* 189**,** 5482-5494.

PUZISS, J. W., HARVEY, R. J. & BASSFORD, P. J., JR. 1992. Alterations in the hydrophilic segment of the maltose-binding protein (MBP) signal peptide that affect either export or translation of MBP. *J Bacteriol,* 174**,** 6488-97.

PYLYPENKO, O., WELZ, T., TITTEL, J., KOLLMAR, M., CHARDON, F., MALHERBE, G., WEISS, S., MICHEL, C. I., SAMOL-WOLF, A., GRASSKAMP, A. T., HUME, A., GOUD, B., BARON, B., ENGLAND, P., TITUS, M. A., SCHWILLE, P., WEIDEMANN, T., HOUDUSSE, A. & KERKHOFF, E. 2016. Coordinated recruitment of Spir actin nucleators and myosin V motors to Rab11 vesicle membranes. *Elife,* 5.

RAMASAMY, S., ABROL, R., SULOWAY, C. J. & CLEMONS, W. M., JR. 2013. The glove-like structure of the conserved membrane protein TatC provides insight into signal sequence recognition in twin-arginine translocation. *Structure,* 21**,** 777-88.

RARAN-KURUSSI, S., KEEFE, K. & WAUGH, D. S. 2015. Positional effects of fusion partners on the yield and solubility of MBP fusion proteins. *Protein Expr Purif,* 110**,** 159-64.

RATTU, M. A., SHAH, N., LEE, J. M., PHAM, A. Q. & MARZELLA, N. 2013. Glucarpidase (voraxaze), a carboxypeptidase enzyme for methotrexate toxicity. *P & T : a peer-reviewed journal for formulary management,* 38**,** 732-744.

RAY, N., NENNINGER, A., MULLINEAUX, C. W. & ROBINSON, C. 2005. Location and mobility of twin arginine translocase subunits in the Escherichia coli plasma membrane. *J Biol Chem,* 280**,** 17961-8.

RAY, N., OATES, J., TURNER, R. J. & ROBINSON, C. 2003. DmsD is required for the biogenesis of DMSO reductase in Escherichia coli but not for the interaction of the DmsA signal peptide with the Tat apparatus. *FEBS Letters,* 534**,** 156-160.

REN, G. H., CAO, L. C., KONG, W., WANG, Z. J. & LIU, Y. H. 2016. Efficient Secretion of the beta-Galactosidase Bgal1-3 via both Tat-Dependent and Tat-Independent Pathways in Bacillus subtilis. *J Agric Food Chem,* 64**,** 5708-16.

RETALLACK, D. M., SCHNEIDER, J. C., MITCHELL, J., CHEW, L. & LIU, H. 2007. Transport of heterologous proteins to the periplasmic space of Pseudomonas fluorescens using a variety of native signal sequences. *Biotechnol Lett,* 29**,** 1483-91.

RICHTER, S. & BRUSER, T. 2005. Targeting of unfolded PhoA to the TAT translocon of Escherichia coli. *J Biol Chem,* 280**,** 42723-30.

RINAS, U. & HOFFMANN, F. 2004. Selective leakage of host-cell proteins during high-cell-density cultivation of recombinant and non-recombinant Escherichia coli. *Biotechnol Prog,* 20**,** 679-87.

RITZ, D., LIM, J., REYNOLDS, C. M., POOLE, L. B. & BECKWITH, J. 2001. Conversion of a peroxiredoxin into a disulfide reductase by a triplet repeat expansion. *Science,* 294**,** 158-60.

RIVAS, L. A., PARRO, V., MORENO-PAZ, M. & MELLADO, R. P. 2000. The Bacillus subtilis 168 csn gene encodes a chitosanase with similar properties to a streptomyces enzyme. *Microbiology,* 146 (Pt 11)**,** 2929-36.

ROBERTS, S. A., WEICHSEL, A., GRASS, G., THAKALI, K., HAZZARD, J. T., TOLLIN, G., RENSING, C. & MONTFORT, W. R. 2002. Crystal structure and electron transfer kinetics of CueO, a multicopper oxidase required for copper homeostasis in Escherichia coli. *Proc Natl Acad Sci U S A,* 99**,** 2766-71.

ROCCO, M. A., WARAHO-ZHMAYEV, D. & DELISA, M. P. 2012. Twin-arginine translocase mutations that suppress folding quality control and permit export of misfolded substrate proteins. *Proceedings of the National Academy of Sciences,* 109**,** 13392.

RODRIGUE, A., CHANAL, A., BECK, K., MULLER, M. & WU, L. F. 1999. Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial tat pathway. *J Biol Chem,* 274**,** 13223-8.

RODRIGUEZ, F., ROUSE, S. L., TAIT, C. E., HARMER, J., DE RISO, A., TIMMEL, C. R., SANSOM, M. S., BERKS, B. C. & SCHNELL, J. R. 2013. Structural model for the protein-translocating element of the twin-arginine transport system. *Proc Natl Acad Sci U S A,* 110**,** E1092-101.

ROIFMAN, C. M., MILLS, G. B., CHU, M. & GELFAND, E. W. 1985. Functional comparison of recombinant interleukin 2 (IL-2) with IL-2-containing preparations derived from cultured cells. *Cell Immunol,* 95**,** 146-56.

ROLLAUER, S. E., TARRY, M. J., GRAHAM, J. E., JAASKELAINEN, M., JAGER, F., JOHNSON, S., KREHENBRINK, M., LIU, S. M., LUKEY, M. J., MARCOUX, J., MCDOWELL, M. A., RODRIGUEZ, F., ROVERSI, P., STANSFELD, P. J., ROBINSON, C. V., SANSOM, M. S., PALMER, T., HOGBOM, M., BERKS, B. C. & LEA, S. M. 2012. Structure of the TatC core of the twin-arginine protein transport system. *Nature,* 492**,** 210-4.

ROSANO, G. L. & CECCARELLI, E. A. 2014. Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol,* 5**,** 172.

ROSE, P., FROBEL, J., GRAUMANN, P. L. & MULLER, M. 2013. Substrate-dependent assembly of the Tat translocase as observed in live Escherichia coli cells. *PLoS One,* 8**,** e69488.

ROTH, R., VAN ZYL, P., TSEKOA, T., STOYCHEV, S., MAMPUTHA, S., BUTHELEZI, S. & CRAMPTON, M. 2017. Co-expression of sulphydryl oxidase and protein disulphide isomerase in Escherichia coli allows for production of soluble CRM197. *J Appl Microbiol,* 122**,** 1402-1411.

RUDNER, D. Z. & LOSICK, R. 2010. Protein subcellular localization in bacteria. *Cold Spring Harb Perspect Biol,* 2**,** a000307.

RUSCH, S. L., CHEN, H., IZARD, J. W. & KENDALL, D. A. 1994. Signal peptide hydrophobicity is finely tailored for function. *J Cell Biochem,* 55**,** 209-17.

RUSSELL, A. B., PETERSON, S. B. & MOUGOUS, J. D. 2014. Type VI secretion system effectors: poisons with a purpose. *Nat Rev Microbiol,* 12**,** 137-48.

SACHELARU, I., PETRIMAN, N. A., KUDVA, R. & KOCH, H. G. 2014. Dynamic interaction of the sec translocon with the chaperone PpiD. *J Biol Chem,* 289**,** 21706-15.

SACK, M., HOFBAUER, A., FISCHER, R. & STOGER, E. 2015. The increasing value of plant-made proteins. *Curr Opin Biotechnol,* 32**,** 163-170.

SALA, A., BORDES, P. & GENEVAUX, P. 2014. Multitasking SecB chaperones in bacteria. *Frontiers in microbiology,* 5**,** 666-666.

SAMALURU, H., SAISREE, L. & REDDY, M. 2007. Role of SufI (FtsP) in cell division of Escherichia coli: evidence for its involvement in stabilizing the assembly of the divisome. *J Bacteriol,* 189**,** 8044-52.

SAMBASIVARAO, D., SCRABA, D. G., TRIEBER, C. & WEINER, J. H. 1990. Organization of dimethyl sulfoxide reductase in the plasma membrane of Escherichia coli. *J Bacteriol,* 172**,** 5938-48.

SAMBROOK, J. & RUSSEL, D. W. 2001. *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press.

SANCHEZ-GARCIA, L., MARTIN, L., MANGUES, R., FERRER-MIRALLES, N., VAZQUEZ, E. & VILLAVERDE, A. 2016. Recombinant pharmaceuticals from microbial cells: a 2015 update. *Microb Cell Fact,* 15**,** 33.

SANTINI, C. L., BERNADAC, A., ZHANG, M., CHANAL, A., IZE, B., BLANCO, C. & WU, L. F. 2001. Translocation of jellyfish green fluorescent protein via the Tat system of Escherichia coli and change of its periplasmic localization in response to osmotic up-shock. *J Biol Chem,* 276**,** 8159-64.

SANTINI, C. L., IZE, B., CHANAL, A., MÜLLER, M., GIORDANO, G. & WU, L. F. 1998. A novel sec-independent periplasmic protein translocation pathway in Escherichia coli. *The EMBO Journal,* 17**,** 101-112.

SARGENT, F., BERKS, B. C. & PALMER, T. 2002. Assembly of membrane-bound respiratory complexes by the Tat protein-transport system. *Arch Microbiol,* 178**,** 77-84.

SARGENT, F., BOGSCH, E., STANLEY, N., WEXLER, M., ROBINSON, C., BERKS, B. C. & PALMER, T. 1998. Overlapping functions of components of a bacterial Sec-independant protein export pathway. *EMBO J,* 17**,** 3640-3650.

SARGENT, F., GOHLKE, U., DE LEEUW, E., STANLEY, N. R., PALMER, T., SAIBIL, H. R. & BERKS, B. C. 2001. Purified components of the Escherichia coli Tat protein transport system form a double-layered ring structure. *Eur J Biochem,* 268**,** 3361-7.

SARGENT, F., STANLEY, N. R., BERKS, B. C. & PALMER, T. 1999. Sec-independent protein translocation in Escherichia coli. A distinct and pivotal role for the TatB protein. *J Biol Chem,* 274**,** 36073-82.

SAWERS, R. G. & BOXER, D. H. 1986. Purification and properties of membrane-bound hydrogenase isoenzyme 1 from anaerobically grown Escherichia coli K12. *Eur J Biochem,* 156**,** 265-75.

SCALES, B. S., DICKSON, R. P., LIPUMA, J. J. & HUFFNAGLE, G. B. 2014. Microbiology, genomics, and clinical significance of the Pseudomonas fluorescens species complex, an unappreciated colonizer of humans. *Clinical microbiology reviews,* 27**,** 927-948.

SCHENBORN, E. & GROSKREUTZ, D. 1999. Reporter gene vectors and assays. *Mol Biotechnol,* 13**,** 29-44.

SHARP, P. M., AVEROF, M., LLOYD, A. T., MATASSI, G. & PEDEN, J. F. 1995. DNA sequence evolution: the sounds of silence. *Philos Trans R Soc Lond B Biol Sci,* 349**,** 241-7.

SHRIVER-LAKE, L. C., GOLDMAN, E. R., ZABETAKIS, D. & ANDERSON, G. P. 2017. Improved production of single domain antibodies with two disulfide bonds by co-expression of chaperone proteins in the Escherichia coli periplasm. *J Immunol Methods,* 443**,** 64-67.

SHU, X., SHANER, N. C., YARBROUGH, C. A., TSIEN, R. Y. & REMINGTON, S. J. 2006. Novel chromophores and buried charges control color in mFruits. *Biochemistry,* 45**,** 9639-47.

SIMONE, D., BAY, D. C., LEACH, T. & TURNER, R. J. 2013. Diversity and evolution of bacterial twin arginine translocase protein, TatC, reveals a protein secretion system that is evolving to fit its environmental niche. *PLoS One,* 8**,** e78742.

SINGH, A., UPADHYAY, V., UPADHYAY, A. K., SINGH, S. M. & PANDA, A. K. 2015. Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. *Microb Cell Fact,* 14**,** 41.

SINGH, P., SHARMA, L., KULOTHUNGAN, S. R., ADKAR, B. V., PRAJAPATI, R. S., ALI, P. S. S., KRISHNAN, B. & VARADARAJAN, R. 2013. Effect of Signal Peptide on Stability and Folding of Escherichia coli Thioredoxin. *PLOS ONE,* 8**,** e63442.

SMITH, M. A., CLEMONS, W. M., JR., DEMARS, C. J. & FLOWER, A. M. 2005. Modeling the effects of prl mutations on the Escherichia coli SecY complex. *J Bacteriol,* 187**,** 6454-65.

SOARES, C. R., GOMIDE, F. I., UEDA, E. K. & BARTOLINI, P. 2003. Periplasmic expression of human growth hormone via plasmid vectors containing the lambdaPL promoter: use of HPLC for product quantification. *Protein Eng,* 16**,** 1131-8.

SOLFOROSI, L., MANCINI, N., CANDUCCI, F., CLEMENTI, N., SAUTTO, G. A., DIOTTI, R. A., CLEMENTI, M. & BURIONI, R. 2012. A phage display vector optimized for the generation of human antibody combinatorial libraries and the molecular cloning of monoclonal antibody fragments. *New Microbiol,* 35**,** 289-94.

SOLOVYEV, V. 2011. V. Solovyev, A Salamov (2011) Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies (Ed. R.W. Li), Nova Science Publishers, p.61-78.

SONG, Y., NIKOLOFF, J. M. & ZHANG, D. 2015. Improving Protein Production on the Level of Regulation of both Expression and Secretion Pathways in Bacillus subtilis. *J Microbiol Biotechnol,* 25**,** 963-77.

SORENSEN, H. P. & MORTENSEN, K. K. 2005. Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microb Cell Fact,* 4**,** 1.

SPIRO, R. G. 2002. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology,* 12**,** 43R-56R.

SPIRO, S. & GUEST, J. R. 1990. FNR and its role in oxygen-regulated gene expression in Escherichia coli. *FEMS Microbiol Rev,* 6**,** 399-428.

STAFFORD, S. J. & LUND, P. A. 2000. Mutagenic studies on human protein disulfide isomerase by complementation of Escherichia coli dsbA and dsbC mutants. *FEBS Letters,* 466**,** 317-322.

STANBURY, P. F. & WHITAKER, A. 1984. Principles of fermentation technology.

STANLEY, N. R., PALMER, T. & BERKS, B. C. 2000. The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in Escherichia coli. *J Biol Chem,* 275**,** 11591-6.

STEELAND, S., VANDENBROUCKE, R. E. & LIBERT, C. 2016. Nanobodies as therapeutics: big opportunities for small antibodies. *Drug Discov Today,* 21**,** 1076-113.

STEWART, E. J., ASLUND, F. & BECKWITH, J. 1998. Disulfide bond formation in the Escherichia coli cytoplasm: an in vivo role reversal for the thioredoxins. *Embo j,* 17**,** 5543-50.

STOLLE, P., HOU, B. & BRUSER, T. 2016. The Tat Substrate CueO Is Transported in an Incomplete Folding State. *J Biol Chem,* 291**,** 13520-8.

STOLT-BERGNER, P., BENDA, C., BERGBREDE, T., BESIR, H., CELIE, P. H. N., CHANG, C., DRECHSEL, D., FISCHER, A., GEERLOF, A., GIABBAI, B., VAN DEN HEUVEL, J., HUBER, G., KNECHT, W., LEHNER, A., LEMAITRE, R., NORDÉN, K., PARDEE, G., RACKE, I., REMANS, K., SANDER, A., SCHOLZ, J., STADNIK, M., STORICI, P., WEINBRUCH, D., ZAROR, I., LUA, L. H. L. & SUPPMANN, S. 2018. Baculovirus-driven protein expression in insect cells: A benchmarking study. *Journal of Structural Biology,* 203**,** 71-80.

STORF, S., PFEIFFER, F., DILKS, K., CHEN, Z. Q., IMAM, S. & POHLSCHRÖDER, M. 2010. Mutational and Bioinformatic Analysis of Haloarchaeal Lipobox-Containing Proteins. *Archaea,* 2010**,** 11.

STUDIER, F. W. 2014. Stable expression clones and auto-induction for protein production in E. coli. *Methods Mol Biol,* 1091**,** 17-32.

SU, L., CHEN, S.-S., YANG, K.-G., LIU, C.-Z., ZHANG, Y.-L. & LIANG, Z.-Q. 2006. High-level expression of human stem cell factor fused with erythropoietin mimetic peptide in Escherichia coli. *Protein Expression and Purification,* 47**,** 477-482.

SUOMINEN, I., MEYER, P., TILGMANN, C., GLUMOFF, T., GLUMOFF, V., KAPYLA, J. & MANTSALA, P. 1995. Effects of signal peptide mutations on processing of Bacillus stearothermophilus alpha-amylase in Escherichia coli. *Microbiology,* 141 (Pt 3)**,** 649-54.

SUTHERLAN, M. L. & WONG, W. K. R. 1993. *Excretion of heterologous porteins from E. coli*. United States patent application US19890395797 19890818.

SUTHERLAND, G. A., GRAYSON, K. J., ADAMS, N. B. P., MERMANS, D. M. J., JONES, A. S., ROBERTSON, A. J., AUMAN, D. B., BRINDLEY, A. A., STERPONE, F., TUFFERY, P., DERREUMAUX, P., DUTTON, P. L., ROBINSON, C., HITCHCOCK, A. & HUNTER, C. N. 2018. Probing the quality control mechanism of the Escherichia coli twin-arginine translocase with folding variants of a de novo-designed heme protein. *J Biol Chem,* 293**,** 6672-6681.

SZELIOVA, D., KRAHULEC, J., SAFRANEK, M., LISKOVA, V. & TURNA, J. 2016. Modulation of heterologous expression from PBAD promoter in Escherichia coli production strains. *J Biotechnol,* 236**,** 1-9.

TAN, S. 2001. A modular polycistronic expression system for overexpressing protein complexes in Escherichia coli. *Protein Expr Purif,* 21**,** 224-34.

TAO, K. 2008. Subcellular localization and in vivo oxidation-reduction kinetics of thiol peroxidase in Escherichia coli. *FEMS Microbiol Lett,* 289**,** 41-5.

TARRY, M. J., SCHÄFER, E., CHEN, S., BUCHANAN, G., GREENE, N. P., LEA, S. M., PALMER, T., SAIBIL, H. R. & BERKS, B. C. 2009. Structural analysis of substrate binding by the TatBC component of the twin-arginine protein transport system. *Proceedings of the National Academy of Sciences,* 106**,** 13284-13289.

TATUM, E. L. & LEDERBERG, J. 1947. Gene Recombination in the Bacterium Escherichia coli. *Journal of bacteriology,* 53**,** 673-684.

TERPE, K. 2003. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol,* 60**,** 523-33.

TERPE, K. 2006. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology,* 72**,** 211.

THEIN, M., SAUER, G., PARAMASIVAM, N., GRIN, I. & LINKE, D. 2010. Efficient Subfractionation of Gram-Negative Bacteria for Proteomics Studies. *Journal of Proteome Research,* 9**,** 6135-6147.

THOMAS, J. D., DANIEL, R. A., ERRINGTON, J. & ROBINSON, C. 2001. Export of active green fluorescent protein to the periplasm by the twin-arginine translocase (Tat) pathway in Escherichia coli. *Mol Microbiol,* 39**,** 47-53.

THOMAS, S., HOLLAND, I. B. & SCHMITT, L. 2014. The Type 1 secretion pathway - the hemolysin system and beyond. *Biochim Biophys Acta,* 1843**,** 1629-41.

THOMSON, N. M., SHIRAI, T., CHIAPELLO, M., KONDO, A., MUKHERJEE, K. J., SIVANIAH, E., NUMATA, K. & SUMMERS, D. K. 2018. Efficient 3-Hydroxybutyrate Production by Quiescent Escherichia coli Microbial Cell Factories is Facilitated by Indole-Induced Proteomic and Metabolomic Changes. *Biotechnol J,* 13**,** e1700571.

THORNE, N., INGLESE, J. & AULD, D. S. 2010. Illuminating Insights into Firefly Luciferase and Other Bioluminescent Reporters Used in Chemical Biology. *Chemistry & Biology,* 17**,** 646-657.

TONG, L., LIN, Q., WONG, W. K. R., ALI, A., LIM, D., SUNG, W. L., HEW, C. L. & YANG, D. S. C. 2000. Extracellular Expression, Purification, and Characterization of a Winter Flounder Antifreeze Polypeptide from Escherichia coli. *Protein Expression and Purification,* 18**,** 175-181.

TROUP, B., JAHN, M., HUNGERER, C. & JAHN, D. 1994. Isolation of the hemF operon containing the gene for the Escherichia coli aerobic coproporphyrinogen III oxidase by in vivo complementation of a yeast HEM13 mutant. *J Bacteriol,* 176**,** 673-80.

TSIEN, R. Y. 1998. The green fluorescent protein. *Annu Rev Biochem,* 67**,** 509-44.

TSIRIGOTAKI, A., DE GEYTER, J., SOSTARIC, N., ECONOMOU, A. & KARAMANOU, S. 2017. Protein export through the bacterial Sec pathway. *Nat Rev Microbiol,* 15**,** 21-36.

TSUKAZAKI, T., MORI, H., ECHIZEN, Y., ISHITANI, R., FUKAI, S., TANAKA, T., PEREDERINA, A., VASSYLYEV, D. G., KOHNO, T., MATURANA, A. D., ITO, K. & NUREKI, O. 2011. Structure and

function of a membrane component SecDF that enhances protein export. *Nature,* 474**,** 235-8.

TULLMAN-ERCEK, D., DELISA, M. P., KAWARASAKI, Y., IRANPOUR, P., RIBNICKY, B., PALMER, T. & GEORGIOU, G. 2007. Export pathway selectivity of Escherichia coli twin arginine translocation signal peptides. *J Biol Chem,* 282**,** 8309-16.

UKKONEN, K., NEUBAUER, A., PEREIRA, V. J. & VASALA, A. 2017. High Yield of Recombinant Protein in Shaken E. coli Cultures with Enzymatic Glucose Release Medium EnPresso B. *Methods Mol Biol,* 1586**,** 127-137.

ULFIG, A. & FREUDL, R. 2018. The early mature part of bacterial twin-arginine translocation (Tat) precursor proteins contributes to TatBC receptor binding. *J Biol Chem,* 293**,** 7281-7299.

VALDERRAMA-RINCON, J. D., FISHER, A. C., MERRITT, J. H., FAN, Y. Y., READING, C. A., CHHIBA, K., HEISS, C., AZADI, P., AEBI, M. & DELISA, M. P. 2012. An engineered eukaryotic protein glycosylation pathway in Escherichia coli. *Nat Chem Biol,* 8**,** 434-6.

VAN DER WOLK, J. P., FEKKES, P., BOORSMA, A., HUIE, J. L., SILHAVY, T. J. & DRIESSEN, A. J. 1998. PrlA4 prevents the rejection of signal sequence defective preproteins by stabilizing the SecA-SecY interaction during the initiation of translocation. *Embo j,* 17**,** 3631-9.

VAN ULSEN, P., RAHMAN, S., JONG, W. S., DALEKE-SCHERMERHORN, M. H. & LUIRINK, J. 2014. Type V secretion: from biogenesis to biotechnology. *Biochim Biophys Acta,* 1843**,** 1592-611.

VARKI, A. 2017. Biological roles of glycans. *Glycobiology,* 27**,** 3-49.

VON HEIJNE, G. 1984. How signal sequences maintain cleavage specificity. *J Mol Biol,* 173**,** 243-51.

WACKER, M., LINTON, D., HITCHEN, P. G., NITA-LAZAR, M., HASLAM, S. M., NORTH, S. J., PANICO, M., MORRIS, H. R., DELL, A., WREN, B. W. & AEBI, M. 2002. N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science,* 298**,** 1790-3.

WALSH, S. T., JEVITTS, L. M., SYLVESTER, J. E. & KOSSIAKOFF, A. A. 2003. Site2 binding energetics of the regulatory step of growth hormone-induced receptor homodimerization. *Protein Sci,* 12**,** 1960-70.

WANG, J., MUKHTAR, H., MA, L., PANG, Q. & WANG, X. 2018. VHH Antibodies: Reagents for Mycotoxin Detection in Food Products. *Sensors (Basel),* 18.

WANG, X. & LAVROV, D. V. 2007. Mitochondrial genome of the homoscleromorph Oscarella carmela (Porifera, Demospongiae) reveals unexpected complexity in the common ancestor of sponges and other animals. *Mol Biol Evol,* 24**,** 363-73.

WARAHO-ZHMAYEV, D., GKOGKA, L., YU, T. Y. & DELISA, M. P. 2013. A microbial sensor for discovering structural probes of protein misfolding and aggregation. *Prion,* 7**,** 151-6.

WEINER, J. H., BILOUS, P. T., SHAW, G. M., LUBITZ, S. P., FROST, L., THOMAS, G. H., COLE, J. A. & TURNER, R. J. 1998. A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins. *Cell,* 93**,** 93-101.

WESOLOWSKI, J., ALZOGARAY, V., REYELT, J., UNGER, M., JUAREZ, K., URRUTIA, M., CAUERHFF, A., DANQUAH, W., RISSIEK, B., SCHEUPLEIN, F., SCHWARZ, N., ADRIOUCH, S., BOYER, O., SEMAN, M., LICEA, A., SERREZE, D. V., GOLDBAUM, F. A., HAAG, F. & KOCH-NOLTE, F. 2009. Single domain antibodies: promising experimental and therapeutic tools in infection and immunity. *Medical Microbiology and Immunology,* 198**,** 157-174.

WEXLER, M., SARGENT, F., JACK, R. L., STANLEY, N. R., BOGSCH, E. G., ROBINSON, C., BERKS, B. C. & PALMER, T. 2000. TatD is a cytoplasmic protein with DNase activity. No requirement for TatD family proteins in sec-independent protein export. *J Biol Chem,* 275**,** 16717-22.

WU, H., BENNETT, G. N. & SAN, K. Y. 2015. Metabolic control of respiratory levels in coenzyme Q biosynthesis-deficient Escherichia coli strains leading to fine-tune aerobic lactate fermentation. *Biotechnol Bioeng,* 112**,** 1720-6.

WU, L. F., IZE, B., CHANAL, A., QUENTIN, Y. & FICHANT, G. 2000. Bacterial twin-arginine signal peptide-dependent protein translocation pathway: evolution and mechanism. *J Mol Microbiol Biotechnol,* 2**,** 179-89.

XIA, Y., ZHAO, J., CHEN, H., LIU, X., WANG, Y., TIAN, F., ZHANG, H. P., ZHANG, H. & CHEN, W. 2010. Extracellular secretion in Bacillus subtilis of a cytoplasmic thermostable beta-galactosidase from Geobacillus stearothermophilus. *J Dairy Sci,* 93**,** 2838-45.

YAHR, T. L. & WICKNER, W. T. 2001. Functional reconstitution of bacterial Tat translocation in vitro. *The EMBO Journal,* 20**,** 2472-2479.

YAMAMOTO, Y., RITZ, D., PLANSON, A. G., JONSSON, T. J., FAULKNER, M. J., BOYD, D., BECKWITH, J. & POOLE, L. B. 2008. Mutant AhpC peroxiredoxins suppress thiol-disulfide redox deficiencies and acquire deglutathionylating activity. *Mol Cell,* 29**,** 36-45.

YARRANTON, G. T. & BEBBINGTON , C. R. 2005. *SECRETION OF ANTIBODIES WITHOUT SIGNAL PEPTIDES FROM BACTERIA*. United States patent application 13/506, 653.

YAU, K. Y., GROVES, M. A., LI, S., SHEEDY, C., LEE, H., TANHA, J., MACKENZIE, C. R., JERMUTUS, L. & HALL, J. C. 2003. Selection of hapten-specific single-domain antibodies from a non-immunized llama ribosome display library. *J Immunol Methods,* 281**,** 161-75.

YEN, M. R., TSENG, Y. H., NGUYEN, E. H., WU, L. F. & SAIER, M. H., JR. 2002. Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system. *Arch Microbiol,* 177**,** 441-50.

YOKOYAMA, K. & LEIMKUHLER, S. 2015. The role of FeS clusters for molybdenum cofactor biosynthesis and molybdoenzymes in bacteria. *Biochim Biophys Acta,* 1853**,** 1335-49.

YOUNGMAN, K. M., SPENCER, D. B., BREMS, D. N. & DEFELIPPIS, M. R. 1995. Kinetic Analysis of the Folding of Human Growth Hormone: INFLUENCE OF DISULFIDE BONDS. *Journal of Biological Chemistry,* 270**,** 19816-19822.

YU, F., YAMADA, H., DAISHIMA, K. & MIZUSHIMA, S. 1984. Nucleotide sequence of the lspA gene, the structural gene for lipoprotein signal peptidase of Escherichia coli. *FEBS Lett,* 173**,** 264-8.

YU, N. Y., WAGNER, J. R., LAIRD, M. R., MELLI, G., REY, S., LO, R., DAO, P., SAHINALP, S. C., ESTER, M., FOSTER, L. J. & BRINKMAN, F. S. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics,* 26**,** 1608-15.

ZAFAR, A., AFTAB, M. N., UD DIN, Z., AFTAB, S., IQBAL, I. & UL HAQ, I. 2016. Cloning, Purification and Characterization of a Highly Thermostable Amylase Gene of Thermotoga petrophila into Escherichia coli. *Appl Biochem Biotechnol,* 178**,** 831-48.

ZHANG, Y., HU, Y., LI, H. & JIN, C. 2014a. Structural basis for TatA oligomerization: an NMR study of Escherichia coli TatA dimeric structure. *PLoS One,* 9**,** e103157.

ZHANG, Y., WANG, L., HU, Y. & JIN, C. 2014b. Solution structure of the TatB component of the twin-arginine translocation system. *Biochim Biophys Acta,* 1838**,** 1881-8.

ZHANG, Z., TANG, R., ZHU, D., WANG, W., YI, L. & MA, L. 2017. Non-peptide guided auto-secretion of recombinant proteins by super-folder green fluorescent protein in Escherichia coli. *Sci Rep,* 7**,** 6990.

ZHENG, M., ASLUND, F. & STORZ, G. 1998. Activation of the OxyR transcription factor by reversible disulfide bond formation. *Science,* 279**,** 1718-21.

ZHOU, Z., DANG, Y., ZHOU, M., LI, L., YU, C.-H., FU, J., CHEN, S. & LIU, Y. 2016. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proceedings of the National Academy of Sciences,* 113**,** E6117.

ZISMANN, V. & NOURBAKHSH, M. 2014. Rapid mapping of RNA 3' and 5' ends. *Methods Mol Biol,* 1182**,** 19-25.

ZOUED, A., BRUNET, Y. R., DURAND, E., ASCHTGEN, M.-S., LOGGER, L., DOUZI, B., JOURNET, L., CAMBILLAU, C. & CASCALES, E. 2014. Architecture and assembly of the Type VI secretion system. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* 1843**,** 1664-1673.

APPENDICES

Appendix 1: List of NTS from *E. coli*

List of 28 natural Tat substrates from Brown et al. (1995) and Tullman-Ercek et al. (2007) supplemented with information from UniProt (www.uniprot.org) using only reviewed materials (Swiss-Prot). The table indicates the short name of the protein, the molecular size in kDa, the number of cysteine (#Cys), the number of disulphide bond (#DSB), the protein's name, the cofactor(s), the oligomeric state in which the said protein is transported via the Tat pathway, the required chaperone and the chromosomal minutes. The listed proteins are sorted by molecular size. The selected candidates are highlighted in green. N/D non-determined.

Name	Size (kDa)	#Cys	#DSB	Protein	Cofactor	Oligomeric state during Tat export	Chaperone	Min
YcbK	20.4	0	0	Putative protein	N/D	N/D	N/D	21.17
PaoA	24.3	9	0	Xanthine dehydrogenase	[2Fe-2S]$_2$	Heterotrimer with PaoB and PaoC	PaoD	6.49
NrfC	24.6	16	0	Iron-sulfur protein	[2Fe-2S]$_4$	N/D	N/D	92.41
NapG	24.9	16	0	Ferredoxin	[2Fe-2S]$_4$	Monomer	N/D	49.57
YdhX	25.1	16	N/D	Putative ferredoxin-like protein	[2Fe-2S]$_4$	N/D	N/D	37.72
YfhG	27.3	4	N/D	Putative protein	N/D	N/D	N/D	57.85
AmiA	31.4	0	0	N-acetylmuramoyl-L-alanine amidase A	Zn	Monomer	N/D	55.06
FhuD	33.0	1	0	Iron(3+)-hydroxamate-binding protein	N/D	N/D	N/D	3.68
HybA	36.0	17	0	Hydrogenase 2	[2Fe-2S]$_4$	Monomer	N/D	67.65
MsrP	37.4	1	0	Methionine sulphoxide reductase	Molybdenum	Heterodimer with MsrQ	N/D	43.95
HybO	39.7	13	0	Hydrogenase 2	[2Fe-2S]$_3$	Heterodimer with HybC	HybE	67.67
HyaA	40.7	14	0		[2Fe-2S]$_3$	Heterodimer with HyaB	HyaE	22.23
AmiC	45.6	0	0	N-acetylmuramoyl-L-alanine amidase C	Zn	Monomer	N/D	63.43
EfeB	46.8	4	0	Peroxidase	Fe, heme	Homodimer	N/D	23.34
YahJ	50.6	4	N/D	Putative protein	N/D	N/D	N/D	7.34
WcaM	51.3	6	0	Colanic acid biosynthesis protein	N/D	N/D	N/D	45.57
FtsP	51.8	2	0	Cell division protein	Cu	Monomer	N/D	68.02
CueO	56.6	3	0	Copper oxidase	4 Cu	Monomer	N/D	2.96
MdoD	62.8	8	N/D	Glucans biosynthesis protein	N/D	N/D	N/D	32.27
PcoA	67.3	1	0	Copper resistance protein A	4 Cu	Monomer	N/D	Pls.
TorZ	89.0	5	0	Trimethylamine-N-oxide reductase 2	Molybdenum	Monomer	TorY	42.13
YnfE	89.8	12	0	Dimethyl sulfoxide reductase	Molybdenum, [2Fe-2S]	N/D	N/D	35.75
YnfF	90.0	13	0	Putative dimethyl sulfoxide reductase	Molybdenum, [2Fe-2S]	heterotrimer with YnfG and YnfH	N/D	35.80
DmsA	90.4	15	0	Dimethyl sulfoxide reductase	Molybdenum, [2Fe-2S]	Monomer	N/D	20.29
NapA	93.0	12	0	Nitrate reductase	Molybdenum, [2Fe-2S]	Heterodimer with NapB	N/D	49.60
TorA	94.5	10	0	Trimethylamine-N-oxide reductase 1	Molybdenum	Monomer	TorD	22.83
FdoG	112.5	14	0	Formate dehydrogenase-O	Molybdenum, [2Fe-2S]	Heterodimer with FdoH	FdhD	76.45
FdnG	113.0	13	0	Formate dehydrogenase	Molybdenum, [2Fe-2S]	Heterodimer with FdnH	FdhD	33.37

Appendix 2: Signal peptides alignment

The magenta, green, cyan and orange residues represent the hydrophobic, polar, negatively charged and positively charged residues respectively. The sections highlighted in grey cover the twin-arginine Tat motif. The last three residues of each sequence are the signal peptidase cleavage motif (V/A-x-A). The last column corresponds to the net charges estimated at pH = 7.0.

Protein	Origin	Signal peptide sequence	Length	Charge
OmpA	*E. coli*	----------M--------KKTAI-AIAVALAG-F--ATV------AQA----------	21	1.9
AmiA	*E. coli*	---MSTF-KPL--------KTLTS-RRQVLKAG-L--AALTL-SGMSQAIA----------	34	4.9
HyaA	*E. coli*	MNNEETFYQAM--------RRQGVTRRSFLKYCSLAATSLGLGAGMAPKIAWA----------	45	3.9
PaoA	*E. coli*	MSNQGEYPEDNRV-GKHEPHDLSLTRRDLIKVSAATA---------ATAVVYPHSTLAASVPA	50	-0.4
TorA	*E. coli*	MNNNDLF-QAS--------------RRRFLAQL----GGLTVAGMLGPSLLTPRRATA-----	39	3.9
YcbK	*E. coli*	---MDKF-DAN--------------RRKLL-AL----GGVALGAAILPTPA---FA-------	30	1.9
PhoD	*B. subtilis*	MAYDSRFDEWVQKLKEESFQNNTFDRRKFIQGA-GKIAGLSLGLTIAQSVGAFEVN------A	55	-0.1

Appendix 3: Genomic comparisons between *E. coli* strains

List of missing genes compared to the W3110 WT reference genome (AP009048.1, NCBI). *data obtained previously (Peters et al., 2003). Names in parentheses correspond to the previous gene names used in the publication. † putative gene discovered after the publication. ‡ partially deleted gene.

Locus	W3110 WT	MC4100 WT	MC4100 WT*	MC4100 Tat-null
A		fimB‡	fimB‡	fimB‡
A		fimE‡	fimE‡	fimE‡
B		fruB	fruB	fruB
B		fruK‡	fruK‡	fruK‡
B		setB	setB (yeiO)	setB
B		yeiP	yeiP	yeiP
B		yeiP	yeiQ	yeiQ
B		yeiR	yeiR	yeiR
B		yeiU‡	yeiU (b2174)‡	yeiU‡
B		yeiW†		yeiW†
C				TatA
C				TatB
C				TatC
C				TatD
D		icd‡	icd‡	icd‡
D		icdC‡	icdC‡	icdC‡
D		intE	intE (b1140)	intE
D		lit	lit	lit
D		mcrA	mcrA	mcrA
D		pin	pin	pin
D		stfE	stfE (b1157)	stfE
D		tfaE	tfaE (ycfA)	tfaE
D		ycfK	ycfK	ycfK
D		ymfD	ymfD (b1137)	ymfD
D		ymfE	ymfE (b1138)	ymfE
D		ymfG	ymfG (b1141)	ymfG
D		ymfH	ymfH (b1142)	ymfH
D		ymfI	ymfI (b1143)	ymfI
D		ymfJ	ymfJ (b1144)	ymfJ
D		ymfK	ymfK (b1145)	ymfK
D		ymfL	ymfL (b1147)	ymfL
D		ymfM	ymfM (b1148)	ymfM
D		ymfN	ymfN (b1149)	ymfN
D		ymfO	ymfO (b1151)	ymfO
D		ymfP	ymfP (b1152)	ymfP
D		ymfQ	ymfQ (b1153)	ymfQ
D		ymfR	ymfR (b1150)	ymfR
D		ymfS	ymfS (b1155)	ymfS
D		ymfT	ymfT (b1146)	ymfT
E				tatE

Locus	W3110 WT	MC4100 WT	MC4100 WT*	MC4100 Tat-null
F		afuB	afuB (b0263)	afuB
F		afuC	afuC (yagC)	afuC
F		argF	argF	argF
F		b0309	b0309	b0309
F		b0332	b0332	b0332
F		betA	betA	betA
F		betB	betB	betB
F		betI	betI	betI
F		betT	betT	betT
F		codA	codA	codA
F		codB	codB	codB
F		cynR	cynR	cynR
F		cynS	cynS	cynS
F		cynT	cynT	cynT
F		cynX	cynX	cynX
F		eaeH	eaeH	eaeH
F		intF	intF (b0281)	intF
F		lacA	lacA	lacA
F		lacI	lacI	lacI
F		lacY	lacY	lacY
F		lacZ	lacZ	lacZ
F		mhpA	mhpA (b0347)	mhpA
F		mhpB	mhpB	mhpB
F		mhpC	mhpC (b0349)	mhpC
F		mhpD‡	mhpD (b0350)‡	mhpD‡
F		mhpR	mhpR (b0346)	mhpR
F		mmuM	mmuM (yagD)	mmuM
F		mmuP‡	mmuP (ykfD)‡	mmuP‡
F		prpB	prpB (b0331)	prpB
F		prpC	prpC (b0333)	prpC
F		prpD	prpD (b0334)	prpD
F		prpE	prpE (b0335)	prpE
F		prpR	prpR (b0330)	prpR
F		yagA	yagA	yagA
F		yagB	yagB	yagB
F		yagE	yagE	yagE
F		yagF	yagF	yagF
F		yagG	yagG	yagG
F		yagH	yagH (b0271)	yagH
F		yagI	yagI	yagI
F		yagJ	yagJ	yagJ
F		yagK	yagK	yagK
F		yagL	yagL	yagL
F		yagM	yagM (b0279)	yagM

Locus	W3110 WT	MC4100 WT	MC4100 WT*	MC4100 Tat-null
F		yagN	yagN (b0280)	yagN
F		yagP	yagP	yagP
F		paoD	paoD (yagQ)	paoD
F		paoC	paoC (yagR)	paoC
F		paoB	paoB (yagS)	paoB
F		paoA	paoA (yagT)	paoA
F		yagU	yagU	yagU
F		yagV	yagv	yagV
F		yagW	yagW	yagW
F		yagX	yagX	yagX
F		yagY	yagY	yagY
F		yagZ	yagZ	yagZ
F		yahA	yahA	yahA
F		yahB	yahB (b0316)	yahB
F		yahC	yahC (b0317)	yahC
F		yahD	yahD (b0318)	yahD
F		yahE	yahE (b0319)	yahE
F		yahF	yahF (b0320)	yahF
F		yahG	yahG (b0321)	yahG
F		yahH	yahH(b0322)	yahH
F		yahI	yahI(b0323)	yahI
F		yahJ	yahJ (b0324)	yahJ
F		yahK	yahK (b0325)	yahK
F		yahL	yahL (b0326)	yahL
F		yahM	yahM (b0327)	yahM
F		yahN	yahN (b0328)	yahN
F		yahO	yahO (b0329)	yahO
F		ykgA	ykgA	ykgA
F		ykgB	ykgB	ykgB
F		ykgC	ykgC	ykgC
F		ykgD	ykgD	ykgD
F		ykgE	ykgE	ykgE
F		ykgF	ykgF (b030)	ykgF
F		ykgG	ykgG	ykgG
F		ykgH	ykgH	ykgH
F		ykgI	ykgI (b0303)	ykgI
F		ykgJ	ykgJ (b0288)	ykgJ
F		ykgK	ykgK (b0294)	ykgK
F		ykgL	ykgL (b0295)	ykgL
F		ykgM	ykgM (b0296)	ykgM
F		ykgN†		ykgN†
G		fnr		
G		uspE		
G		ynaJ		

Locus	W3110 WT	MC4100 WT	MC4100 WT*	MC4100 Tat-null
Count	0	123	118	125

Appendix 4: PureFrac detailed protocol

Keep samples and buffers on ice throughout the whole protocol and all centrifugation steps are carried out at 4°C. Use a single channel pipette with 1 ml tips to remove or transfer small liquid volumes without disturbing the pellet. All resuspension steps are performed by pipetting up and down gently until all of the pellet is resuspended without creating air bubbles.

- At harvest, transfer a volume of $\frac{8}{OD_{600}}$ of the culture to a 50 ml tube.
- Centrifuge for 15 min at 4816 g.
- Optional: transfer the supernatant to a new 1.5 ml microfuge tube without disturbing the pellet and store at-20°C (media fraction).

Periplasmic extraction

- Resuspend the pellet in 850 µl of PBS (NaCl 137 mM, KCl 2.7 mM, Na2HPO4 24.2 mM, KH2PO4 5.2 mM at pH7.4, 100 µM N-Ethylmaleimide) and transfer to a 1.5 ml microfuge tube. Do not vortex or fractions will become cross-contaminated.
- Centrifuge for 3 min at 20,817g and carefully discard the supernatant.
- Resuspend the pellet in 900 µl of Buffer 1 (100 mM Tris pH 8.0, 500 mM sucrose, 0.5 mM EDTA pH 8.0, 1 cOmplete™ Protease Inhibitor Cocktail tablet (per 50 ml), 100 µM N-Ethylmaleimide) and incubate on ice for 5 min.
- Centrifuge for 3 min at 20,817g and carefully discard the supernatant.
- Resuspend the pellet in 400 µl of Mg-water (Ultrapure sterile water, 100 µM N-Ethylmaleimide, 1 mM MgCl2) by pipetting for about 30 sec and incubate on ice for 2 min.
- Centrifuge for 3 min at 20,817g.
- Transfer the supernatant to a new 1.5 ml microfuge tube without disturbing the pellet and store at-20°C (periplasmic fraction). Make sure not to disturb the pellet to keep the periplasmic fraction pure. The protocol can be stopped at this point and the pellet frozen down at –20°C.

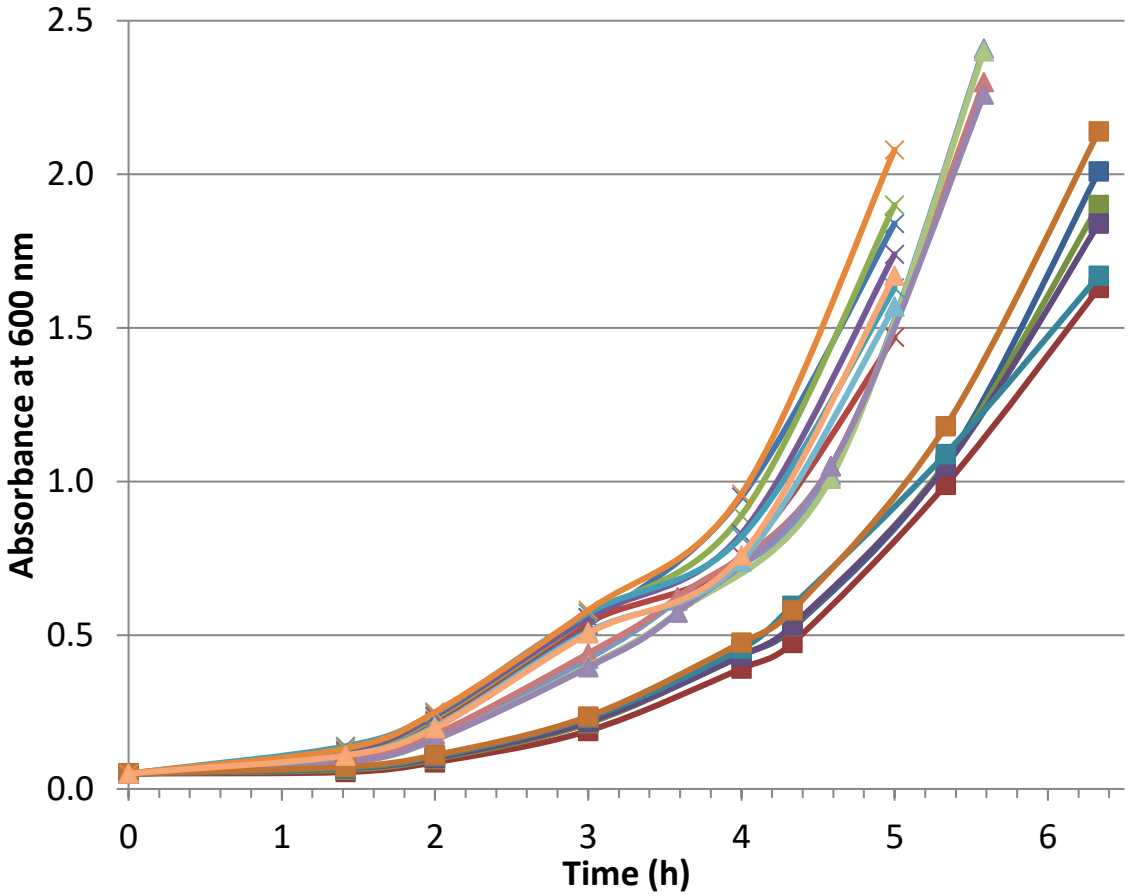
Cell disruption

- Wash the cells by resuspending the pellet in 1 ml of Buffer 2 (50 mM Tris pH 8.0, 250 mM sucrose, 10 mM MgSO4, 1 cOmplete™ Protease Inhibitor Cocktail tablet (per 50 ml), 100 µM N-Ethylmaleimide).
- Centrifuge for 3 min at 20,817g and discard the supernatant.
- Resuspend the pellet in 750 ml of Buffer 3 (50 mM Tris pH 8.0, 2.5 mM EDTA pH 8.0, 1 cOmplete™ Protease Inhibitor Cocktail tablet (per 50 ml), 100 µM N-Ethylmaleimide) and immediately proceed to the next step.
- Ultrasonicate with the following cycle parameters while the sample is incubated in an ice bath.
    - 10 sec amplitude 8 µm
    - 10 sec resting

  } 5 cycles
- Transfer the lysate to ultra-centrifuge tubes.
- Ultra-centrifuge at 186 000 g for 30 min.
- Transfer the supernatant to a new 1.5 ml microfuge tube and store at -20°C (cytoplasmic fraction).
- Resuspend the pellet in 1 ml of Buffer 3 and store at -20°C (insoluble fraction containing membrane fragments and potential inclusion bodies)

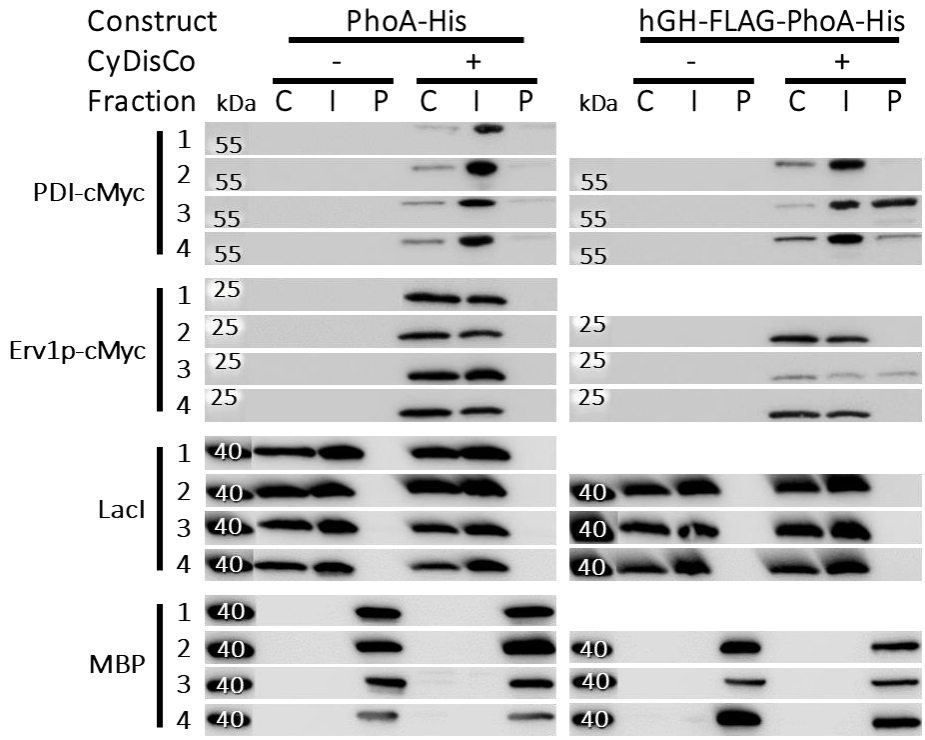
Appendix 5: Growth curves of the initial expression of the 5 NTS

The His-tagged AmiA (dark blue), HyaA (red), PaoA (green), TorA (purple) and YcbK (light blue) were expressed with their respective signal peptide in the W3110 WT (square markers), MC4100 WT (cross markers) and MC4100 Tat-null (triangle markers) strains as well as the empty vector as control (orange). The culture was performed as described in Table 2-3, condition 1.



Appendix 6: Controls for the localisation of hGH fusions with PhoA as the protein of interest

The samples obtained in Figure 7-2 were further analysed. The Western-blots represented the detection of PDI-cMyc, Erv1p-cMyc, LacI and MBP using anti-His, anti-FLAG, anti-cMyc, anti-LacI and anti-MBP antibodies respectively. PageRuler Plus Prestained Protein Ladder (Life Technologies) and MagicMark™ (Life Technologies) molecular weight markers are indicated in kilodalton (kDa).



Appendix 7: Table of the sizes of the proteins expressed in Chapter 7.

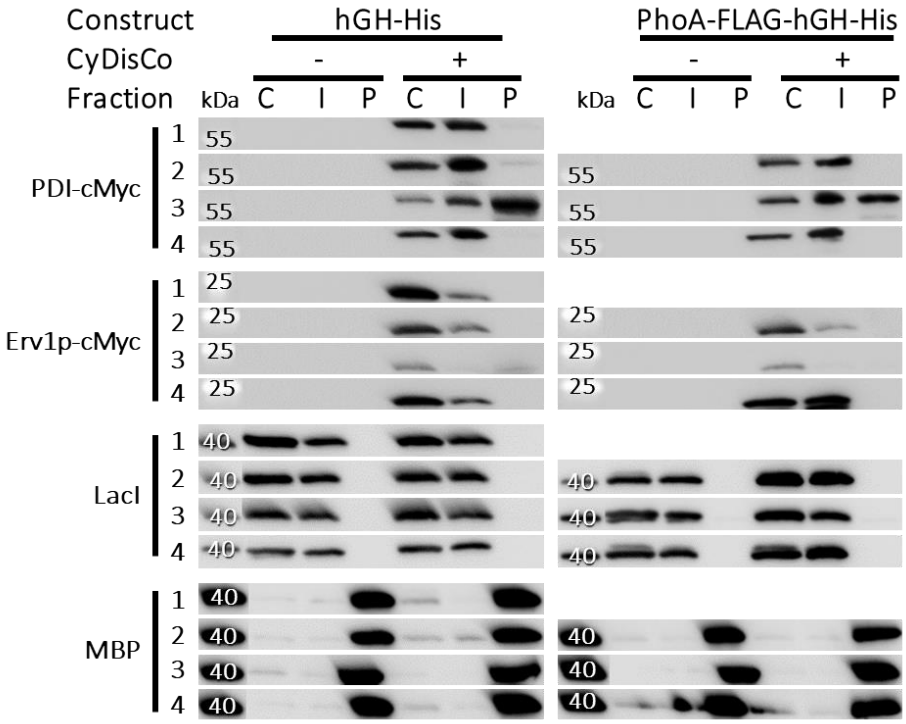
The composition of the constructs is given in terms of signal peptide, soluble partner and protein of interest when applicable. Then, the corresponding protein sizes are indicated for the mature and premature forms when relevant in kilodalton (kDa) and well as the number of disulphide bond (DSB).

Construct			Protein size (kDa)		DSB
Signal peptide	Soluble partner	Protein of interest	Premature	Mature	
-	-	PhoA	-	48.2	2
$OmpA_{SP}$	-	PhoA	50.2	48.1	2
$TorA_{SP}$	-	PhoA	52.8	48.6	2
-	hGH	PhoA	-	74.2	4
$OmpA_{SP}$	hGH	PhoA	76.2	74.2	4
$TorA_{SP}$	hGH	PhoA	78.9	74.7	4
-	-	hGH	-	23.1	2
$OmpA_{SP}$	-	hGH	25.1	23.1	2
$TorA_{SP}$	-	hGH	27.8	23.6	2
-	PhoA	hGH	-	74.3	4
$OmpA_{SP}$	PhoA	hGH	76.2	74.2	4
$TorA_{SP}$	PhoA	hGH	78.9	74.7	4
-	-	FABP4	-	15.5	0
$OmpA_{SP}$	-	FABP4	17.6	15.6	0
$TorA_{SP}$	-	FABP4	20.3	16.1	0
-	hGH	FABP4	-	41.7	2
$OmpA_{SP}$	hGH	FABP4	43.7	41.7	2
$TorA_{SP}$	hGH	FABP4	46.4	42.2	2
-	PhoA	FABP4	-	66.8	2
$OmpA_{SP}$	PhoA	FABP4	68.7	66.7	2
$TorA_{SP}$	PhoA	FABP4	71.4	67.2	2
-	-	VHH1	-	13.8	1
$OmpA_{SP}$	-	VHH1	15.7	13.7	1
$TorA_{SP}$	-	VHH1	18.4	14.2	1
-	hGH	VHH1	-	39.8	3
$OmpA_{SP}$	hGH	VHH1	41.9	39.9	3
$TorA_{SP}$	hGH	VHH1	44.6	40.4	3
-	PhoA	VHH1	-	64.9	3
$OmpA_{SP}$	PhoA	VHH1	66.8	64.8	3
$TorA_{SP}$	PhoA	VHH1	69.5	65.3	3
-	-	VHH2	-	13.8	2
$OmpA_{SP}$	-	VHH2	15.7	13.7	2

Construct			Protein size (kDa)		DSB
Signal peptide	Soluble partner	Protein of interest	Premature	Mature	
$TorA_{SP}$	-	VHH2	18.4	14.2	2
-	hGH	VHH2	-	39.8	4
$OmpA_{SP}$	hGH	VHH2	41.9	39.9	4
$TorA_{SP}$	hGH	VHH2	44.6	40.4	4
-	PhoA	VHH2	-	64.9	4
$OmpA_{SP}$	PhoA	VHH2	66.8	64.8	4
$TorA_{SP}$	PhoA	VHH2	69.5	65.3	4

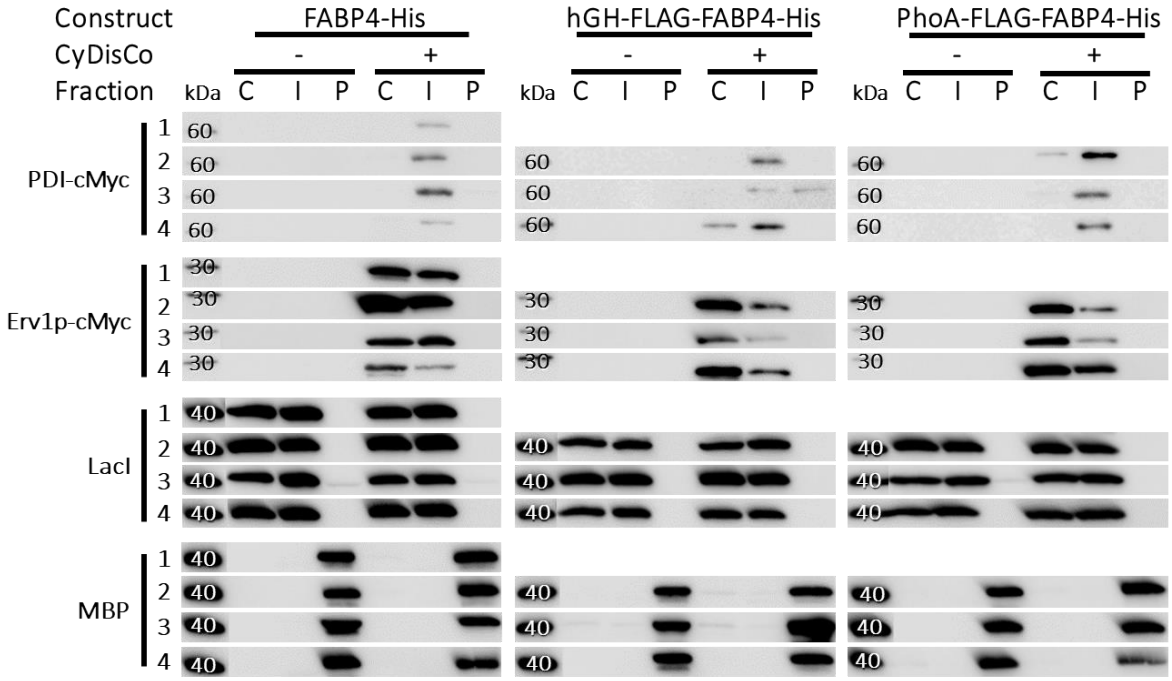
Appendix 8: Controls for the localisation of PhoA fusions with hGH as the protein of interest

The samples obtained in Figure 7-4 were further analysed. The Western-blots represented the detection of PDI-cMyc, Erv1p-cMyc, LacI and MBP using anti-His, anti-FLAG, anti-cMyc, anti-LacI and anti-MBP antibodies respectively. PageRuler Plus Prestained Protein Ladder (Life Technologies) and MagicMark™ (Life Technologies) molecular weight markers are indicated in kilodalton (kDa).



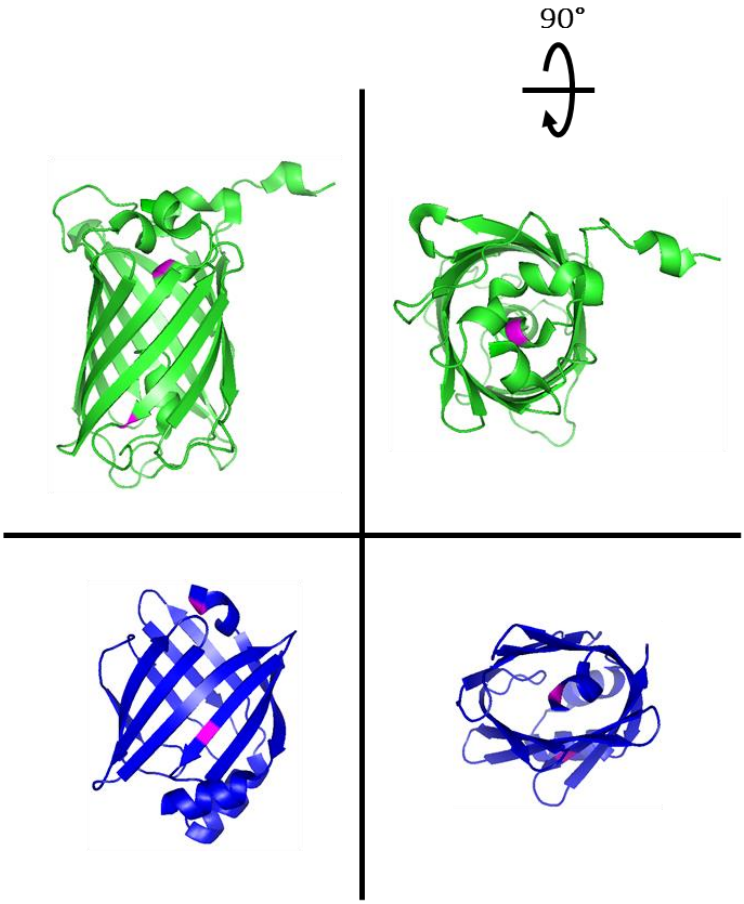
Appendix 9: Controls for the localisation of fusions with FABP4 as the protein of interest

The samples obtained in Figure 7-6 were further analysed. The Western-blots represented the detection of PDI-cMyc, Erv1p-cMyc, LacI and MBP using anti-His, anti-FLAG, anti-cMyc, anti-LacI and anti-MBP antibodies respectively. PageRuler Plus Prestained Protein Ladder (Life Technologies) and MagicMark™ (Life Technologies) molecular weight markers are indicated in kilodalton (kDa).



Appendix 10: Structures of sfGFP and FABP4

The structures of sfGFP and FABP4 were solved at 1.85 and 1.08 Å by X-ray diffraction respectively (Dippel et al., 2016, González and Fisher, 2015). The top figures represented sfGFP and the bottom ones represented FABP4. The right images corresponded to a 90° rotation form the left ones. The cysteine residues were indicated in magenta, the flat arrows represented.



Appendix 11: Primary sequence alignment of the VHHs

The alignment was performed using the online tool EMBOSS Needle with default parameters (Madeira et al., 2019). The cysteine residues are indicated in red.

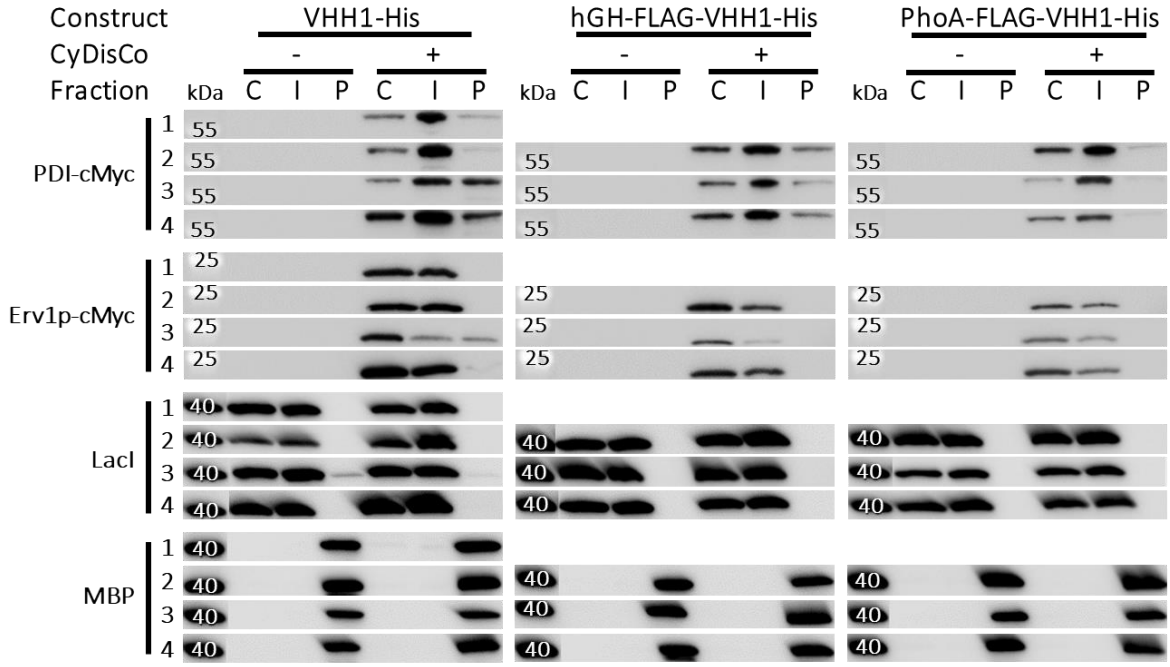
```
VHH1     1 QVQLVESGGGVVQAGGSLRLSCAASERTFSSYTMGWFRQAPGKEREFVAA 50
           |||||||||||:|.:||||.:||.||..:|..|.|.|.|||||||.|::|.
VHH2     1 QVQLVESGGGLVHSGGSLTVSCVASGFSFPYYYMSWVRQAPGKEIEWIAG 50


VHH1    51 ISWSGGSTYYADSVKGRFTISRDNAKNTLYLHMNDLKPEDTAVYYCNGFG 100
           ||.| |..:|..|||||||||||||||||.|:.|||||||:|||...|
VHH2    51 ISPS-GDRHYTSSVKGRFTISRDNAKNTLYLQMSSLKPEDTALYYCARDG 99


VHH1   101 Y-RG---EYWGQGTQVTVSS 116
           . ||   |..||||||||||
VHH2   100 ICRGDNCETPGQGTQVTVSS 119
```

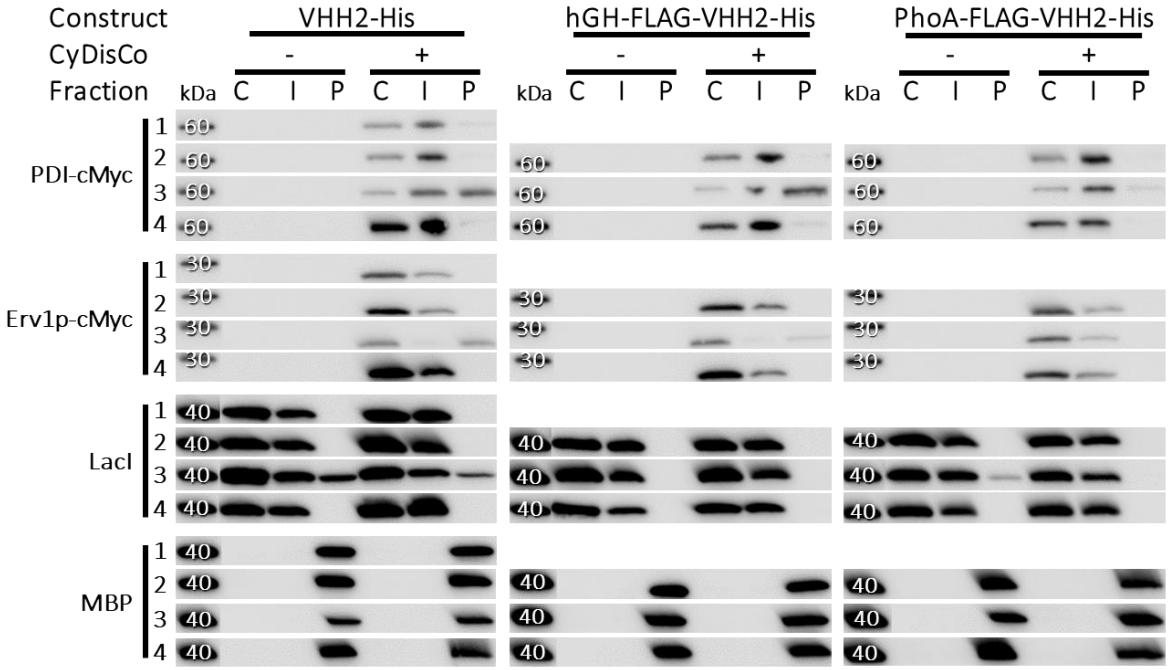
Appendix 12: Controls for the localisation of fusions with VHH1 as protein of interest

The samples obtained in Figure 7-10 were further analysed. The Western-blots represented the detection of PDI-cMyc, Erv1p-cMyc, LacI and MBP using anti-His, anti-FLAG, anti-cMyc, anti-LacI and anti-MBP antibodies respectively. PageRuler Plus Prestained Protein Ladder (Life Technologies) and MagicMark™ (Life Technologies) molecular weight markers are indicated in kilodalton (kDa).



Appendix 13: Controls for the localisation of fusions with VHH2 as protein of interest

The samples obtained in Figure 7-12 were further analysed. The Western-blots represented the detection of PDI-cMyc, Erv1p-cMyc, LacI and MBP using anti-His, anti-FLAG, anti-cMyc, anti-LacI and anti-MBP antibodies respectively. PageRuler Plus Prestained Protein Ladder (Life Technologies) and MagicMark™ (Life Technologies) molecular weight markers are indicated in kilodalton (kDa).



Appendix 14: Alignment of the SecY protein sequences extracted from the project's strains

The nucleotide sequences of *secY* were pulled from the genomic sequencing detailed in section 3.6. and translated into their corresponding protein sequence. The latter were then aligned to the W3110 WT reference genome (AP009048.1, NCBI) using Clustal Omega (Madeira et al., 2019).

```
WT_Ref              MAKQPGLDFQSAKGGLGELKRRLLFVIGALIVFRIGSFIPIPGIDAAVLA 50
W3110_WT            MAKQPGLDFQSAKGGLGELKRRLLFVIGALIVFRIGSFIPIPGIDAAVLA 50
MC4100_WT           MAKQPGLDFQSAKGGLGELKRRLLFVIGALIVFRIGSFIPIPGIDAAVLA 50
MC4100_Tat-null     MAKQPGLDFQSAKGGLGELKRRLLFVIGALIVFRIGSFIPIPGIDAAVLA 50

WT_Ref              KLLEQQRGTIIEMFNMFSGGALSRASIFALGIMPYISASIIIQLLTVVHP 100
W3110_WT            KLLEQQRGTIIEMFNMFSGGALSRASIFALGIMPYISASIIIQLLTVVHP 100
MC4100_WT           KLLEQQRGTIIEMFNMFSGGALSRASIFALGIMPYISASIIIQLLTVVHP 100
MC4100_Tat-null     KLLEQQRGTIIEMFNMFSGGALSRASIFALGIMPYISASIIIQLLTVVHP 100

WT_Ref              TLAEIKKEGESGRRKISQYTRYGTLVLAIFQSIGIATGLPNMPGMQGLVI 150
W3110_WT            TLAEIKKEGESGRRKISQYTRYGTLVLAIFQSIGIATGLPNMPGMQGLVI 150
MC4100_WT           TLAEIKKEGESGRRKISQYTRYGTLVLAIFQSIGIATGLPNMPGMQGLVI 150
MC4100_Tat-null     TLAEIKKEGESGRRKISQYTRYGTLVLAIFQSIGIATGLPNMPGMQGLVI 150

WT_Ref              NPGFAFYFTAVVSLVTGTMFLMWLGEQITERGIGNGISIIIFAGIVAGLP 200
W3110_WT            NPGFAFYFTAVVSLVTGTMFLMWLGEQITERGIGNGISIIIFAGIVAGLP 200
MC4100_WT           NPGFAFYFTAVVSLVTGTMFLMWLGEQITERGIGNGISIIIFAGIVAGLP 200
MC4100_Tat-null     NPGFAFYFTAVVSLVTGTMFLMWLGEQITERGIGNGISIIIFAGIVAGLP 200

WT_Ref              PAIAHTIEQARQGDLHFLVLLLVAVLVFAVTFFVVFVERGQRRIVVNYAK 250
W3110_WT            PAIAHTIEQARQGDLHFLVLLLVAVLVFAVTFFVVFVERGQRRIVVNYAK 250
MC4100_WT           PAIAHTIEQARQGDLHFLVLLLVAVLVFAVTFFVVFVERGQRRIVVNYAK 250
MC4100_Tat-null     PAIAHTIEQARQGDLHFLVLLLVAVLVFAVTFFVVFVERGQRRIVVNYAK 250

WT_Ref              RQQGRRVYAAQSTHLPLKVNMAGVIPAIFASSIILFPATIASWFGGGTGW 300
W3110_WT            RQQGRRVYAAQSTHLPLKVNMAGVIPAIFASSIILFPATIASWFGGGTGW 300
MC4100_WT           RQQGRRVYAAQSTHLPLKVNMAGVIPAIFASSIILFPATIASWFGGGTGW 300
MC4100_Tat-null     RQQGRRVYAAQSTHLPLKVNMAGVIPAIFASSIILFPATIASWFGGGTGW 300

WT_Ref              NWLTTISLYLQPGQPLYVLLYASAIIFFCFFYTALVFNPRETADNLKKSG 350
W3110_WT            NWLTTISLYLQPGQPLYVLLYASAIIFFCFFYTALVFNPRETADNLKKSG 350
MC4100_WT           NWLTTISLYLQPGQPLYVLLYASAIIFFCFFYTALVFNPRETADNLKKSG 350
MC4100_Tat-null     NWLTTISLYLQPGQPLYVLLYASAIIFFCFFYTALVFNPRETADNLKKSG 350

WT_Ref              AFVPGIRPGEQTAKYIDKVMTRLTLVGALYITFICLIPEFMRDAMKVPFY 400
W3110_WT            AFVPGIRPGEQTAKYIDKVMTRLTLVGALYITFICLIPEFMRDAMKVPFY 400
MC4100_WT           AFVPGIRPGEQTAKYIDKVMTRLTLVGALYITFICLIPEFMRDAMKVPFY 400
MC4100_Tat-null     AFVPGIRPGEQTAKYIDKVMTRLTLVGALYITFICLIPEFMRDAMKVPFY 400

WT_Ref              FGGTSLLIVVVVIMDFMAQVQTLMMSSQYESALKKANLKGYGR 443
W3110_WT            FGGTSLLIVVVVIMDFMAQVQTLMMSSQYESALKKANLKGYGR 443
MC4100_WT           FGGTSLLIVVVVIMDFMAQVQTLMMSSQYESALKKANLKGYGR 443
MC4100_Tat-null     FGGTSLLIVVVVIMDFMAQVQTLMMSSQYESALKKANLKGYGR 443
```