Development of *E. coli* strains with novel Tat-based export systems for therapeutic protein production

Kelly May Frain

A thesis submitted for the degree of Doctor of Philosophy

University of Kent

Department of Biosciences
September 2018

Table of Contents

List of figures

List of tables

Acknowledgments

Firstly, I would like to thank my supervisor, Professor **Colin Robinson**. After knowing me for 8 weeks during my final undergraduate lab project, he believed in me enough to guide me through a 4 year PhD. More fool him.

Kelly Walker also had a great deal of input, without her subtle hints, I wouldn't have applied for the position. Kelly you were (are still are) an inspiration, you not only taught me most of the science I know, but you also taught me invaluable life lessons.

My MedImmune team including, **Dr Ray Field**, **Dr Benjamin Kemp**, **Dr Trevor Wilkinson**, **Dr Soojin Han**, **Dr Mark Berge** and **Dr Ronald Schoner**, have also helped along the way. Thank you for your advice during our meetings and my secondment.

Thank you to all CR lab members, past and present, I often think of us all as one big crazy family. **Jo Roobol** and **Ally Walter**, you kept and keep the ship afloat, thank you. **Cristina Matos**, without your data my PhD wouldn't exist, so thank you too. Other supportive Post-docs include **Kirsty Richards** and **Wayne Millar**. **Julie Zedler** and **Doris Gangl**, you both set amazing precedents, your work ethics and values will stick with me.

Amber Peswani and **Conner Webb**, I'm sure your punderful jokes will aid you both through your new PhD's, like they did mine. **Sarah Bischoff**, **Serena Lima** and **Chi Jawara** good luck too, although I know you'll do great! **Isabel Guerrero-Montero**, thank you for your outreach activities, book club and hand knitting were welcome distractions.

Alex Jones, you have known me from the very start and will know me until the very end. Thank you for your patience, encouragement and advice (even though you know I won't listen). "Home safe" would not be the same without you or **Daphne Mermans**. Daphne, you are always giving, your kind nature will take you far, thank you for being you throughout my PhD, I look forward to visiting Stockholm in the near future.

Thank you to my close family including; **Mum**, **Dad**, **Mama**, **Jack**, **Callum** and **Ben**. You taught me nothing is impossible and your undivided support has helped me through University, but I'm sure your continual love will help me through many more years.

Not least, I'd like to thank **Charlie Hope**, as a highly determined and motivated individual, he's taught me to dream big. Who would have thought 7 years ago in Flat 19 Kemsdale Court, I might one day be a Dr?

Declaration

The work presented in this thesis is original, and was conducted myself (unless started otherwise) under the supervision of Professor Colin Robinson. All sources of information have been acknowledged by means of references. None of this work has been used in any previous application for a degree.

I would like to acknowledge Dr Sarah Smith (University of Warwick) for collecting data via CryoEM of my SMALPed TatAyCy samples.

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

Frain, K. M., Jones, A. S., Schoner, R., Walker, K. L., & Robinson, C. (2017). The *Bacillus subtilis* TatAdCd system exhibits an extreme level of substrate selectivity. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1864*(1), 202–208.

Frain, K., Gangl, D., Jones, A., Zedler, J.A., & Robinson, C. (2016). Protein translocation and thylakoid biogenesis in cyanobacteria. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1857*(3), 266–273.

This research was a CASE studentship co-funded by MedImmune (AstraZeneca) and the BBSRC.

Summary

The Twin Arginine Translocase is a protein transport system in the inner membrane of both Gram-negative (*E. coli*) and Gram-positive (*B. subtilis*) bacteria. The translocase is unique as it exports *fully folded* substrates to the periplasm, moreover, it preferentially exports correctly folded proteins. Less information is known about the minimal Gram-positive translocases, TatAdCd and TatAyAy, but their 'proofreading' abilities and structure, respectively, are investigated alongside the usefulness of Tat in the biopharmaceutical industry.

Initial studies (chapter 3) addressed which pharmaceutically recombinant proteins TatAdCd could export with a Tat specific signal peptide, TorA, to the periplasm. Surprisingly, the translocase only exported two scFv's, but by testing export of mutant scFv's I was able to determine, for the first time, the level of proofreading selectively by a Gram-positive Tat system.

Given this knowledge, 8 scFv constructs in varying orientations from MedImmune were tested exploiting TatAdCd, Tat express and CyDisCo (chapter 4). I endeavored to optimize conditions for their expression and export through means of alternative media's and the fermentation process.

Finally, given that Gram-positive Tat has unusual specificities and a minimal translocase, I used novel approaches to purify and characterise TatAyCy with SMA and CryoEM techniques (Chapter 5). The detailed insight provided in this study could help towards understanding the elusive translocation mechanism of Tat.

Abbreviations

°C	Degrees celsius
Å	Ångströms
AIM	Auto Induction Media
APH	Amphipathic Helix
APS	Ammonium persulphate
AT	Alpha toxin
ATP	Adenosine triphosphate
BBSRC	Biotechnology and Biological Sciences Research Council
BSA	Bovine serum albumin
C	Cytoplasmic fraction
C-	Carboxy terminus
CASE	Collaborative Awards in Science and Engineering
CEA	Carcinoembryonic antigen
CV	Column volume
CyDisCo	Cytoplasmic Disulphide bond formation in *E. coli*
DDM	n-Dodecyl-ß-D-maltoside
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DS	Disulphide stabilised
E. coli	*Escherichia coli*
ECL	Enhanced chemifluoresence
EDTA	Ethylenediaminetetraacetic acid

EM	Electron microscopy
ER	Endoplasmic reticulum
Erv1p	Yeast mitochondrial thiol oxidase
FT	Flow through
g	grams
GF	Gel filtration
GFP	Green fluorescent protein
hGH	Human growth hormone
His	Hexa-histidine tag
hPDI	Human protein disulphide isomerase
hr(s)	Hour(s)
HRP	Horseradish peroxidase
IFN	Interferon
IPTG	Isopropylthiogalactoside
IMAC	Immobilised metal ion affinity chromatography
kDa	Kilodaltons
L	litre
LB	Luria Bertani broth
LBA	Luria Bertani agar
M	Membrane fraction
M	Molar
mAU	Milli-absorbance unit
mg	Millgram
min	Minutes

mL	Millilitre
mM	Millimolar
N-	Amino terminus
NBD	Nucleotide Binding Domain
NEB	New England Biolabs
nM	Nanometers
NMR	Nuclear Magnetic Resonance
OD	Optical density
O/N	Overnight
P	Periplasmic fraction
PAGE	Polyacrylamide gel electrophoresis
PBS	Phoshate buffered saline
PBS-T	Phoshate buffered saline and tween
PCR	Polymerase Chain Reaction
PMF	Proton motive force
POI	Protein(s) of interest
PVDF	Polyvinylidene fluoride membrane
RNA	Ribonucleic acid
RNC	Ribosome nascent chain complex
PY	Phtyone
Rpm	Revolutions per minute
RR	Twin-arginine motif
RT	Room temperature
SB(s)	Salt bridge(s)

scFv	Single-chain variable-fragment from an antibody
scFvM	scFv raised against C-met
scFvO	scFv raised against the omega peptide of β-galactosidase
SDS	Sodium dodecylsuphate
Sec	General secretory pathway
SEC	Size Exclusion Chromatography
SM6	Synthetic medium 6
SMA	Styrene- maleic acid
SMALP	Styrene- maleic acid lipid particle
sp	Signal peptide
SRP	Signal recognition particle
Strep-tag	Strep II tag
Tat	Twin arginine translocation
TB	Terrific broth
TEMED	Tetramethylethylenediamine
TF	Trigger Factor
TM	Transmembrane
T_m	Melting temperate
TorA	Trimethylamine-N-oxide reductase
TorA-	TorA signal peptide fused to following protein
Tris	Tris (hydroxymethyl) aminomethane
Tween20	Polyoxyethylenesorbitan monolaurate
UV	Ultra violet

VH	Variable heavy region
W	Wash fraction
WT	Wild type
w/v	Weight per volume
Δ (Delta)	Gene deletion
µL	Microlitre
µM	Micromolar

1. Introduction

In order to function correctly and efficiently, every cell needs to be highly organised, tightly regulated and compartmentalised. Although prokaryotic cells are simpler than eukaryotic cells, bacteria too require specialised machinery to grow and thereafter thrive in niche environments.

Proteins are essential macromolecules synthesised by ribosomes in the cytoplasm that often require localisation before they can carry out their particular purpose. Their proper formation, targeting and function are imperative to the survival of the cell. To name but a few functions, proteins replicate DNA, catalyse reactions, allow cellular locomotion and transport substrates from one location to the next.

Prokaryotic bacteria can be broadly classified into two groups, Gram-positive and Gram-negative. This is based on their ability to retain Gram-stain, which varies due to distinctive compartmentalisation differences. Gram-positive bacteria have a single cell membrane enclosed with a thick peptidoglycan wall whereas Gram-negative bacteria have a more complicated double cell membrane with a thin peptidoglycan wall. Hence, protein transport and destination can vary between the two classifications, see Figure 1.

20-30% of proteins synthesised in the cytoplasm of bacteria are destined for extra-cytosolic locations (Holland 2004), they therefore have to pass a cell membrane composed of a tightly sealed lipid bilayer intent on keeping the cell structurally sound and impenetrable. Specialised transport systems exist within the cell and the cell membrane to cross this barrier, including the Twin Arginine Translocase (Tat) pathway.

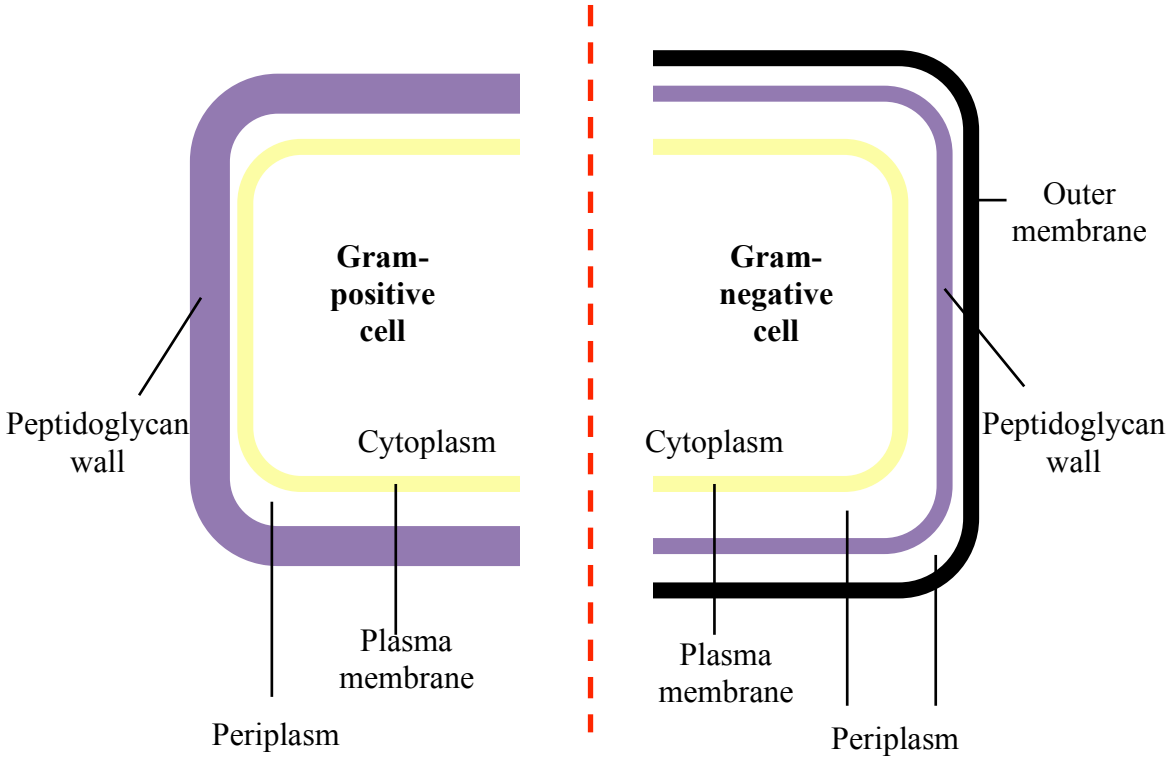


Figure 1. **The differences between Gram-positive and Gram-negative cells**
A schematic representation of a Gram-positive cell (left) compared to a Gram-negative cell (right). Both cell types have a cytoplasm contained by the plasma membrane, encompassed by periplasm and the peptidoglycan cell wall. A Gram-positive cell has a thicker peptidoglycan wall compared to a Gram-negative cell. A Gram-negative cell has an additional cellular compartment, the outer membrane, which gives rise to an addition periplasmic space.

1.1.1 Protein translocation in bacteria

A variety of protein transport mechanisms ensure proteins are translocated across or into the phospholipid bilayer without compromising its structural integrity or function. Each mechanism made up of essential proteins is as specialised as the protein substrate intended to use each system. However common features tie the pathways together, which guarantee cell regulation and safety, including: A gated pore (which stops unnecessary substance leakage), an energy requirement (to push the substrate through the membrane) and the use of signal peptides (that direct the protein to the correct translocase and intimately, the correct location).

Two major transport systems exist for protein translocation and membrane insertion, The General Secretory (Sec) Pathway and the Twin Arginine Translocase (Tat) Pathway. Most proteins use the more abundant Sec pathway that is shared throughout all domains of life. In Eukarya the pathway functions in the ER whereas in Prokarya, it resides mainly in the plasma membrane to export unfolded proteins. The Tat pathway, the focus of this thesis, is much more exclusive, it has only around 30 native *E.* coli substrates (Palmer et al. 2011) and each protein is fully folded in the cytoplasm prior to Tat export. Figure 2, a schematic diagram, outlines the two alternative pathways by depicting their differences.

To ensure proteins are directed to the appropriate pathway, signal peptides are encoded on the N-terminus of each protein. Their unique codes, or more exactly, amino acids guide the polypeptide through the cell to its correct location.

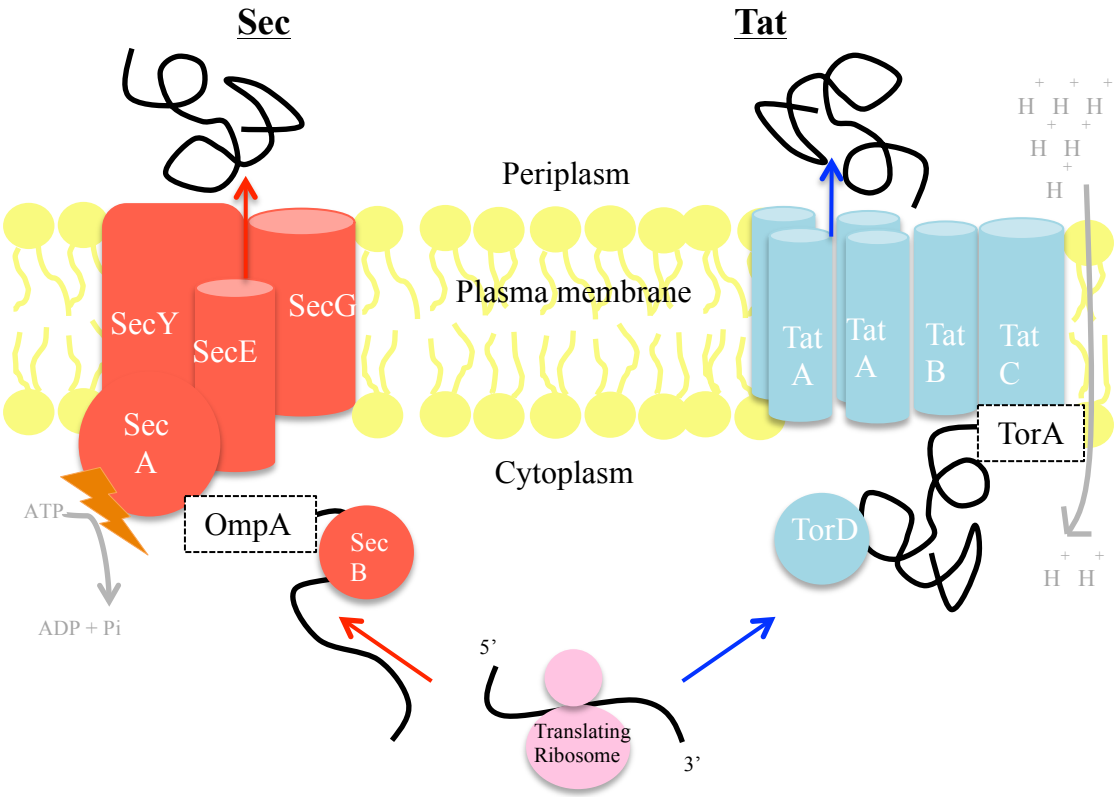


Figure 2. **Two alternative protein translocation pathways, Sec and Tat.**
Proteins always originate from translating ribosomes and their N-terminal signal peptide (OmpA or TorA in this overview) will direct the nacent polypeptide chain to the correct translocase (Sec or Tat, respectively), aided by chaperones. The unfolded Sec protein is transferred to SecA where it is threaded through the SecYEG channel in the plasma membrane, powered by repeated cycles of ATP hydrolysis. In the oxidising periplasm, the unfolded protein matures into its tertiary fully folded protein. The Tat protein is fully folded within the cytoplasm (perhaps with a cofactor) and once directed to TatBC, TatA protomers recruit to translocate the protein across the plasma membrane. The energy required for this process is created by proton motif force.

1.1.2 Targeting via the signal peptide

Substrates are targeted to the correct translocase by virtue of an N-terminal signal peptide with its mature protein in tow. The signal peptide and mature protein create a precursor protein. On the *trans* side of the membrane the precursor is cleaved by a signal peptidase to release just the mature protein (von Heijne 1990) (Lüke et al. 2009). However resident plasma membrane proteins possess an N-terminal signal anchor sequence (usually in the first TM domain) which lacks a signal peptidase cleave site (Sakaguchi et al. 1992).

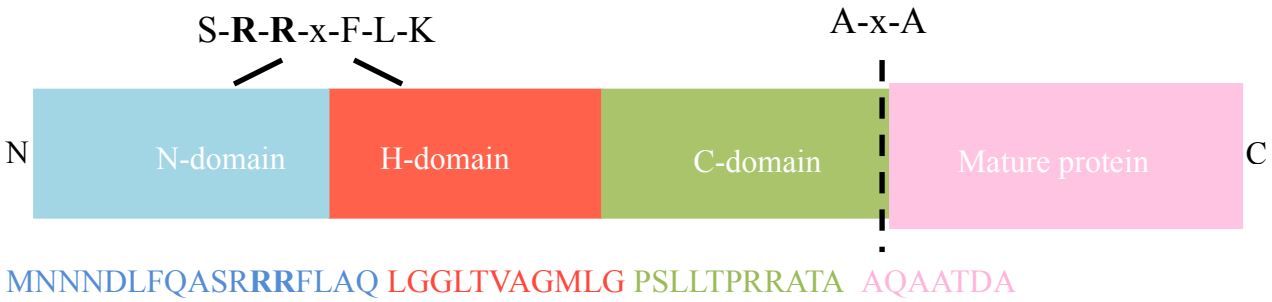
Signal peptides differ in amino acid composition, but all bacterial signal peptides have a core tripartite structure: a positively charged N-terminal domain, a hydrophobic H-domain and a carboxylase C-terminal domain (Berks 1996). The aforementioned 'cleavage site' which is usually an Ala-x-Ala motif resides in the extreme of the C - domain. Figure 3 outlines these distinct regions with both Tat and Sec signal peptide examples.

The signal peptides are nevertheless distinguishable, a distinct difference is defined in the Tat pathway name, a highly conserved twin arginine motif is present in nearly all Tat signal peptides. A larger motif, S-R-R-x-F-L-K (where x is any polar amino acid) is present in >50% of Tat signal peptides and spans both the N- and H- domains (Sargent et al. 1998). The importance of the other preserved amino acids in the larger motif depends upon the substrate and is varied in different bacteria (Berks et al. 2003). However, RR residues are invariant and essential for protein export, for example the substitution of RR > KK (a charge neutral change) blocks Tat export completely (Stanley et al. 2000). A single Arg to Lys mutation in most bacteria only slows down the rate of translocation (Buchanan et al. 2002) but interestingly in chloroplasts RR > KR is tolerated but RR > RK stops transport

altogether (Chaddock et al. 1995)(Halbig et al. 1999). Indeed, the TtrB subunit of the tetrathionate reductase in *Enterobacteriaceae* is the only known native Tat substrate to have a KR motif (Hinsley et al. 2001) and experimentally the single substitution of Arg to Glu has been reported as tolerated too (DeLisa et al. 2002). Aside from the RR motif, other residues within the larger twin arginine motifs are important. The phenylalanine amino acid is important and present in 80% of Tat signal peptides, substitutions showed its high hydrophobicity is essential (Stanley et al. 2000). But what other differences are there between the Tat and Sec signal peptides?

Tat signal peptides are at least 30 residues long, which is longer than Sec signal peptides, which are between 17 and 24 residues in length. The extra residues reside in the basic N-domain and the hydrophobic H-domain (Kipping et al. 2003). Tat signal peptides are also overall less hydrophobic than Sec signal peptides, if the hydrophobicity of the signal peptide is increased, a Tat substrate can be directed to the Sec pathway (Cristóbal et al. 1999). Additionally, the C-domain of Tat signal peptides often contains basic residues (before the A-x-A motif) that are rarely found in Sec signal peptides. Several studies have suggested these basic residues are the 'Sec avoidance motif' whilst another study also including the AxA motif and the first five amino acids of the mature protein as 'Sec avoiding' (Bogsch et al. 1997)(Tullman-ercek et al. 2009).

Tat, TorA



Sec, OmpA

Figure 3. **Comparison of Tat (TorA) and Sec (OmpA) signal peptide.**

The structure of both signal peptides contains three regions a basic N-domain (blue), a hydrophobic H-domain (red) and a polar C-domain (green). A peptidase cleavage site (AxA) follows prior t the mature protein (pink). Below in corresponding colours are the amino acid sequences of TorA and OmpA signal peptides. Tat signal peptides (above) have a consensus motif containing the twin arginines, Sec signal peptides do not contain this. The Sec signal peptides tend to be shorter, with fewer residues in the N and H-domain.

1.2 The General Secretory Pathway (Sec)

The Sec translocase is a ubiquitous membrane protein complex that serves to transport unfolded protein across or into the plasma membrane. The SecYE core channel is conserved across all Kingdoms of life, highlighting its importance, but this chapter will focus mainly on the bacterial Sec system. Indeed, *E. coli* used the Sec pathway for around 96% of its exportome (Orfanoudaki & Economou 2014).

Around 30 years ago genetic screens identified the main components of the *E. coli* Sec translocase and in the '90s the translocation pathway was reconstituted in vitro, which lead to the discovery of the intricate reactions creating the translocation event (Wickner & Leonard 1996). Several other Sec components were identified too, such as dedicated chaperones, SecB and SRP (Valent et al. 1998) alongside the motor which drives translocation, SecA ATPase (Lill et al. 1989). To understand the details of Sec function, its known structure must be revealed first.

1.2.1 Sec membrane components

The bacterial Sec translocon consists of three main components: SecY, SecE and SecG. SecY (48 kDa) is composed of 10 TM helices that can be divided into two sub-domains, TM 1-5 and TM 6-10. These form a clamp that are related to each other by a two-fold pseudo-symmetry axis and hence constitute the aqueous pore polypeptides are translocated through (Van Den Berg et al. 2004). The proposed translocation channel is located in the centre of the Y-subunit, which is blocked by an extension of TM2a, 'the plug'. Its distorted helix extends halfway to the centre of the membrane but cross-linking studies have shown that upon substrate binding, 'the plug' is displaced to open the channel (Van Den Berg et al. 2004)(Tsukazaki et al. 2008). 6 isoleucine residues line the tightest part of the channel

which is an hour-glass shape and this hydrophobicity is predicted to form the seal around the transporting polypeptide to prevent any unwanted leakage and guide the substrate through (Park & Rapoport 2012). The interface between TM2 and TM7 form the 'lateral gate', contact with the pre-protein here induces the 2 halves to separate, allowing signal sequences into the lipid bilayer whilst TM5 and TM6 form 'the hinge' (Maillard et al. 2005).

Another component of the Sec translocon is SecE (14 kDa) (Schatz et al. 1989), without SecE, SecY is unstable and degraded, so SecE is essential for cell viability (Kihara et al. 1995). *E. coli* SecE folds into 3 TM helices and although 2 are non essential (Murphy & Beckwith 1994) the other TM docks across the interface of the two SecY-domains, clamping the 'clamshell' closed. Specific protein-protein contact was identified at the C-terminal end of SecE and TM2a (Harris & Silhavy 1999). SecG (12 kDa) in contrast to SecYE is not essential and does not exist outside the Bacterial Kingdom (Cao & Saier 2003). With 2 TM helices the subunit stabilises SecYE and enhances translocation rate by stimulating SecA activity (Matsumoto et al. 1998). Indeed SecG becomes more important when SecA is compromised by mutations or when SecA is critical (for example at low temperatures or low Transmembrane PMF) (Hanada et al. 1996).

Described above and in Figures 2 and 4 are the membrane components of the Sec pathway, other soluble proteins are yet to be explained. These include proteins such as the Signal Recognition Particle (SRP), SecA, SecB, and other components that aid the translocation event, seen in Figure 4.

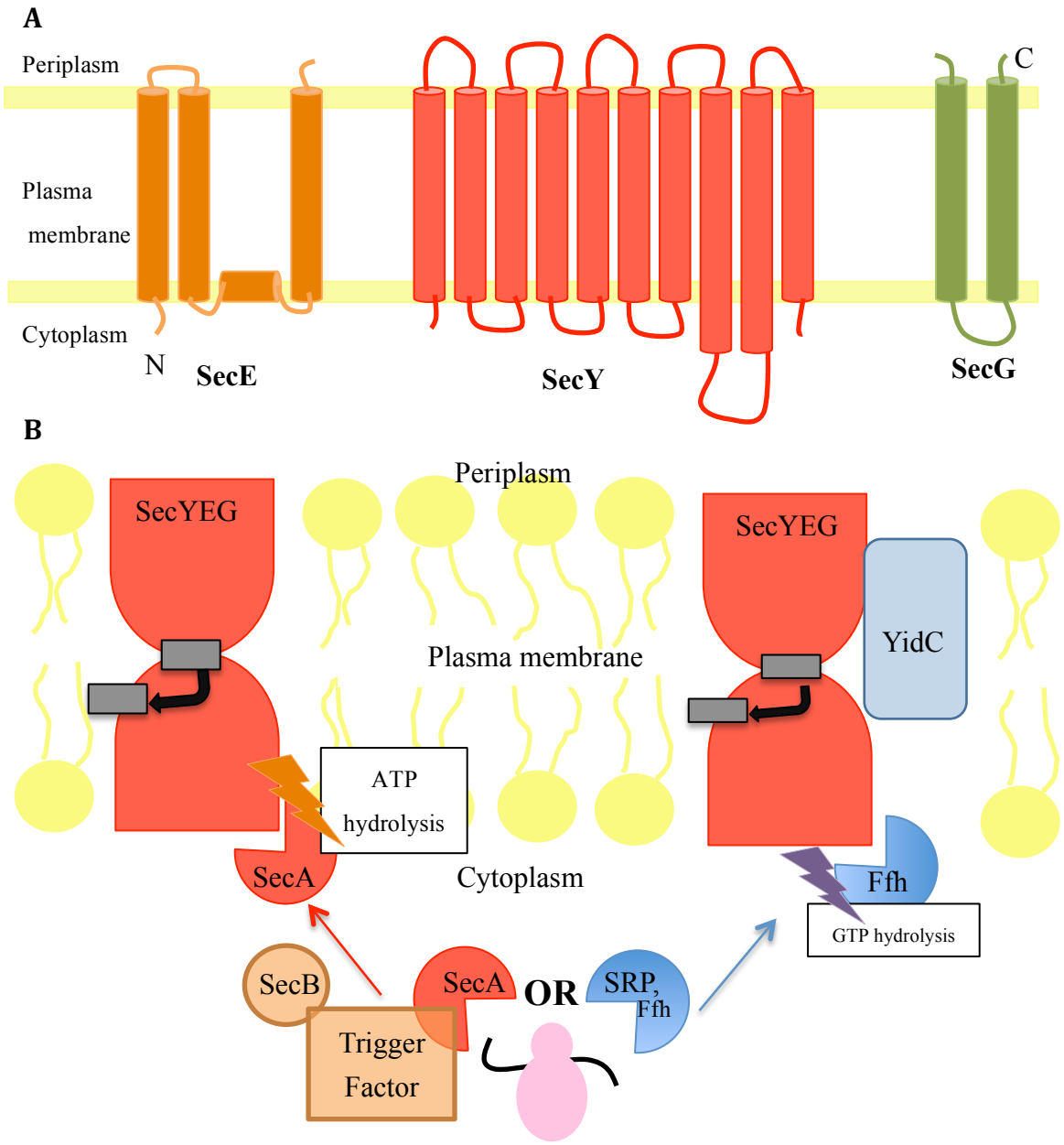


Figure 4. **The Secretory pathway is either SecA or SRP dependent.**
A. Schematic of *E.coli* SecYEG. SecE (orange) has three transmembrane helices, SecY (red) has ten transmembrane helices and SecG (green) has two transmembrane helices. **B. N**ascent polypeptide chain emerges from the ribosome tunnel (pink), if it has high hydrophobicity and is destined to be a membrane protein, a Signal Recognition Particle named Ffh will bind. Once bound, the complex is directed to SecYEG (red) and the GTPase activity of Ffh drives the disassociation of the Ribosome Nascent Chain from the its own complex onto SecYEG. The RNC is inserted int the SecY channel and when the 'plug' (grey) is displaced, the protein can enter its final destination, the inner membrane. Sometimes YidC is involved. If the protein is less hydrophobic and destined for the periplasm, the cytosolic chaperone Trigger Factor binds to the signal peptide of the protein. SecB also plays a roll in preventing protein folding but SecA is the motor of this pathway. Its association with SecYEG stimulates the hydrolysis of ATP, the 'plug' (grey) moves from SecY and protein translocated to the periplasm.

1.2.2 The SRP pathway

As the nascent polypeptide chain emerges from the ribosome tunnel, its encoded signal sequence dictates its rightful pathway. If there's a high degree of hydrophobicity (von Heijne 1986) and helicity (Bruch et al. 1989), which are usually indicative of membrane protein, RNA is bound to the SRP. The SRP pathway is best described for Eukaryotes but it exists in Bacteria too.

In *E. coli* the SRP protein subunit is named Ffh (Fifty-four homologue). The 48 kDa GTPase protein, Ffh and 4.5S RNA make up the ribonucleoprotein complex, in this co-translocational process. Ffh has 3 domains, the C-terminal M domain binds substrate via its methionine rich region, which is predicted to provide a hydrophobic surface for signal sequence interaction (Zheng & Gierasch 1997) (Bernstein et al. 1989). Once bound, the Ribosome nascent-chain (RNC) complex is targeted to the inner membrane where it associates with its membrane anchored receptor FtsY that is located on the ribosome binding site of SecY (Kuhn et al. 2011).

The GTPase activity of Ffh is unusual; it is still unclear what ultimately drives the completion of substrate handover. But it is understood that GTP hydrolysis drives the dissociation of the RNC-SRP complex so that the RNC fully docks onto the SecYEG machinery. The hydrolysis also drives the disassembly and recycling of the constituents, allowing them to continue protein targeting (Akopian et al. 2013).

The signal anchor (SA) sequence of the RNC is inserted into the SecY channel and moves towards the lateral gate (Egea & Stroud 2010). When the 'plug' is displaced in the SecY translocon, the protein enters its final destination, the inner membrane. 60% of all Sec-

transported proteins are destined for the plasma membrane (Orfanoudaki & Economou 2014). See Figure 4.

1.2.3 The Sec pathway and YidC

On the other hand, the nascent polypeptide chain may emerge from the ribosome tunnel less hydrophobic, in which case the cytosolic chaperone Trigger Factor (TF) binds to the signal peptide to prevent SRP binding (Beck et al. 2000). The protein is hence fully synthesised prior to locating its destination in this post-translocational pathway.

Other chaperones too play a role, for example, SecB. Multiple copies of SecB bind to a pre-protein that lacks tertiary structure to maintain its unfolded state and to prevent/delay aggregation (Kumamoto 1989)(Bechtluft et al. 2007). Moreover, SecB relays pre-proteins to either the smaller portion of cytosolic dimer SecA or the monomer SecA, associated with the phospholipid head groups and SecYEG (Lill et al. 1990). To note, when not bound to SecB, SecA only transiently interacts with SecYEG. However SecA-SecYEG interaction increases the affinity of SecA-SecB complex (Hartl et al. 1990), which is explained soon.

SecA is an essential, large 102 kDa protein which acts as the Sec translocase motor. It's Nucleotide Binding Domain (NBD) 1 and NBD2, the intra-molecular regulator of ATPase 2, sandwich a single nucleotide (Tsirigotaki et al. 2017). NBD2 docks SecA onto SecY at which stage, the greater interactions between SecA and SecB leads to an exchange of ADP for ATP at NBD1, SecB is then released. Hydrolysis of ATP at NBD1 promotes NBD2 ATP hydrolysis. The pre-protein stimulates ATP hydrolysis through the triggered SecA protein and the proteins' mature domain becomes trapped in the SecA-SecYEG complex, SecA monomerises (Gouridis et al. 2009). Through unknown mechanisms, the signal

peptide relocated to the 'lateral gate' and the 'plug' from SecY is moved to translocate the protein through the channel approximately 30-40 amino acids at a time (Schiebel et al. 1991).

The translocation is achieved by either SecA acting as a piston to push the protein through by ATP binding and hydrolysis, like power strokes (Bauer et al. 2014) or SecA acting as an 'allosteric channel regulator, SecA conformations control SecYEG through cycles of ATP hydrolysis. The latter would facilitate a Brownian-ratchet mechanism, the pre-protein diffuses across the membrane without energy requirement and SecA prevents chain backsliding via PMF (Allen et al. 2016).

Once most of the mature protein is translocated, SecA loses contact with SecY and ATP stimulation is lost. Signal peptidases cleave the signal peptide and the mature protein is released to the *trans* side of the plasma membrane (Auclair et al. 2012).

Some proteins require the aid of YidC, a 62 kDa homologue to both Oxa1p and Alb3 found in mitochondrial inner membranes and thylakoid membranes, respectively (Saaf et al. 1998). Their job is to insert proteins into the membrane, YidC is essential for the F_0 domain of the *E. coli* F_1F_0 ATPase and cytochrome *o* oxidase of the electron transport machinery. With 6 TM domains, YidC receives these SRP substrates and mediates their insertion to SecY. Indeed, it has been shown that YidC interacts directly with FtsY (Welte et al. 2012) and that it can occupy the SecYEG translocon (Sachelaru et al. 2013). However, this method of translocation is poorly understood.

1.3 The Twin Arginine Translocase (Tat) Pathway

The Twin Arginine Translocase (Tat) pathway differs from the other protein pathways discussed so far because Tat transports fully folded proteins.

In the early 1990's, the alternative translocase was discovered in the thylakoid membrane of chloroplasts, which worked in parallel to the Sec pathway (Mould & Robinson 1991). Initially named the ΔpH-dependent pathway in thylakoids (due to its unusual sole requirement of a proton gradient for translocation), (Klösgen et al. 1992), three membrane proteins were soon identified in thylakoids as essential for translocation of fully folded proteins (Clark & Theg 1997). Tha4 (Mori et al. 1999), Hcf106 (Settles et al. 1997) and cpTatC (Cline & Mori 2001).

Not long after, homologues to the complex were found expressed in some bacteria, archaea and mitochondria (Yen et al. 2002). In *E. coli*, Tat substrates are exported across the plasma membrane into the periplasm via TatA, TatB and TatC (homologues to thylakoid proteins Tha, Hcf106 and cpTatC, respectively). (Bogsch et al. 1998)(Sargent et al. 1998)(Sargent et al. 1999). This system was named Mtt, Membrane Targeting and Translocation and the importance of the complex in *E. coli* was highlighted when several integral proteins were mislocalised in Mtt knockouts (Weiner et al. 1998). Finally however, the system was renamed the Twin Arginine Translocase (Tat) pathway as virtually every protein targeted to it has a twin arginine motif in its signal peptide.

The function of the Tat pathway is to transport a subset of more complex proteins that require cofactor insertion or immediate oligomerisation (Berks 1996). Therefore, the

pathway is essential for many processes including energy metabolism, cell division, cell envelope biogenesis, quorum sensing, motility, symbiosis and pathogenesis (Palmer et al. 2005)(Bernhardt & De Boer 2003)(Ize et al. 2003)(Ding & Christie 2003). Tat can even export more complex heterologous proteins than Sec. Tightly folded proteins such as dihydrofolate reducase (bound to methotrexate) and GFP with a Tat signal peptide are some early examples (Hynds et al. 1998)(Santini et al. 2001), now more bio-pharmaceutically relevant proteins have too been exported via Tat (Alanen et al. 2014). But another intriguing attribute of the pathway lies in its innate ability to detect unfolded or mutated proteins. Indeed, Tat can 'proofread' proteins and reject them for export if they have not been synthesised and folded correctly (Delisa et al. 2003)(Richter & Brüser 2005a).

This unique platform means recombinant proteins are produced with high quality; products are homogeneous and relatively pure from the periplasm. However after nearly three decades of studies, little is still known about Tat's export mechanism, its proof reading ability and its potential in the Biotechnology industry.

1.3.1 Bacterial Tat genes

The Tat system in *E. coli* is expressed from the *tatABCD* operon and while they are all constitutively expressed, TatA is expressed 50 times and 25 times more than TatC and TatB, respectively (Jack et al. 2001). This difference reflects the final component make-up of the Tat translocase in the plasma membrane. *tatD*, is expressed further downstream and whilst originally its function was unknown, more recently TatD is thought to aid proofreading proteins (Matos et al. 2008). Elsewhere in *E. coli's* genome, *tatE* is expressed, TatE is thought to be a late gene duplication of TatA (Baglieri et al. 2012). Although Δ*tatABCDE* strains are viable, as the Tat system is important for many essential cellular functions, pleiotropic cells defects including impaired septation, decreased motility and an increased sensitivity to detergent are apparent (Stanley et al. 2001).

The Tat complex can minimally function with TatA-like and TatC-like proteins, although this varies greatly between organisms (Goosens et al. 2014b). *Bacillus subtilis* expresses two minimal pathways, *tatAdCd* and *tatAyCy*, from different operons (Jongbloed et al. 2004). TatAdCd is expressed with its only known substrate, PhoD during phosphate starved conditions (Beck et al. 2009). In contrast, TatAyCy is expressed constitutively with several substrates such as YwnN (EfeB), QcrA and YkuE (van der Ploeg et al. 2011)(Monteferrante et al. 2012). A third TatA gene, *tatAc* is expressed elsewhere in the genome, it is thought TatAc could represent an intermediate step between TatA and TatB (Goosens et al. 2015).

The organisation of both *E. coli* and *B. subtilis tat* genes are shown in Figure 5.

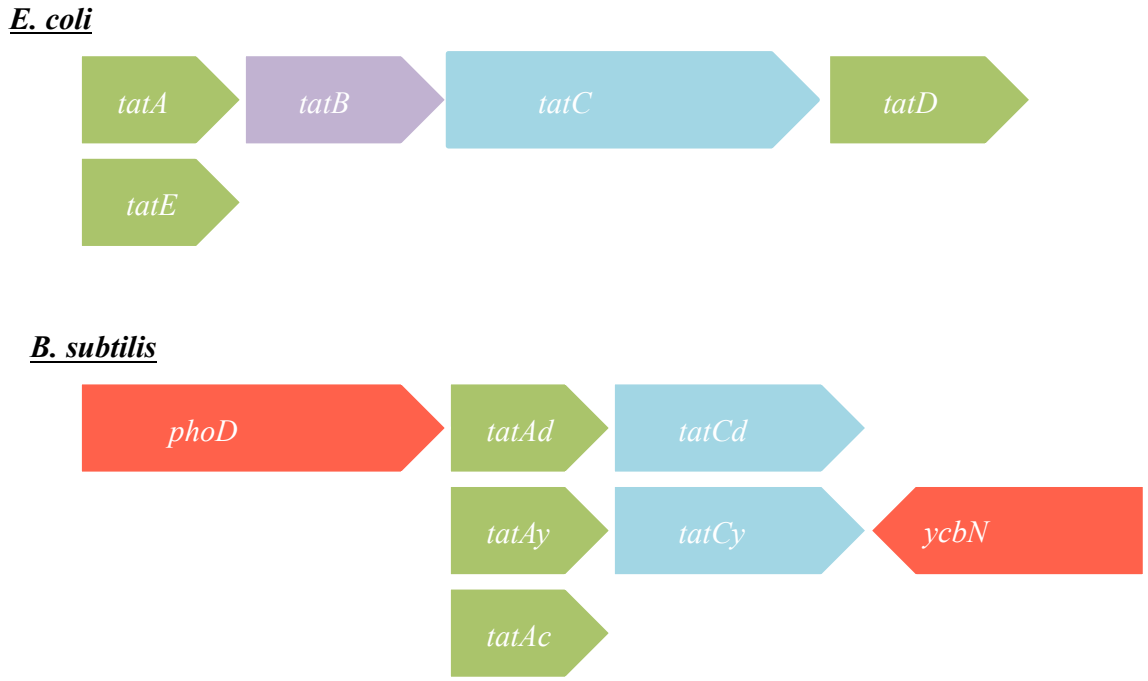


Figure 5. **Organisation of bacterial Tat genes in *E. coli* and *B. subtilis***

E. coli tatABC genes (above) are constitutively expressed from a single gene, these make up the essential translocase. *tatD* is expressed from the same operon, but further downstream, whereas *tatE* is expressed elsewhere in the genome. *B. subtilis tat* genes (below) are expressed in separate operons alongside their respective substrates. *tatAdCd* is expressed alongside *phoD* and *tatAyCy* is expressed with *ycbN*. *tatAc* is expressed alone.

1.3.2 *Escherichia coli* Tat components

1.3.2.1 TatA

TatA is the most abundant component of the Tat complex, most likely responsible for forming the translocase channel (Sargent et al. 2001). *E. coli* has a core TatA protein but it also has TatA-like proteins, TatB and TatE which arose due to gene duplication events, but following sequence divergence now have specialised functions, discussed later (Yen et al. 2002).

B. subtilis has three TatA proteins, TatAd, TatAy and TatAc (Goosens et al. 2014). TatA and TatA-like proteins can be interchanged and function both intra / inter- bacteria. For example, TatE can substitute TatA (Baglieri et al. 2012), TatAd can substitute TatA and TatB (Barnett et al. 2008) and TatAc can form active translocases with TatCd and TatCy (Monteferrante, Baglieri, et al. 2012) and functionally replace TatA and TatE (Beck et al. 2013). This is interesting as the functional overlaps suggest a universal translocation mechanism independent of TatA protein size as *E. coli* TatA is 9.6 kDa (Porcelli et al. 2002) and *B. subtilis* TatAd is 7 kDa (Barnett et al. 2008).

Table 1 displays *E. coli* and *B. subtilis* Tat protein components sizes and their complex sizes.

TatA and TatA-like proteins have similar structures, a short N-terminal domain which is exposed to the periplasm (Koch et al. 2012), followed by a single TM helix linked to an amphipathic helix (APH) which lies against the membrane (Hu et al. 2010) and the C-domain which forms an unstructured and hydrophilic tail in the cytoplasm (Rodriguez et al. 2013). NMR experiments on *B. subtilis* TatAd have confirmed the L shape arrangement

and have indicated the TM helix (14-16 residues in length) is tilted by 17^{o}, which has translocation implications (Müller et al. 2007). This arrangement was supported and enhanced with the knowledge that the following APH is aligned with a comparatively steep tilt angle against the membrane (Walther et al. 2010). Contrary to the previous predicted conformation, which implied the APH was flat against the membrane (Lange et al. 2007), the APH now too has a key role in transformation assembly.

Surprisingly, not many substitutions in TatA block export, but there are a few instances. In *E. coli* a glycine residue in the "hinge region" connects the TM helix and the APH, indeed, the mutation G33A stops export, suggesting this a key residue alongside others (Barrett et al. 2003). The mutant F39A also inactivates the translocase, this residue anchors the APH to the membrane (Hicks et al. 2003). Lastly, using cysteine-scanning mutagenesis other APH residues and Q8 in the TM helix have been identified as imperative (Greene et al. 2007).

1.3.2.2 TatE

TatE (7 kDa), is a much smaller TatA-like protein that shares 57% similarity with TatA (Sargent et al. 1998). Given it is smaller and is 100 times less abundant than TatA, it was thought TatE had no real function in the Tat translocon (Jack et al. 2001). More recently however, it was shown TatE could substitute TatA (Baglieri et al. 2012) and that TatE was a regular constituent of the Tat translocase (Eimer et al. 2015). The role of TatE is much more active than thought, TatE has been shown to interact with the Tat signal peptide and even partially prevent premature cleavage of TorA signal peptide (Eimer et al. 2018). In fact, TatE homologues have been found in many other bacteria species too, *B. subtilis*

TatAc has 45% amino acid sequence similarity to TatE, suggesting unique TatE-paralogs exist which may bridge the gap between the functions of TatA and TatB.

1.3.2.3 TatB

TatB is functionally different to TatA, its role is to bind the Tat signal peptide and subsequent mature protein. Despite only sharing 20% sequence identity to TatA and being nearly double TatA's size (18.5 kDa), TatB is predicted to have a very similar structure and topology (Hicks et al. 2003). However TatB does have a slightly longer APH (by 12 amino acids) (Zhang, et al. 2014) and a longer unstructured C-terminal (Lee et al. 2002). Only mutations in "the hinge" region and APH cause translocation defects (Barrett et al. 2003). Single N-terminal amino acid substitutions in TatA allow complementation to Δ*tatB* strains (Blaudeck et al. 2005)(Barrett et al. 2007) which perhaps validates the knowledge that TatB came from a TatA gene duplication event, but later specialised. The bifunctionality of TatA mirrors the minimal TatAC systems apparent in Gram-positive bacteria.

1.3.2.4 TatC

TatC is the largest (32 kDa) protein in Tat complex that aids substrate binding (Tarry et al. 2009). TatC is also has the most conserved sequence across bacteria and plants, with most conservation in its cytoplasmic loops (Kneuper et al. 2012), this preservation highlights the central role TatC has in the translocation event.

Its structure is very different to other Tat components as it has 6 TM helices with an N-in C-in topology (Behrendt et al. 2004). The crystallisation of TatC from *Auifex aeolicus* (which shares 40% sequence identity to *E. coli* Tat C) in 2013 gave new insights into the structural basis of each TM domain (Ramasamy et al. 2013). TM1 is perpendicular to the membrane whilst TM2 is at a tilted angle until a conserved proline residue creates a kink. TM3 is even more tilted and contacts TM2/4/6, reminiscent of SecE, suggesting a role in overall stabilization. TM4 is very similar to TM2, it has a proline kink and a conserved glycine (Gly170). TM5 is the shortest helix, it contacts TM4 and ends with a highly conserved proline turn. TM6 is also very short, it has a relatively shallow angle and contacts TM3/4/5. Overall, these structures result in a membrane protein that resembles the shape of a baseball glove or a cupped hand. How? Most helices are titled at angles between 20^{o} and 40^{o} to the membrane, TM2/3/5 are strongly kinked and TM5/6 are too short to span the whole membrane. TatC shows very restricted structural flexibility (usually seen in small molecule transporters), so it's unlikely it undergoes a big conformation change upon translocation (Rollauer et al. 2013).

A notable surface feature of TatC is the conserved side chain, Glu165 in *A. aeolicus* or Glu170 in *E. coli* (Rollauer et al. 2013)(Buchanan et al. 2002). Its intriguing location in the interior of the bilayer perturbs membrane bilayer structure and indeed a recent study identified this residue as part of or close to the binding pocket for Tat signal peptides

(Blümmel et al. 2017). Other important residues are localised to the cytoplasmic N-region and its first cytoplasmic loop (Holzapfel et al. 2007)(Kneuper et al. 2012).

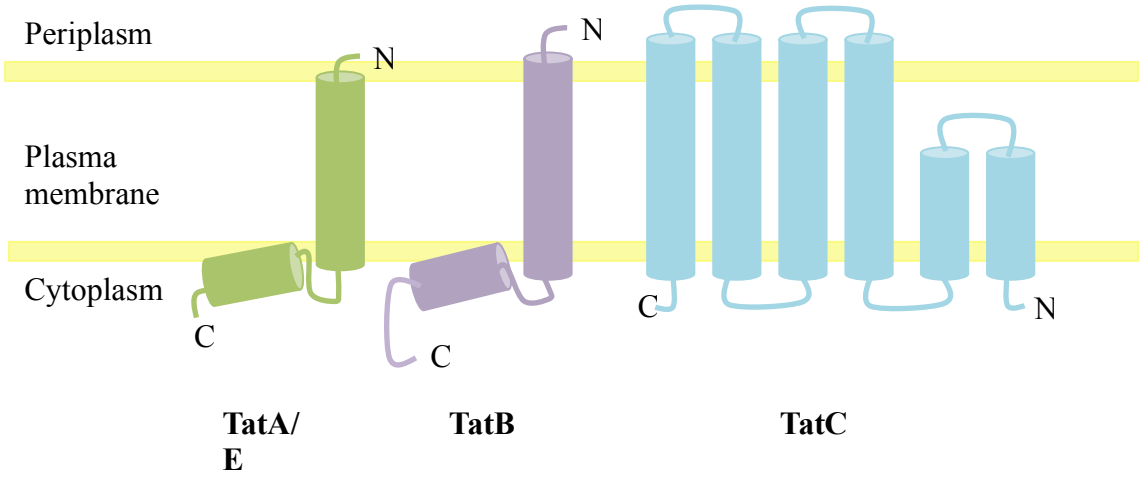


Figure 6. **Schematic of *E. coli* TatABC components in the plasma membrane**

Three essential components form the Gram-negative Tat complex, TatA, TatB and TatC. TatA/E and TatB have similar topologies in that they have one TM domain with a short periplasmic N-terminal region, a tilted APH and an unstructured C-terminus on the cytoplasmic side of the plasma membrane. To note, TatB is larger than TatA with a longer C-terminal tail. TatC is significantly bigger as it containing 6 TM's with both C and N terminals residing in the cytoplasm. TM 5 and 6 do not fully span the membrane, which may contribute to its function.

1.3.2.5 TatD

TatD is the final constituent of the *E. coli* Tat system; it's different because it's a soluble cytoplasmic protein (28.9 kDa). Neither mutant or overproduction of TatD has effects on substrate export nor as purified TatD exhibits a magnesium-dependent DNase activity, it's difficult to see its role in the Tat translocase (Wexler et al. 2000).

1.3.3 *Bacillus subtilis* Tat components

B. subtilis is the model organism for Gram-positive bacteria, it's both easy to genetically manipulate and industrially relevant (as it secretes proteins to medium), but little is known about its minimal Tat system. As mentioned in Section 1.3.1 Bacterial Tat genes, *B. subtilis* has only TatA and TatC subunits, like all Gram-positive bacteria (except *Actinomycetes)* (Schaerlaekens et al. 2001). *B. subtilis* TatA is bifunctional by acting as both *E. coli* TatA and TatB. The more homogeneous complexes TatAd-TatCd, TatAy-Cy and TatAc work in parallel with different substrate specificities (Jongbloed et al. 2004).

1.3.3.1 TatAdCd

TatAdCd exists as a complex of approximately 230 kDa, TatAd alone is 6.5kDa (but ~ 270 kDa in a complex) and TatCd alone is 25 kDa (Barnett et al. 2008). TatAd has the same essential structure and regions as *E. coli* TatA (Hu et al. 2010). However, when TatAdCd is expressed in *E. coli*, a TatAd mutant with substitutions in the C-terminal part of its TM dramatically reduces translocation, but the same *E. coli* TatA mutation only leads to a drop in translocation activity (Barnett et al. 2011). The same study showed mutagenesis of the extreme N-terminus (including the short periplasmic domain and the N-terminal half of the TM span) of TatAd blocked translocation in Δ*tatB* cells. Clearly then, these are important to the ability of TatAd to substitute for TatB in *E. coli* and hence, might be linked to its bi-functionality. Structural data has also changed the way TatA has been perceived. It was thought TatA formed the pore that proteins went through. However, EM images show TatAd complexes too small and too homogeneous for Tat substrates (Beck et al. 2009), translocation mechanisms are discussed in the next chapter.

TatCd both has an affinity for and stabilises TatAd, perhaps the residues responsible for interaction are in the loops flanking the fifth domain (Schreiber et al. 2006). TatCd is also very similar to *E. coli* TatC with 6 TM helices which are slightly tilted with respect to the bilayer normal (Nolandt et al. 2009).

1.3.3.2 TatAyCy

TatAyCy is 200 kDa, TatAy alone is 6 kDa, a TatAy complex is about 200 kDa and TatCy alone is 28 kDa (Goosens et al. 2014).

TatAy (like TatAd) has a similar structure to *E. coli* TatA, for example the conserved hinge region of TatAy is as important in TatAd and *E. coli* TatA (Barnett et al. 2011). However, only in Gram-positive bacteria, a conserved residue (P2) exists in the N-terminal extracytoplasmic region. Indeed, the equivalent residue (P8A) in TatAd was shown to be particularly important for the bifunctionality of TatAd (Barnett et al. 2011). In TatAy, P2 mutants either completely or partially stop protein export (Goosens et al. 2015). Moreover, a different study with the same P2A mutant showed via EM and advanced atomic force microscopy that when mutated large fibrils accumulated which are chains of numerous complexes (Patel et al. 2014). The combined data suggest the mutant is impaired in pore formation, not docking complex formation.

TatCy is predicted to have 6 TM helices like other TatC's discussed, but the extreme C-terminus of TatCy is extended by 5 amino acids compared to TatCd, these residues are important for YwbN export (Eijlander et al. 2009). Indeed, the N-terminus and the first cytoplasmic loop and the C-terminal tail of TatC, TatCd and TatCy are important for

translocation, but intriguingly, the importance of the conserved residues varies and the effects of amino acid substitutions differ depending on the substrate (Eijlander et al. 2009).

1.3.3.3 TatAc

TatAc is the third TatA like protein in *B. subtilis*, it's constitutively expressed and until recently thought to be redundant as Δ*tatAc* cells showed no phenotype. However, more recently the protein has been shown to be active. When expressed in *E. coli*, TatAc is able to make complexes with TatCd and TatAy to form TatAcCd and TatAcAy which can export endogenous Tat substrates such as AmiA, AmiC and TorA (Monteferrante et al. 2012). TatAc is also capable of substituting TatA and TatB in *E. coli* (Beck et al. 2013).

E. coli	Size of gene product (kDa)	Size of complex (kDa)	Reference
TatA	11	50-500	(Oates et al. 2005)
TatB	18	100	(Orriss et al. 2007)
TatC	30	220	(Orriss et al. 2007)
TatE	7	50-110	(Baglieri et al. 2012)
TatABC		440-580	(Behrendt et al. 2007)
B. subtilis	Size of gene product (kDa)	Size of complex (kDa)	Reference
TatAd	7	200*	(Barnett et al. 2008)
TatCd	27	100	(Monteferrante et al. 2012)
TatAy	6	200*	(Barnett et al. 2009)
TatCy	28	66	(Monteferrante et al. 2012)
TatAc	6	100	(Monteferrante et al. 2012)
TatAdCd		230	(Monteferrante et al. 2012)
TatAyCy		200	(Monteferrante et al. 2012)

Table 1. **Tat components and complexes sizes of *E. coli* and *B. subtilis.***

The predicted protein sizes of *E. coli* and *B. subtilis* Tat components according to gene sizes. Complex sizes were determined mostly by blue-native gel electrophoresis, except those determined by gel filtration chromatography, indicated with a *

1.3.4 The Tat complex

The Tat system transports folded proteins; its largest native substrates are heterodimeric formate dehydrogenases of nearly 150 kDa (Berks et al. 2000). However, Tat components themselves are relatively small so the translocase must coalesce with multiple copies of each individual subunit. Live fluorescence microscopy shows the TatABC system assembling on demand to newly synthesised Tat substrates, reorganising individual subunits in an uniform manner in the plasma membrane (Rose et al. 2013). Two complexes exist, the TatBC complex and the TatA complex. It is widely accepted that TatBC forms a receptor complex that binds Tat substrates initially at the membrane. Substrate binding triggers PMF-dependent TatA recruitment and oligomerisation to form an active translocase, suggesting TatA has a role in the translocation event (Berks et al. 2014). The *in vitro* cycle is slow, starting from substrate bound TatBC, the translocation event takes 1-3 mins (Celedon & Cline 2012). Disassembly of TatA from TatBC requires substrate export (Alcock et al. 2013) and the mechanism of the translocation is discussed in the next chapter.

1.3.4.1 The TatBC complex

TatBC interacts in a 1:1 stoichiometry (Bolhuis et al. 2001) and several TatBC complexes bind the substrate(s) (Behrendt & Brüser 2014). Low resolution EM structures of the TatBC complex revealed a hemispherical morphology, with an internal cavity which could accommodate the signal peptide (Tarry et al. 2009). 7 copies of TatBC are able to fit into the 11-17nm reconstruction, however when a Tat substrate was also expressed, only 2 substrates bound at a single time (not 7). This at least supports that more than one substrate can bind at once (Ma & Cline 2010). Though the question still remains, how does the Tat

signal peptide and protein interact with the TatBC complex?

Cross-linking studies show TatC is the primary interaction site of the signal peptide (Gérard & Cline 2006)(Alami et al. 2003)(Kreutzenbeck et al. 2007). More specially, the N-region (containing the RR motif) of the signal peptide binds to residues in the TatC *cis* 1 domain and *cis* 2 loop (Zoufaly et al. 2012). The signal peptide is inserted deep into the TatC (Gérard & Cline 2006) by perhaps adopting a hairpin loop confirmation (Ramasamy et al 2013). In which case, the signal peptide could extend to expose its C-region signal peptidase site at the periplasmic side of the membrane (Fröbel et al. 2012) (Blümmel et al. 2015). However, this model does not account for TatB's role.

Cross linking experiments have also shown the signal peptide H domain can interact with the TatB TM (Alami et al. 2003) and Tat precursors can bind to TatB at multiple sites early on (Maurer et al. 2010). Although high resolution structural information is lacking regarding TatB, NMR data can confirm its extended C-terminal is relatively flexible and able to facilitate substrate binding (Zhang et al. 2014). Thereafter, each TatC monomer has 2 contact sites for TatB. One is located around the TM5 of TatC (Kneuper et al. 2012), notably TM5/6 of TatC are shorter and therefore able to accommodate the TM of TatB, cross-linking data supports this (Rollauer et al. 2012). Whilst the other binding site is positioned between TM4 and TM2 of TatC (Blümmel et al. 2015), both these locations are important for TatBC oligomeric structure, discussed shortly.

Of interest, other than binding substrates and binding to TatC, TatB could also have a role in preventing premature signal peptide cleavage (Fröbel et al. 2012). When TatB was absent the signal peptide of untransported substrate was cleaved by the *trans* facing signal peptidase.

The exact arrangement of the functioning TatBC structure is still unclear, but photo and disulphide cross-linking studies have provided some insight. As mentioned previously, TatB can cross-link to TM2/4 and TM5 of TatC (Blümmel et al. 2015), they interpreted this as each TatB TM bridges TM2/4 of one TatC and TM5 of another TatC, supported by other data, see Figure 7 (Maurer et al. 2010)(Lee et al. 2006). This is intriguing as it places the TM of TatB close to where TatA initiates translocase oligomer assembly, supporting the r ole of TatB as a regulatory surrogate of TatA, especially due to their high level of homology (Cline 2015).

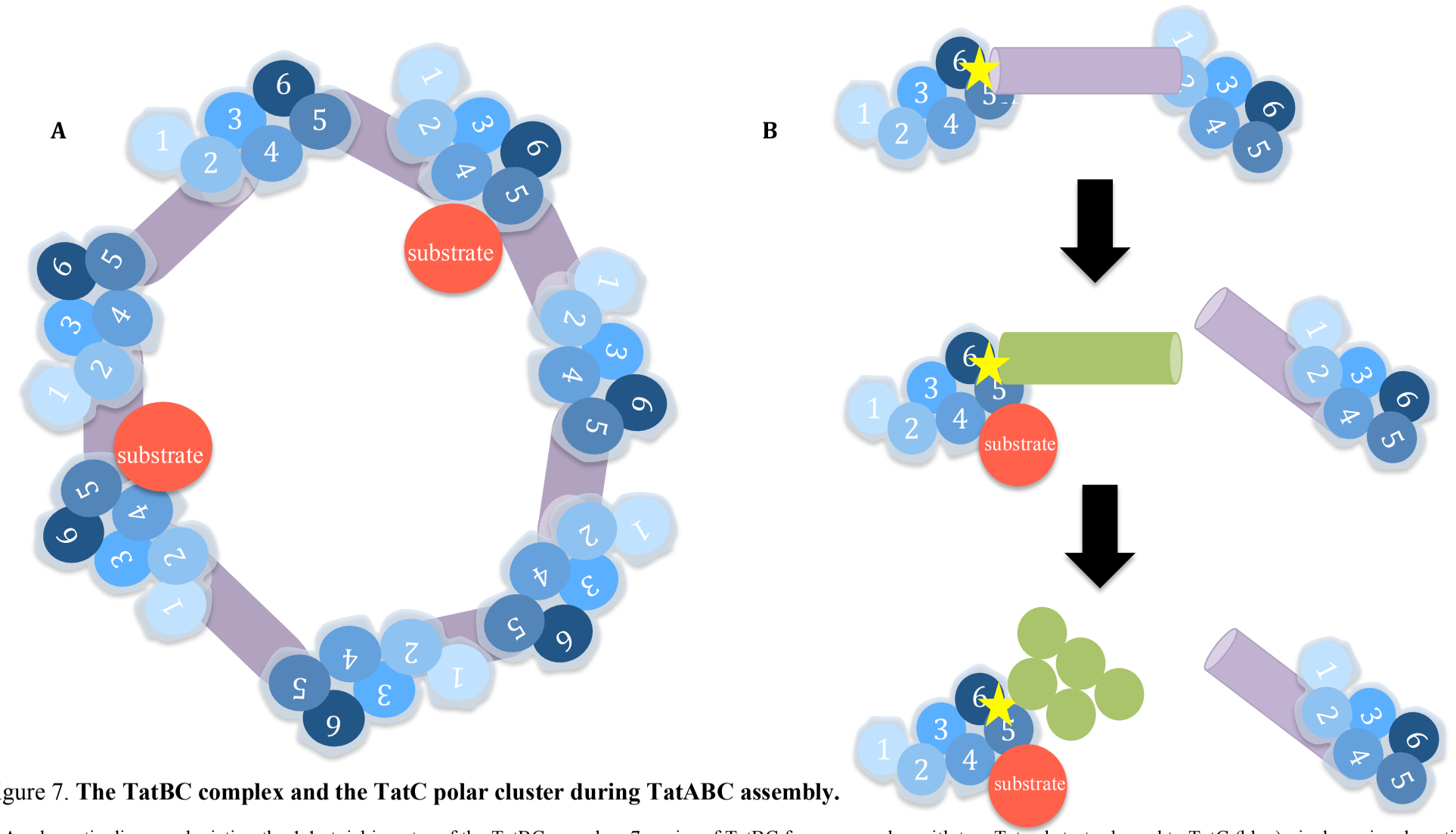


Figure 7. **The TatBC complex and the TatC polar cluster during TatABC assembly.**

A. A schematic diagram depicting the 1:1 stoichiometry of the TatBC complex. 7 copies of TatBC form a complex with two Tat substrates bound to TatC (blue) via deep signal peptide insertion. Each TatC has 6 TMH's (numbered) and is bridged by TatB (purple) via TM2/4 and TM5 of TatC.

B. TatC has a polar cluster of amino acids at TM5/6 (yellow star), without a bound substrate this site is occupied by TatB. Substrate binding brings on a position switch; TatB vacates the site for TatA to bind. Thereafter, the nucleation of TatA (green) promoters starts to form the translocation unit. Figure based on Rollauer *et al* 2012.

1.3.4.2 TatC's 'polar cluster'

More recently, coevolution analysis predicted a location for TatA and TatB binding along TM5/6 of TatC, indicating a 'polar cluster' of amino acids on TatC likely form contact with a polar side chain of TatA or TatB. Both TatA and TatB can occupy the same TatC TM5/6 site at different stages of transport (Alcock et al. 2016). Disulphide cross-linking has confirmed the 'polar cluster' interactions (Habersetzer et al. 2017). The paper further demonstrates that upon substrate binding, TatA cross linking to the TM6 site is reduced, TatB cross linking at the polar site is reduced and TatA binding at the polar site is increased. This position switch is probably bought on by signal peptide binding deep into the complex; the polar cluster site is adjacent to the docking site of the signal peptide (Eimer et al. 2015) and this could cause the conformational rearrangements. Signal peptide binding also altered TatC's resting state arrangement of head-to-tail to tail-to-tail. This opens up the complex and allows TatA access to the vacated polar cluster site. Binding of a TatA molecule to this site places it adjacent to the concave face of TatC where the nucleation of TatA has been suggested (Rollauer et al. 2012). See figure 7.

1.3.4.2 The TatA complex

TatA is the subunit that facilitates the translocation event, either by forming a pore by which the substrate can go through (Gohlke et al. 2005) or by weakening the lipid bilayer (Brüser & Sanders 2003).

While TatA is not obligatory for TatBC formation or substrate docking (Orriss et al. 2007), trace amounts of TatA are observed in association with the TatBC complex which may the nucleation points for other TatA oligomers to join and form the translocase (Aldridge et al.

2014). TatA is expressed 20 times more than TatBC (Sargent et al. 2001) and if TatB is overexpressed compared to TatAC, Tat is inactivated (Sargent et al. 1999) perhaps because it is blocking the 'polar cluster' on TM5 of TatC.

The exact concentration of TatA required to catalyse membrane export in thylakoids was calculated in the 100 nM range, but of more interest, TatA's kinetics suggests the Tat translocase operates as an allosteric enzyme complex (Hauer et al. 2013), this fits with the 'polar cluster' position switch theory (Habersetzer et al. 2017). However, the exact number of TatA subunits that oligomerise within the complex is more complicated. TatA complexes in *E. coli* demonstrate a high heterogeneity as TatA complexes vary in size from 100 kDa to 500 kDa as identified by Blue Native PAGE (Oates et al. 2005). These complexes can be resolved at 34 kDa intervals, supporting an idea of modular formation of 3 or 4 complexes at a time. But after substrate export, the transient complex disassembles, so it's difficult to capture the event. One study controlled oligomer size by detergent ratio and solution NMR indicated self oligomerisation is through TatA's TMH (Rodriguez et al. 2013). Multi-experimental techniques supports this (White et al. 2010)(Zhang, Wang, et al. 2014) whilst the mechanics of how TatA translocates protein is still debated.

1.4 Tat's mechanism of translocation

Although the primary role of each Tat component is understood, the TatBC complex binds Tat substrate, which triggers TatA oligomerisation and mediates the translocation event, the mechanistic process is still not fully resolved. The initial and well accepted model, the 'Translocation pore model' theorises that TatA forms a channel/pore which proteins translocate through (Gohlke et al. 2005). However, more recent data has cast doubt on this idea (Baglieri et al. 2012) and an alternative model is now more supported, the 'Membrane-destabilisation model' (Brüser & Sanders 2003). The latter was a hypothesis with little evidence until very recently, Hou et al proved the TatA destabilises the membrane upon Tat substrate binding (Hou et al. 2018).

1.4.1 Translocation pore model

Low-resolution images of TatA revealed a pore like complex of varying diamaters (8.5 – 13 nm) that would explain the significant heterogeneity seen on BN-PAGE gels and more importantly, accommodate Tat's varying substrate sizes (Gohlke et al. 2005)(Sargent et al. 2006). To prevent ion leakage, a lid-type feature was also identified on the inner side of the cytoplasmic membrane, as if the transporter were a trap door. Moreover, the flexibility implied by the conserved glycine hinge in TatA suggested that such a rearrangement was plausible (Gouffi et al. 2004). Perhaps, TatA's APH swings down (into a hairpin) to align with its TMS so the polar residues on the APH would constitute the inner lining of a pore and allow hydrophilic cargo passage (Walther et al. 2013). In this scenario, after substrate docking onto TatBC, TatA protomers are recruited to suit the size of the substrate in an oligomeric ring format (Dabney-Smith et al. 2006)(Chan et al. 2007).

Using TatAd, NMR data complemented this theory by highlighting the importance of its hinge region and its topology (Walther et al. 2010). The group enhanced the model by uncovering a 'charge zipper mechanism' (Walther et al. 2013) Again based on TatAd, sequence charges suggested each TatA protein could be 'zipped together' by 7 salt bridges; the length of the APH hairpin, which matched the lipid bilayer thickness. Indeed, salt dependence for the translocase was reported (Van Der Ploeg et al. 2011), see Figure 8.

However, now the 'translocation pore model' is less attractive for a multitude of reasons. The insertion of the APH into the bilayer was not found in other topological studies (Koch et al. 2012)(Aldridge et al. 2012) and solution NMR indicate the APH fans outwards from a central TMH core (Rodriguez et al. 2013). Also, low-resolution EM structures of a TatA homolog, TatE precludes its capability to possess a pore forming role as it is too small (Baglieri et al. 2012). Intriguingly, the unique translocase must therefore function otherwise.

1.4.2 Membrane-destablisation model

A second model proposes TatA complexes assemble and weaken/destabilise the lipid bilayer (Brüser & Sanders 2003) and unlike the previous model, more recent topologies of TatA support this idea (Rollauer et al. 2012)(Rodriguez et al. 2013). Through molecular dynamic stimulations, Rodriguez et al pinpointed the destabilisation to the fact that TatA has a short TM domain. Fast forward five years and the theory was verified and enhanced with more information in *E. coli in vivo* (Hou et al. 2018).

More specifically, in this model it's the N-terminal hydrophobic anchor of TatA that alone destabilises the membrane; its short length (as predicted) restricts the membrane thickness

to its own length, only 12 amino acids. However, this destabilisation only happens when a Tat substrate is bound. At resting state, when no substrate is bound, TatA's APH immerses itself into the membrane, elongating the membrane. As a consequence of substrate association, TatA re-orientates its APH (Lys24-Met43 and Asp 45-Lys49) outside of the membrane. TatA's rearrangement has been identified previously in thylakoids, its conformational transition of APH upon substrate induction is very similar which supports this data (Aldridge et al. 2012).

The 'switch' from one state to the other is most likely also a consequence of the interactions between the substrate and TatA's APH. TatA does not recognise the synonymous 'RR' motif, but TatA's APH does interact with the C-domain of the signal peptide (Taubert et al. 2015). Indeed, and in thylakoids, the APH and TatA have been seen interacting too (Pal et al. 2013) and with *E. coli* evidence also demonstrates mature protein interacting with TatA (Alami et al. 2003)(Maurer et al. 2010)(Taubert & Brüser 2014) .

As a whole, this model cooperates with data regarding the TatBC complex (discussed above) and proposes both TatA and TatBC complexes work together synergistically on distinct regions of the transported Tat substrate.

The Tat translocase is distinguished from other protein translocation systems (like Sec) due to the fact it exports fully folded proteins across the lipid bilayer. However other qualities also make the pathway unique such as its sole use of PMF for energy and its ability to 'proof read' substrates to ensure they're correctly folded. These assets make Tat exploitation in the biotechnology industry appealing and hence will be discussed in further detail below.

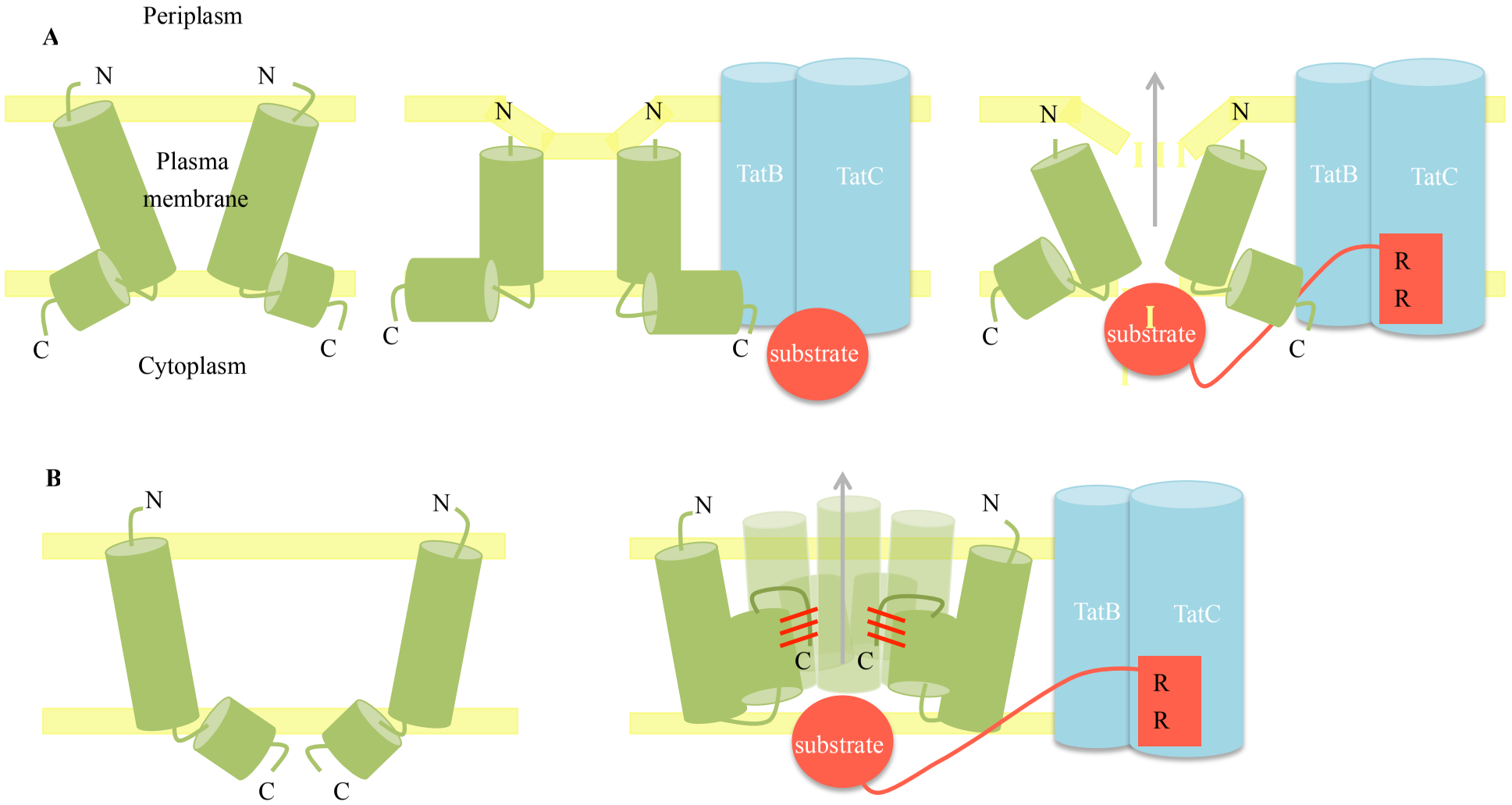


Figure 8. **The membrane destabilisation and translocation pore models.**

A. The TM of TatA (green) is very short, at resting state TatA's APH is tilted to elongate the complex and stabilise the membrane. When substrate binds Tat (blue), the APH's N-terminal anchor switches position and immerses into the membrane, shortening the complex, causing membrane destabilisation so the substı (red) can translocate. **B.** The APH of TatA is flexible, upon substrate binding it can fold into a hairpin stabilised by a cluster of intramolecular salt bridges (lines). TatA's together form a pore with polar residues at the APH, which constitute the inner lining of the pore and hence allow hydrophilic substrate passage. Adapted from Patel *et al* 2014.

1.5 The Unique Tat pathway

The Tat translocase is distinguished from other protein translocation systems (like Sec) due to the fact it exports fully folded proteins across the lipid bilayer. However other qualities also make the pathway unique such as its sole use of PMF for energy and its ability to 'proof read' substrates to ensure they're correctly folded. These assets make Tat exploitation in the biotechnology industry appealing and will be discussed in below.

1.5.1 Energy requirements

A hallmark of the Tat system is that it is powered solely by the PMF and not nucleotide hydrolysis (ATP or GTP) like most other protein transportation machinery. PMF consists of a pH gradient (ΔpH) and the electric potential (Δψ) at the membrane (Mould et al. 1991). Early studies into the energetics of Tat were carried out in the plant thylakoid system, in the presence of light and ΔpH (but in the absence of nucleotides) the oxygen evolving Tat substrate was still exported (Mori & Cline 2002). In addition a phage shock protein PspA, involved in maintaining PMF, increased Tat translocation in bacteria (DeLisa et al. 2004). However, other studies in the green alga *Chlamydomonas reinhardtii* in vivo have shown the system can transport proteins without a thylakoidal ΔpH (Finazzi et al. 2003). With the knowledge that the Tat pathway can use both ΔpH and Δψ equally, this can be explained (Braun et al. 2007). As a consequence of this equivalency, the Theg group have suggested an antiporter mechanism where a coupling of H^+ flow and protein transport has been suggested (Musser & Theg 2000). It's interesting to note the group also calculated the counterflow of protons necessary for Tat protein export, it's in the region of 7.9×10^4 protons per molecule (Alder & Theg 2003). This is equivalent to 10,000 ATP molecules, 3% of the energy produced by chloroplast, so it is a considerable cost to the cell.

With regards to individual steps of the mechanism, the PMF is not required for protein targeting or protein binding to TatBC, but is required for more advanced binding stages and the oligomerisation of TatA (Gérard & Cline 2007) (Alami et al. 2003). For thylakoids, the ΔpH component of PMF could protonate TatA at Glu10, making it energetically feasible to move up in the membrane to its docking site on TM4 and hydrogen-bond with TatC (Gln-234) (Aldridge et al. 2014). It's more unclear how the assembly of TatA in *E. coli* is facilitated by PMF as it's been shown largely, if not entirely, the transport energy driving force is from Δψ alone (Bageshwar & Musser 2007). Indeed, the *in vitro* study concluded two distinct Δψ steps allowed transport, one Δψ of relatively high magnitude that may be short-lived and a second Δψ requiring step minimally requires a long duration, but relatively low magnitude Δψ. If Δψ was increased, so was transport speed. This begs the question, how does the mechanism work and why is it different to thylakoid data? Perhaps these inverted membrane vesicles (IMV's) (Alami et al. 2003) are missing something as the nature of Δψ and protein export by Tat (not least in *E. coli)* remains unclear. A reasonable scenario is that movement of certain charged regions within Tat proteins could be induced by a Δψ, but more research is required in this area.

1.5.2 Proofreading and Quality control

Another feature of the Tat pathway is its quality control, proofreading ability. Tat substrates only can exit the cytosol once they are fully folded and (if necessary) contain their allocated cofactor. This complicated task requires the aid of specific chaperones known as Redox Enzyme Maturation Proteins (REMPs) (Turner et al. 2004). An example Tat substrate with a REMP is the oxidoreductase trimethylamine-*N*-oxide (TMAO) reductase (TorA); it is encoded by the *torCAD* operon. *torA* encodes the TMAO reductase, *torC* encodes its haem-

binding quinol oxidase and *torD* is its REMP. REMPs often dimerise and bind to their cognate partner via the C-domains of its signal peptide (Tranier et al. 2002). The role of TorD, like most specific chaperones, is to block the AXA cleavage site (to prevent premature protein degradation), coordinate cofactor assembly and foresee other maturation steps such as membrane targeting and interaction (Chan et al. 2009). This coordination occurs until the apo-pre-protein interacts with Tat machinery. As Tat substrates are especially complex, to ensure malfunctioning proteins are not exported, the Tat complex rejects and sometimes even degrades the protein within the cytosol (Halbig et al. 1999)(Delisa et al. 2003)(Matos et al. 2008). To note, the thylakoid Tat system seems to have a less stringent 'proofreading' system as unfolded proteins are imported (Hynds et al. 1998). The quality control exhibited by the *E. coli* Tat pathway is highly unusual; hence different approaches have investigated when and how it happens.

Cofactor insertion has been explored in Tat proofreading. Native *E. coli* Tat substrates NrfC and NapG were mutated to prevent their central cofactor FeS binding, indeed this alteration blocked export (Matos et al. 2008). The *B. subtilis* Rieske iron-sulphur substrate QcrA was mutated too, to either stop cofactor binding or disulphide formation (Goosens et al. 2014). A proofreading hierarchy was uncovered; mutant's defective in disulphide bonding were quickly degraded, whereas those defective in cofactor binding accumulated in the cytoplasm and membrane. Two heme-binding proteins have also been tested, cytochrome C required heme insertion for export (Sanders et al. 2001) and another heme binding protein, the synthetic BT6 maquette showed interesting results. The latter binds 2 hemes and is tested with a TorA signal peptide. Histidines were mutated into alanines to prevent heme binding, if no hemes bound, the protein was not exported, one heme bound

allowed some protein export, whereas if two hemes bound, good export was seen (Sutherland et al. 2018). These findings suggest Tat can sense the protein conformational stability.

Conformational stability has been studied *in vivo* and *in vitro* using non-native Tat substrates (PhoA, scFv and a Fab fragment). These proteins only exported in oxidising conditions (Cox strains) when their disulphide bonds could form (Delisa et al. 2003). However in another study some proteins (hGH, scFv and IFN) without their disulphides formed were exported to the periplasm (Alanen et al. 2015). hGH at least can form a near native state in the absence of its two disulphide bonds. This resembles the protein CueO, it can export without its cofactor copper, but UV CD shows CueO without copper is nearly structurally identical to CueO with copper (Stolle et al. 2016).

Several studies in both bacteria and plants have used varying lengths of FG repeats from the yeast nuclear pore protein Nsp1p; they intrinsically lack structure and are hydrophilic. Fused to a Tat signal peptide, early export results demonstrated that with increasing protein length, translocation efficiency decreased, 100-120 amino acids was tolerated but a short hydrophobic stretch stopped export (Cline & McCaffery 2007)(Richter et al. 2007). Unstructured linkers have also been placed between the signal peptide and the N-terminus of a mature Tat substrate, surprisingly an unstructured linker length of 110 amino acids was exported (Lindenstrauß & Brüser 2009). However, the degree of instability and hydrophobicity the Tat system will tolerate had never been investigated with a well characterised test protein.

A study used scFv mutants (which were structurally defined) to identify what Tat

recognised as ‘unfolded’ and rejected for export (Jones et al. 2016). Tat tolerated significant changes in hydrophobicity and charge but did not export the scFv with an unstructured tail or without disulphide bond formation via CyDisCo. CyDisCo comprises of yeast mitochondrial thiol oxidase, Erv1p and human protein disulfide isomerase (PDI) that confer the ability to catalyse cytoplasmic disulphide bonds. In conclusion therefore the study indicated proofreading has evolved to sense conformational flexibility and detect even very transiently exposed internal regions.

Although it’s still unclear what the Tat complex rejects as misfolded, the major question arises, how does the Tat complex reject the proteins? Tat proteins, misfolded or not, both interact with the Tat translocase, for example the native Tat substrate PhoA has been co-purified with TatBC (Richter & Brüser 2005). This gave rise to the idea Tat does not innately have an inbuilt ‘proofreading’ mechanism, but an efficient degradation system clears the Tat translocase. However, site specific photo cross-linking revealed perturbed interaction between the signal peptide and the misfolded protein as if the substrate was not correctly inserted into the binding socket, although physical interaction took place (Panahandeh et al. 2008).

Other approaches have looked into the Tat translocon too. Quality Control Suppression mutants (QCS) were used to screen mutations within the TatABC which gave rise to less stringent proofreading (Rocco et al. 2012). Synthesised using 3-helix-bundles with various conformational flexibilities fused to TEM-1 β-lactamase, key residues were identified. The majority of the QCS mutations were confined to the unstructured or loop regions of TatABC, their existence proves some proofreading is undertaken by the Tat translocon.

1.5.3 The Biotechnology Industry

The ability of Tat to export fully folded proteins to the periplasm and proofread they are correctly folded is advantageous in the Biotechnology industry. High value therapeutic proteins produced via Tat can be exported and processed to their mature state with high homogeneity.

E. coli is already used as a host organism for protein production, owing to its rapid growth, high yield, ease of scale up and cost-effectiveness, around 30% of biotherapeutics are synthesised in bacteria (Overton 2014). The standard location of recombinant protein accumulation is in the cytoplasm, however this is not ideal due to its reducing environment, high levels of proteases and high numbers of contaminating proteins. An alternative method used in industry involves the Sec system (described in Section 1.2), protein is exported to the oxidising periplasm with an N-terminal signal peptide. This alternative approach decreases chances of proteolysis and eases downstream purification as the periplasm only contains 4% of the proteome (Pooley et al. 1996). Usually, proteins are released from the periplasm by osmotic shock, similar to Section 2.4.2 (French et al. 1996). The Sec system, however, does not export folded proteins and it cannot export proteins that fold rapidly in the cytoplasm, such as GFP.

TorA- GFP was expressed in *E. coli* and found to export via the Tat system, as proof of concept that Tat can export proteins to the periplasm Sec cannot (Barrett et al. 2003). Biotherapeutics can too be exported via the Tat pathways, especially with the aid of CyDisCo which catalyses disulphide bond formation in the cytoplasm (Matos et al. 2014)(Alanen et al. 2015). More recently, Tat was overexpressed from the chromosome

(with a pTac promoter) to engineer TatExpress cells during fermentation, a 5 fold increase of hGH was found in the periplasm compared to wildtype cells (Browning et al. 2017).

The 'gold standard' of bacterial protein production, protein secretion to the media, was successfully achieved when Tat *B. subtilis* TatAdCd was expressed in Δ*tatABCDE* strains (Albiniak et al. 2013). In batch fermentation ~90% of GFP was present in the medium as mature protein after 16 hours of fermentation. The cells were otherwise intact and export levels were found to be about 0.35 $g.L^{-1}$. This release resembles both mammalian and yeast systems (Kim et al. 2012)(Mattanovich et al. 2012) but without the considerable time and cost implications. Current efforts are focused on testing different biotherapeutics and increasing protein yield in Tat.

1.6 Aims of this project

The primary aim of this investigation is to develop and understand a novel means of exporting recombinant proteins to the periplasm of *E. coli* via the Tat translocase. The experiments utilised both the Gram-negative *E. coli* TatABC translocase and two minimal Gram-positive TatAC translocases, TatAdCd and TatAyCy. To achieve this overall aim a series of questions and tests were explored:

1. Which biopharmaceutical recombinant proteins will TatAdCd export?
2. How does the Gram-positive Tat translocase compare to the Gram-negative translocase, in terms of proof-reading ability?
3. Which orientations and conditions does Tat prefer for scFv (and possible diabody) export to periplasm? Can this be scaled up to fermentation?
4. As the *E. coli* and *B. subtilis* Tat systems are so distinctive, can I use a novel means of membrane protein purification (SMA) to purify the smaller Gram-positive Tat complexes?
5. Can CryoEM be used to identify additional structural features of a Gram-positive Tat system?

2. Materials and Methods

2.1 Suppliers of chemicals, reagents and equipment used

All reagents, materials and equipment were obtained from the companies listed below.

Beckman Coulter Inc (USA): TL-100 Ultracentrifuge TLA 100.2 rotor; Avanti J-25, JA10 rotor; DU 730 UV/Vis Spectrophotometer
Bio-Rad (USA): Agarose; Bio-Rad gel doc; Bio-Rad Mini-PROTEAN® Tetra System; Acrylamide (15%, 0.3% bis-acrylamide, 37:5:1); Coomassie Brilliant Blue R-250; Claritytm Western ECL substrate; T1000™ Thermo Cycler PCR machine
Eppendorf: 1.5 mL centrifuge tubes; PCR tubes; Centrifuge 5417R; pipette tips
Eurofins (UK): Oligo's/Primers
Fisher Scientific (UK): Anti-his (C-term) unconjugated antibody; dNTP nucleotide mix; GeneRuler (100 bp and 1 kb); Mircobiological spreaders; Inoculation loops; Nanodrop 2000c
Formedium (UK): PBS
GE healthcare (UK): Superdex 200 10/300GL prep-packed gel filtration column; ECL™ detection reagents; IMAC™ sepharose 6 fastflow; Vivaspin™ protein concentration spin column; PVDF and nitrocellulose membrane
GenPure: UV/UF water purification system for MilliQ H_2O
Gilson: Pipettes, P2, P20, P200, P1000
Grant: Heat block; SUB Aqua 12 Plus water bath
Greiner Bio-one: Universals; Easy load tips;
IBA Lifescienes: Anti-strep HRP conjugate;
Infors (Switzerland): Multitron Pro shaking incubators; Minifors 2 fermenters
Invitrogen (USA): Anti-hexahistidine (C-term) conjugated;
Life technologies (USA): BSA
MSE: Soniprep 150 plus
New England Biolabs (UK): Phusion high fidelity DNA polymerase (2U/ µL) ; Restriction enzymes; NEB® Turbo compotent cells; NEB® BL21 DE3 competent cells; Coloured protein standard broad range(11-245 kDa)
Panasonic: Heated static incubator
Promega (UK): dNTP mix (200µM); Anti-mouse HRP conjugate; Anti-rabbit HRP conjugate
Qiagen (Germany): QIAprep Spin Miniprep Kit; QIAprep Gel Extraction Kit
Roche applied science (UK): Complete™ protease inhibitor tablets; T4 DNA ligase
Sarstedt: Serological pipettes; 15 mL and 50 mL falcon tubes; Petri-dishes 92x16 mm; Filter tips; curvettes
Sigma-Aldrich (UK): Lysozyme; Anti-flag HRP conjugate
Starlab: Starguard comfort gloves;
Thermo Fisher Scientific Inc (USA): Nanodrop 2000c; Megafuge 16R Centrifuge, 75003181 rotor; MaxQ 800 Shaking incubators
VWR: Microcentrifuge

2.2 DNA techniques

2.2.1 Preparation of plasmid DNA

Plasmids were purified using the Qiagen QIAprep Spin Miniprep Kit as per the manufacturer's instructions. Plasmids were eluted in 50 µL elution buffer (10 mM Tris-Cl pH 8.5) with their concentration determined using the Nanodrop.

2.2.2 Amplification of DNA via Polymerase Chain Reaction

PCR's were achieved using Bio-Rad T100^{TM} Thermal Cycler machine. Each reaction contained, 1 µL template DNA (80-100 ng), 0.5 µL Phusion high fidelity DNA polymerase, 1 µL dNTP mix, 1.5 µL DMSO (3%), 10 µL GC buffer (10%), 0.5 µL of forward and reverse primer (0.5 µM) made up to 50µL with autoclaved milliQ H_2O.

An example cycle is given below:

Initial denaturation	98°C	30 sec	
Denaturation	98°C	10 sec	x 25 cycles
Annealing	Primer T_m	30 sec	
Extension	72°C	30 sec	
Final extension	72°C	5 mins	
Pause	4°C	∞	

Primer name	**Sequence 5' → 3'**
TatAdCd study	
preFhuDF	GCATATGAATTCATGAGCGGCTTACCTC
preFhuDR	GGTACGCTGCAGTCACGCTTTACCTCCGATG
scFv export study	
ColinscFvNdeIF	GCTATGCCATATGAGGTGCAGCTGGTGGAGTCTGGGGG
ColinscFvBamHIR	CCGCATGGATCCCTATGCGGCCCATTCAGATCCTCTTCTG
GA_{VH}_R	TTCCCGACTGGAAAGCGGGCAGTG
GA_F	GAAGGAAGGCCGTCAAGGCCACG
SMALP study	
TatE_F_SMALP	CGCCAAGAATTCATGGGTGAAATCTCTATTACCAAACTGCTG
TatAy_F_SMALP	CAATTAGAATTCATGCCAATCGGGCCGGGCAGTTGGC
TatAc_F_SMALP	GGACTAGAATTCATGGAACTGTCCTTCACTAAGATCC
10xHistag+XbaI_R	CTGCTGTCTAGACTACTAATGGTGATGGTGATGGTGATGGTGATGGTG
TatAd_Kpn1_F	CAGTCAGGTACCGAATTCATGTTTTCGAACATC
TatAd_SalI_R	CTGCTGGTCGACTCTAGACTACTAATGGTGATC
OLTatAyCy10H_F	ATACCCGTTTTTTTGGGCTAGCAGGAATTCGGGCTAGCAGGAGGAATT
OLTatAyCy10H_R	CTACTAATGGTGATGGTGATGGTGATGGTGATGGTGATTCGATTGCCCAGAAGACA

Table 2. **Amplification primers**

2.2.4 Agarose gel electrophoresis

Agarose gels were prepared by adding 1% (w/v) agarose into 1 x TAE buffer (0.04 M Tris-acetate, 0.001M EDTA, pH 8.2). Typically 0.5g aragrose and 50 mL 1 x TAE buffer were put into a conical flask and microwaved for 1 min until the agarose had dissolved. The solidified gel was then placed in the tank, submerged in 1 x TAE buffer. DNA samples (including the DNA marker) were mixed with SYBER Green Nucleic Acid Gel stain and 6 x Gel loading buffer. Electrophoresis was carried out at 120V for 45 mins and was visualised using Bio-Rad Gel doc.

2.2.5 Purification of DNA from Agarose gels

Appropriate bands were excised from agarose gels under UV transilluminescence using a scalpel blade. DNA was recovered from the agarose gel using Quigen QIAprep Gel Extraction Kit as per the manufacturer's instructions. DNA was eluted in 30 µL elution buffer (10 mM Tris-Cl pH 8.5) with their concentration determined using the Nanodrop.

2.2.6 Restriction Digests of DNA

Typically, 2 µg DNA was mixed with 5 µL 10 x CutSmart (or appropriate) NEBuffer, 20units (generally 1 µL) of one or two enzymes and the total reaction volume was made to 50 µL with autoclaved MilliQ water. The reaction was then incubated for 1 or 2 hours at 37°C and visualised/purified as described above.

Enzyme name	**Sequence 5' → 3'**
BamHI HF	G-GATCC
EcoRI	G-AATTC
HindIII	A-AGCTT
KpnI	GGTAC-C
NdeI	CA-TATG
NheI HF	G-CTAGC
SacHI HF	GAGCT-C
SalI	G-TCGAC

Table 3. **Restriction enzymes used in this study**

2.2.7 Ligation of DNA fragments into plasmid vector backbone

DNA fragments were cloned into vectors with 1 µL T4 D NA ligase, 1 µL 10 x ligase buffer and an insert to vector concentration ratio of 3:1 (typically to make a 10 µL reaction mix). The ligation incubation was either carried out at 2 hrs RT or 0/N at 4°C before transformation.

2.2.8 Sequencing of plasmid DNA

All sequencing was carried out using GATC sequencing service. 5 µLpurified plasmid (typically 100 ng) was mixed with 5 µL (5 mM) sequencing primer.

Primer name	Sequence 5' → 3'
pTac_F	GAGCGGATAACAATTTCACACAGG
pBAD_F	TATGCCATAGCATTTTTATCC

Table 4. **Sequencing primers used in this study**

2.2.9 Constructs used in this study

Plasmid name	Function	Tag	Reference
pCM233	pLysBAD TatABC	Strep	This study
pCM232	pLysBAD TatAdCd	Strep	(Albiniak et al. 2013)
pKF14	pBAD24 TatAyCy	H10	This study
pKF9	pBAD24 TatE	H10	This study
PKRK40c	pBAD24 Erv1p and hPDI (CyDisCo)	Flag	This study
pHA17	pET23/pTac TorA-scFvO	H6	(Alanen et al. 2014)
pHA15	pET23/pTac TorA-IFN	H6	(Alanen et al. 2014)
pHA14	pET23/pTac TorA-hGH	H6	(Alanen et al. 2014)
pHA23	pET23/pTac TorA-VH	H6	(Alanen et al. 2014)
pHAK13	pET23/pTac TorA-scFvM wild-type + CyDisCo	H6	(Jones et al. 2016)
pAJ15	pET23/pTac TorA-scFvM wild-type	H6	(Jones et al. 2016)
pAJ5	pET23/pTac TorA-3SB (L11Q, Q13K, A88E, L112K, T114E, S116 K) + CyDisCo	H6	(Jones et al. 2016)
pAJ6	pET23/pTac TorA-5Lys (L11 K, Q13K, A88K, L112K, S116 K) + CyDisCo	H6	(Jones et al. 2016)
pAJ7	pET23/pTac TorA-5Glu (L11E, Q13E, A88E, L112E, S116E) + CyDisCo	H6	(Jones et al. 2016)
pAJ8	pET23/pTac TorA-1SB (L112K, T114E + CyDisCo	H6	(Jones et al. 2016)
pAJ9	pET23/pTac TorA-5R > K (R19K, R87K, R150K, R203K, R240K) + CyDisCo	H6	(Jones et al. 2016)
pAJ10	pET23/pTac TorA-7 K > R (K65R, K76R, K162R, K183R, K186R, K235R, K239R) + CyDisCo	H6	(Jones et al. 2016)
pAJ11	pET23/pTac TorA-P172S (P172S) + CyDisCo	H6	(Jones et al. 2016)
pAJ24	pET23/pTac TorA-2SB (A88E, L112K, T114E, S116K) + CyDisCo	H6	(Jones et al. 2016)

pAJ31	pET23/pTac TorA-4NLeu (Q13L, R19L, K65L, K76L) + CyDisCo	H6	(Jones et al. 2016)
pAJ35	pET23/pTac TorA-5Arg (L11R, Q13R, A88R, L112R, S116R) + CyDisCo	H6	(Jones et al. 2016)
pAJ36	pET23/pTac TorA-5NLeu (Q13L, R19L, K65L, K76L, A88L) + CyDisCo	H6	(Jones et al. 2016)
pAJ39	pET23/pTac TorA-6NLeu (Q13L, R19L, K65L, K76L, A88L, S116L) + CyDisCo	H6	(Jones et al. 2016)
CM1	pBAD24 preNrfC	H6	(Matos et al. 2008)
CM11	pBAD 24 preFhuD	H6	This study
pKF15	pEXTII TorA 001VH-VL	H6	This study
pKF16	pEXTII TorA-002 VH-VL (DS)	H6	This study
pKF17	pEXTII TorA-003 VH-VL Short linker	H6	This study
pKF18	pEXTII TorA-004 VH-VL Short linker (DS)	H6	This study
pKF19	pEXTII TorA-005 VL-VH	H6	This study
pKF20	pEXTII TorA-006 VL-VH (DS)	H6	This study
pKF21	pEXTII TorA-007 VL-VH Short linker	H6	This study
pKF22	pEXTII TorA-008 VL-VH Short linker (DS)	H6	This study
pKF23	pEXTII TorA 001VH-VL + CyDisCo	H6	This study
pKF24	pEXTII TorA-002 VH-VL (DS) +CyDisCo	H6	This study
pKF25	pEXTII TorA-003 VH-VL Short linker + CyDisCo	H6	This study
pKF26	pEXTII TorA-004 VH-VL Short linker (DS) + CyDisCo	H6	This study
pKF27	pEXTII TorA-005 VL-VH +CyDisCo	H6	This study
pKF28	pEXTII TorA-006 VL-VH (DS) +CyDisCo	H6	This study
pKF29	pEXTII TorA-007 VL-VH Short linker + CyDisCo	H6	This study
pKF30	pEXTII TorA-008 VL-VH Short linker (DS) +CyDisCo	H6	This study

Table 5. **Constructs used in this study**

Strain	Genotype	Reference/source
MC4100	*Ara^R, F2 arD139 DlacU169 rspL150 relA1 flB5301 deoC1 ptsF25 rbs^R*	(Barrett, Ray, et al. 2003)
Δ*tatABCDE*	MC4100 strain lacking *tatABCDE* genes, *Ara^R*	(M Wexler et al. 2000)
BL21 (DE3) Competent *E. coli*	*fhuA2 [lon] ompT gal (λ DE3) [dcm]* Δ*hsdS* *λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21* Δ*nin5*	NEB®
TatExpress BL21	BL21 carrying a *pTac* promotor upstream of *tatABCD*	(Browning et al. 2017)
NEB® Turbo Competent *E. coli*	F' *proA$^+$B$^+$ lacI^q* Δ*lacZM15 / fhuA2* Δ*(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)*Tet^S *endA1 thi-1* Δ*(hsdS-mcrB)5*	NEB®

Table 6. ***E.coli* strains used in this study**

2.3 Maintenance of *E. coli* cultures

2.3.1 Media and supplements

Most liquid cultures were grown in LB medium, although other media's were used too, see table 7. LBA (LB with the addition of 10 g/L bacto-agar) was used for growth on plates. All were supplemented with appropriate antibiotic. Ampicillin (100 µg/mL) and Kanamycin (50 µg/mL) were dissolved in water whereas Chloramphenicol (35 µg/mL) was dissolved in ethanol.

Media	Components
LB	10 g/L sodium chloride, 10 g/L tryptone, 5 g/L yeast extract
TB	24 g/L yeast extract, 12 g/L tryptone, 4 mL/L glycerol Potassium phosphate buffer: K_2PO_4 12.5 g/100mL and KH_2PO_4 g/100mL
AIM	10 g/L tryptone, 5 g/L yeast extract, 3.3 g/L $(NH_4)2s0_4$, 6.8 g/L KH_2PO_4, 7.1 g/L Na_2HPO_4, 0.5 g/L glucose, 2 g/L lactose, 0.15 g/L $MgS0_4$
PY	48 g/L phytone, 30 g/L yeast extract, 5 g/L NaCl
SM6	10 mL 10X SM6 trace elements, 95 g/L glycerol, 5.2 g/L $(NH_4)SO_4$, 4.4 g/L $NaH_2PO_4.H_2O$, 4.16 g/L citric acid, 4.03 g/L KCl, 1.04 g/L $MgSO_4.7H_2O$, 0.25 g/L $CaCl_2.H_2O$
10X SM6 trace elements	104 g/L citric acid, 10.06 g/L $FeCl_3.6H_2O$, 5.22 g/L $CaCl_2.H_2O$, 2.72 g/L $MnSO_4.7H_2O$, 2.06 g/L $ZnSO_4.4H_2O$, 0.81 g/L $CuSO_4.5H_2O$, 0.42 g/L $CoSO_4.7H_2O$, 0.03 g/L H_3BO_3, 0.02 g/L $Na_2MoO_4.2H_2O$

Table 7. **Media's used and their components**

2.3.2 Transformation of competent *E. coli* cells

The *E. coli* strains listed in table 5 were inoculated in 5 mL LB media and cultured O/N at 37°C, 200rpm. The following morning, 100 µL of the pre-culture was inoculated into 10 mL of fresh LB media and incubated in the same conditions until OD_{600} reached 0.4- 0.5. Cells were harvested via centrifugation at 3000 rpm, 4°C for 10 mins. Supernatant was discarded and cell pellet was re-suspended in 10 mL ice-cold 100 mM $MgCl_2$, then left to incubate on ice for 5 mins. Cells were again harvested as described previously. Supernatant was discarded and cell pellet was re-suspended in 1 mL ice-cold 100 mM $CaCl_2$, then left on ice for 1 hr or in the fridge O/N.

To transform, on ice, 50 µL of competent *E.* coli cells were incubated with 2 µL DNA (100 ng/uL) for 30 mins. The cells were then heat shocked at 42°C in the water bath for 30 secs and immediately placed back on ice for 5 mins. Cells recovered in 200 µL LB media for 1 hr at 37°C, 200 rpm and plated on LBA with appropriate antibiotics. These were placed inverted O/N in the plate incubator, set at 37°C.

2.3.3 Storage of *E. coli* cells

Glycerol stocks of *E. coli* cultures were prepared by mixing 2 parts 50% glycerol with 1 part stationary phase culture, stored in the -80°C freezer.

2.4 Protein production and *E. coli* cell fractionation

2.4.1 Cell culture and induction of plasmids

A 5 mL pre-culture of *E. coli* cells with appropriate plasmid and antibiotic were grown O/N at 37°C, 200 rpm. The following morning, cultures were diluted to OD 0.05 with fresh LB media (or other media if stated) and appropriate antibiotic. Typically, 50 mL culture was used to culture *E. coli* cells in a 250 mL Erlenmeyer flask at 37°C, 200rpm until $0D_{600}$ reached 0.4 – 0.6 (stationary phase). Plasmid(s) were then induced and kept in the same conditions as pre-induction. pYU49 (pET23/pTac) and pEXT22 were typically induced with 0.5 mM IPTG, whereas pBAD24 was typically induced with 0.5 mM arabinose, unless otherwise stated. AIM tests were never induced with IPTG.

2.4.2 *E. coli* fermentation

A 50mL pre-inoculant of PY media (with appropriate antibiotic) were grown for 6 hours at 37°C, 250 rpm. 1 mL of pre-inoculation was transferred to 200 mL of SM6 fermentation media to grow the pre-culture O/N at 30°C, 200 rpm. The volume equivalent to 300 OD/L inoculated the fermenters when timings suited.

The fermenter was clean and autoclaved with 500 mL SM6 media and all probes (pH, temperature, pO) inside. Temperature was set to 30°C and sustained via a chiller. Whilst pH was set to pH 7 and kept constant by 25% sulphuric acid base 25% ammonia; connection lines were too primed. The pO probe was calibrated by maxing the stirrer and

airflow. A cascade was set so when pO_2 reached 30%, the stirrer was activated (800-1600 rpm) and then total flow (1.5 0 2 L/min). When both stirrer and airflow were maxed, culture temperature was dropped to 25°C approximately 15 hrs post inoculation. At specific OD's, different supplements were given or induction occurred:

OD 38-42: 8 ml/L 1M $MgS0_4.7H_2O$

OD 54-58: 5 ml/L 232.8 g/L $NaHPO.H_2O$

OD 66-70: 7 ml/L $NaHPO.H_2O$ and start glycerol feed (80% w/w glycerol at a rate of 0.35% pump capacity

>OD 75: Induce with 9 ml/L of 4. 31 g/L IPTG (made fresh)

Cell fractionation took place pre and post induction at different time points while fermentation growth was monitored.

2.4.3 *E. coli* cell fractionation

Cells were fractionated into periplasm, cytoplasm and membrane fractions using the osmotic/cold shock method as described (Randall & Hardy 1986). Post induction, protein expression occurred for typically 3 hrs, until cells equal to $0D_{600}$ were harvested via centrifugation at 3000 rpm, 4°C for 10 mins. Samples were placed on ice, supernatant was discarded and cell pellet was resuspended in 0.5 mL ice cold Buffer 1 (100 mM Tris-acetate pH 8.2, 500 mM sucrose, 5 mM EDTA). 0.5 mL ice cold MilliQ H_2O was added followed by 40 µL lysozyme (1 mg/mL stock) and incubated on ice for 5 mins. To stabilise the inner membrane, 20 µL $MgCl_2$ (1M stock) was added prior to centrifugation at 14,000 rpm, 4°C for 2 mins. 0.5 mL of supernatant was collected as the periplasmic fraction. Remaining supernatant was discarded and the pellet was washed with 1 mL ice cold Buffer 2 (50mM Tris-acetate pH 8.2, 250 mM sucrose, 10 mM $MgS0_4$) prior to centrifugation at 14,000 rpm,

4°C for 5 mins. All supernatant was discarded and pellet was resuspended in Buffer 3 (50mM Tris acetate pH 8.2, 2.5mM EDTA). Sonication (10 secs on/ 10 secs off x 3) at 8 µM ensured full cell lysis. Ultracentrifugation at 70,000 rpm, 4°C for 30 mins separated the spheroplast. 0.5 mL supernatant was collected as the cytoplasmic fraction and the remaining was discarded. The pellet was resuspended in 0.5 mL Buffer 3 and collected as the membrane fraction. All fractions were stored at -20°C with and without Protein Gel Loading Buffer.

2.4.4 Membrane isolation and solubilisation for Tat purification

For a 0.4L culture, Δ *tat* cells expressing Tat (for 3 hrs) were harvested by centrifugation at 3,000 rpm, 4°C for 30 mins. *E. coli* cells were resuspended in 24 ml pre-chilled Disruption buffer (100 mM Tris-acetate pH 8.2, 500 mM sucrose and 5 mM EDTA) with 1 protease inhibitor tablet added to the suspension. To disrupt the *E. coli* outer membrane, 1 mg/ml lysozyme was added, followed by immediate addition of ice-cold dH_2O and 1 ml of 1M $MgS0_4$. The spheroplasts were harvested by centrifugation at 3,000 rpm, 4°C for 20 mins and resuspended in 10 mL Sonication buffer (50 mM tris-acetate pH 8.2 and 2.5 mM EDTA). Sonication (30 secs on/ 30 secs off x 5) at 8 µM ensured full cell lysis and ultracentrifugation at 70,000 rpm, 4°C for 30 mins separated the spheroplast. The resulting membrane pellet was solubilised in SMA solubilisation buffer (50 mM Tris-HCl pH 8.0, 500 µM NaCl and 10% Glycerol).

2.4.5 Membrane SMA solubilisation

The resulting membrane pellet was solubilised in SMA solubilisation buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl and 10% Glycerol). 2.5% (w/v) SMA polymer was added and

incubated for 2 hrs at RT. SMA polymer was synthesised as described (Lee et al. 2016). The insoluble fraction was removed by centrifugation at 44, 000 rpm, 4°C for 1 hr.

2.5 Protein purification

2.5.1 Nickel column purification

Supernatant from the final step in SMA solubilisation was incubated O/N at 4°C with IMAC Sepharose™ 6 Fast Flow beads in a column. The column was made with charged beads (0.2 M NiCl), glass wool that acted as a filter and was equilibrated with 3 CV SMA solubilisation buffer. The following day, FT was collected and unbound matter was removed with an increasing imidazole gradient (0.5 mM, 20mM, 60mM and 200 mM). Each concentration wash was 2 CV and collected as W. 6x histidine-tagged protein was eluted and collected in 5 x 1 mL fractions using 0.5 M imidazole. The column was washed stripped to remove all residual protein with 2 CV 100mM EDTA and stored in 20% ethanol. The appropriate elution fraction as visualised by SDS-PAGE Coomassie stain in Section 2.6.2 was then taken forward for SEC.

2.5.2 Gel-filtration chromatography

Size exclusion chromatography was used as a final purification step to isolate the Tat proteins. A Superdex200 10/300 GL column with a protein separation range of 10-600 kDa was used with an ÄKTA™ FPLC system (operated by UNICORN® v.4.00 software). The column was equilibrated and ran with 2CV Gel filtration buffer (50mM Tris and 150mM NaCl). The sample was centrifuged at 10,000 rpm, 4°C for 10 mins prior to loading. Column flow was usually 0.5 mL/min and elution's were collected in 0.5 mL fractions. Between runs, the column was washed with MilliQ, sometimes stripped with 1 M NaOH and stored in 20% ethanol according to the manufacturer's guide.

2.6 Protein electrophoresis

2.6.1 SDS poly-acrylamide gel electrophoresis (SDS-PAGE)

Gel electrophoresis was used to resolve protein samples, carried out using vertical gel apparatus from Bio-Rad Mini-PROTEAN® Tetra System, gels were ran and cast according to the manufacturer's instructions. 0.75 mm gels were made from a separation and a stacking gel. Separation gel was composed of 15% acrylamide, 0.3% bis-acrylamide (37:5:1), 375 mM Tris-HCl pH 8.85, 0.1% SDS, 0.1% APS and 0.06% TEMED. Staking gel was composed of 5% acrylamide, 0.0375 % bis-acrylamide, 125 mM Tris-HCL pH 6.8, 0.001% SDS, 0.6% APS and 0.06% TEMED. Before loading, samples were mixed with protein gel loading buffer (125 mM Tris-HCl pH 6.8, 20% glycerol, 4% SDS, 0.02% bromophenol blue and 5% β- mercaptoethanol) and heated for 10 mins at 50°C. Gels were ran at 60 mA for typically 40 mins in protein gel running buffer (25 mM Tris, 192 mM Glycine and 0.1% SDS at pH 8.3).

2.6.2 Coomassie staining

Proteins were visualised on SDS-PAGE by incubation with 50 mL Coomassie stain (10% acetic acid, 40% methanol and 1 g/L Coomassie blue) for 1 hr at RT and rocking. The stain was removed from the background by incubation by D-stain (5% ethanol, 7.5% acetic acid) for up to 8 hours, RT and rocking.

2.7 Protein detection

2.7.1 Western-blotting

A parallel SDS page was also run via electrophoresis for a wet-western blot, which visualised C-terminal hexa-histidine residues. To transfer proteins onto PVDF membrane,

the SDS-PAGE was placed on top of the membrane (post activation with methanol). This was sandwiched between two sheets of Whatman paper and two sponges, presoaked in Western blot transfer buffer (192 mM Glycine, 25 mM Tris and 10 % ethanol) placed in a cassette between electrodes. The tank was filled with Western blot transfer buffer while proteins were transferred at 80 V for 1 hr.

2.7.2 Immunoblotting

The PVDF membrane was blocked overnight at 4°C in a solution containing 2.5% skimmed milk powder in 1X PBS-T, the following was carried out at RT on a shaking table. Membranes were washed for 3 x 5 mins PBS-T before the incubation of primary antibody for 1 hr. Then, the membrane is washed again for 3 x 5 mins PBS-T before the incubation of the appropriate secondary antibody for 1hr. Finally, the membrane was washed 8 x 5 mins before imaging. For more information on antibodies, see Table 6. Immunoreactive bands were detected using the ECL™ kit according to the manufacture's guide and the bands were visualised using BioRad Gel-doc chemiluminescence imager with associated software.

Antibody	**Concentration**	**Source**
Anti-hexahistidine (C-terminal)	1: 8,000	Invitrogen
Anti-mouse HRP conjugate	1: 5,000	Promega
Anti-strep HRP conjugate	1: 32,000	IBA life sciences
Anti-Ay	1: 2,000	Jan Maarten van Dijl
Anti-flag	1:8,000	Sigma
Anti-rabbit HRP conjugate	1: 5,000	Promega

Table 8. **Antibodies used in this study**

2.8 Protein imaging

2.8.1 Negative staining

5 µl of purified TatAyCy-SMALP was pipetted onto a freshly glow-discharged Formvar/carbon coated copper grid (300 mesh), for 1 min. Washing excess sample off the grid was achieved by touching the grid to the convex face of a 8 µl drop of GF buffer for ~10 secs with gentle agitation. Post-washing, excess liquid was removed from the grid by blotting with filter paper. This process of washing and blotting was repeated again. Next, the grid was washed for ~10 secs in an 8 µl drop of uranyl acetate stain (1 or 2% (w/v)) before blotting and staining in a fresh drop of uranyl acetate for 20 secs. The grid was blotted for a final time and then left to air dry for ~10 min before insertion into the electron microscope. This protocol is an adaptation of a previously published method (Rubinstein, 2007) using a Parafilm® drop support and Whatman® paper for blotting.

Negative stain data was collected on a JEOL 2100Plus at nominal magnification of 60 000X with a pixel size of 1.84 Å.

2.8.2 Cryogenic plunging of biological samples

Cryogenic samples of TatAyCy-SMALP were prepared using the Leica automatic freeze plunger: EM GP 2. 3 µl of purified TatAyCy-SMALP was pipetted onto a freshly glow-discharged Quantifoil grid (1.2 µm dia x 1.3 µm 300 mesh) and blotted for 6 seconds in an environment of 4 ^{o}C and 100% humidity.

2.9 Transmission electron microscopy (TEM)

2.9.1 JEOL 2100Plus

200 kV JEOL transmission electron microscope with a LaB_6 electron source. Fitted with a Gatan OneView IS camera. The microscope was controlled by JEOL TEM Center (ver.1.7.11.1834) and the OneView camera was operated using Gatan Digital Micrograph software (ver.3.22.1461.0).

2.9.2 JEOL 2200FS

200 kV JEOL transmission electron microscope with a field emission gun electron source. Fitted with an in-column energy filter and Gatan K2 Summit direct electron detector. The microscope was controlled by JEOL SightX software (ver.1.2.3.538) and the K2 camera was operated using Gatan DigitalMicrograph software (ver.3.21.1374.0).

2.10 Data processing

Negative stain data was processed entirely using EMAN2 software (ver.2.2).

3. TatAdCd export and proofreading study

3.1 Introduction

The Gram-positive Tat differs from that of Gram-negative translocase due to the lack of TatB, it's a minimal system where TatA acts bi-functionally in a discrete TatAC complex (Barnett et al. 2008). *Bacillus subtilis* has three variants of the TatA protein, Ad, Ay and Ac, and two variants of the TatC protein, Cd and Cy. Two TatAC systems work in parallel, TatAdCd and TatAyCy, each with different substrate specificities. The only native substrate of TatAdCd is PhoD, a protein with phosphodiesterase and alkaline phosphatase activity which is only expressed during phosphate starvation conditions (Jongbloed et al. 2000). To investigate the substrate selectivity of TatAdCd, the translocase was expressed in mutant Δ*tatABCDE E. coli* where partial complementation was evident (Barnett et al. 2008). The discrete system expressed in *tat* null *E. coli* was also able to export the heterologous protein GFP with a TorA signal peptide (Albiniak et al. 2013). Moreover, during batch-fermentation, GFP protein showed export to the periplasm and release to the culture medium. In this study the TatAdCd translocase was expressed in *tat* null *E. coli* at shake-flask level to investigate which other proteins it could export to the periplasm.

Initially partial complementation at shake-flask was checked with microscopy, then two native *E. coli* substrates were tested and finally a set of heterologous proteins (including a completely synthetic protein) were utilised. Some of the latter biotechnologically relevant molecules including hGH, IFN and an scFv had been tested with TatABC prior to this study. A completely synthetic protein, a maquette was also tested (G. Sutherland et al. 2018). With a TorA signal peptide at the N-terminus and by expressing CyDisCo components, export to periplasm was evident in wild type *E. coli* cells (Alanen et al. 2014). The polycistronic expression of CyDisCo (Erv1p and hPDI) downstream of the POI allows

disulphide bond formation prior to Tat export (Hatahet et al. 2010; Nguyen et al. 2011) . The interesting export results using TatAdCd in *tat* null *E. coli* lead to the second investigation, the proofreading capability of the more discrete complex.

The Tat system has long been known to reject (not export to periplasm), proteins that are not correctly or fully folded (Delisa et al. 2003)(Halbig, Wiegert, et al. 1999)(Richter & Brüser 2005a). Proteins tested tended to be native Tat substrates such as FeS proteins fused to a TorA signal peptide (Matos et al. 2008a)(Delisa et al. 2003). A heterologous yeast nuclear pore protein Nsp1p was tested too, unfolded FG repeats were rejected to varying degrees raising the possibility hydrophobicity may cause Tat rejection (Richter et al. 2007). However, like tests in plant thylakoids, a correlation was seen between physical protein lengths, not just protein instability (Cline & McCaffery 2007). Other unfolded heterologous proteins were tested to study Tat's remarkable ability (Alanen et al. 2014) but the degree of mis-folding and/or gross unfolding made deducing what Tat 'sensed' difficult to ascertain.

An scFv (termed scFvM) was sourced, fused with a TorA signal peptide and histidine tagged to become an ideal tool to study wild-type TatABC proofreading (Austerberry et al. 2017; Jones et al. 2016). See Figure 9. The collaborating group generated a number of variants, all classed as 'folded' as determined by very similar T_m values, peak florescence and secondary structure values (via circular dichromism spectra), which were mutated originally for a solubility study. These mutants therefore were useful candidates to quantify Tat's quality control based solely on conformation flexibility and surface features. Moreover scFvM has two disulphide bonds so CyDisCo could be used both to test Tat's

structural flexibility acceptance (due to export efficiency of oxidized and reduced forms).

The mutants were grouped into three main subsets:
1) 1, 2 or 3 Lys-Glu salt bridges were introduced in place of un- charged residues, increasing surface charge in a net-neutral manor (SB1, SB2, SB3).
2) To create a more positively- or negatively- charged area, 5 uncharged surface residues were substituted with Lys, Glu or Arg to create a 5Lys, 5Glu, and 5Arg.
3) Hydrophobic domains were introduced to the N-terminal region via substitution of 4, 5 or 6 neutral, polar surface residues with Leu (4N-Leu, 5N-Leu and 6N-Leu).

The somewhat surprising result was that TatABC could export the all of these mutants to the periplasm (export was only greatly reduced in the 5Arg mutant) suggesting that the Gram-negative system does tolerate major changes in surface charge and hydrophobicity (Jones et al. 2016). scFvM export was however highly dependent on CyDisCo expression, without it, only 10% of protein was exported in wild type cells. This dependency proposes the theory TatABC can sense conformational flexibility. Another mutant (scFvM 26tail) with an unstructured tail at the C-terminus blocks export confirming that unfolded regions are not tolerated. In this study the same scFvM mutants were tested for export using the Gram- positive translocase in *tat* null *E. coli,* with the aim of testing TatAdCd's proofreading capability.

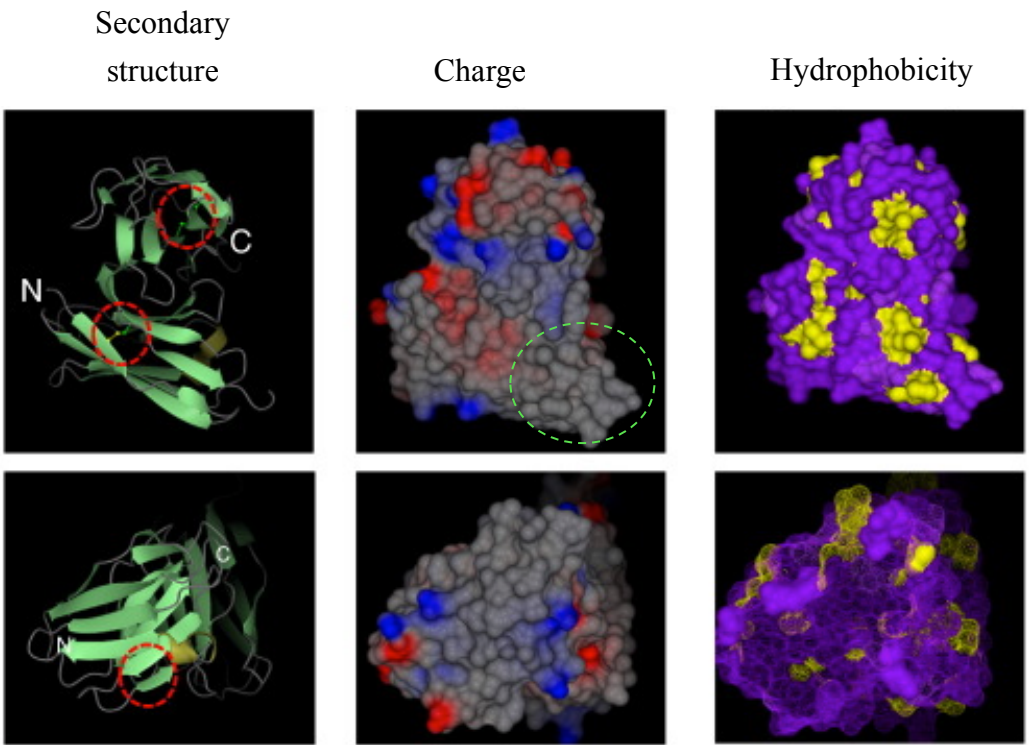


Figure 9. **scFvM wild-type modeled from two projections.**

The ribbon model shows β-sheets in green, α-helices in gold and loop regions in grey. 2 disulphide bridges on the top panel are circled red. Uncharged protein/neutral amino acids are coloured grey, while blue indicates positive and red indicates negative amino acids. The uncharged patch circled green is where salt bridge or charge substitutions took place. Finally, hydrophobicity is coloured yellow and hydrophilic areas are coloured purple. Images were made using CCP4 Molecular Graphics and this figure was adapted from Jones 2017.

3.2 Results

3.2.1 Partial complementation of Δ*tatABCDE* by TatAdCd expression

The *E. coli* Tat pathway is predicted to have relatively few substrates (Palmer et al. 2011) but mutant *E. coli ΔtatABCDE* (Δ*tat*) cells exhibit a range of phenotypes, highlighting the importance of Tat. These defects include impaired septation, decreased motility and an increased sensitivity to detergent (Stanley et al. 2001). Failed full septation causes *E. coli* cells to divide improperly and hence, form chains up to 10 cells long that can be viewed under a light microscope (Harrison et al. 2005; Ize et al. 2003). A possible reason for this effect is the mislocalisation of cell wall amidases AmiA and AmiC, both Tat substrates (Bernhardt & De Boer 2003). Of importance to this study the filamentous phenotype is complemented by expression of TatABC machinery, but only partially complemented by expression of Gram-positive TatAdCd machinery in (Δ*tat*) *E. coli* cells (Albiniak et al. 2013). As *B. subtilis* Tat can export one of the most complex *E. coli* Tat substrates (TMAO reductase, TorA), it was assumed it would complement the mutant phenotype (Barnett et al. 2008)(Mendel et al. 2008). However, the filamentous phenotype, at least during the fermentation process was still present when TatAdCd was expressed in Δ*tat E. coli* cells, suggesting the complementation is not complete (Albiniak et al. 2013). In this study, we sought to visualise by light microscopy the phenotype of Δ*tat E. coli* cells and the same cells expressing TatABC and TatAdCd from the pLys-BAD plasmid at shake flask, 2 and 5 hours post induction- see Figure 10. The images show full complementation (single cell *E. coli*) when TatABC was expressed, but largely the characteristic ‘chaining’ phenotype when TatAdCd was expressed.

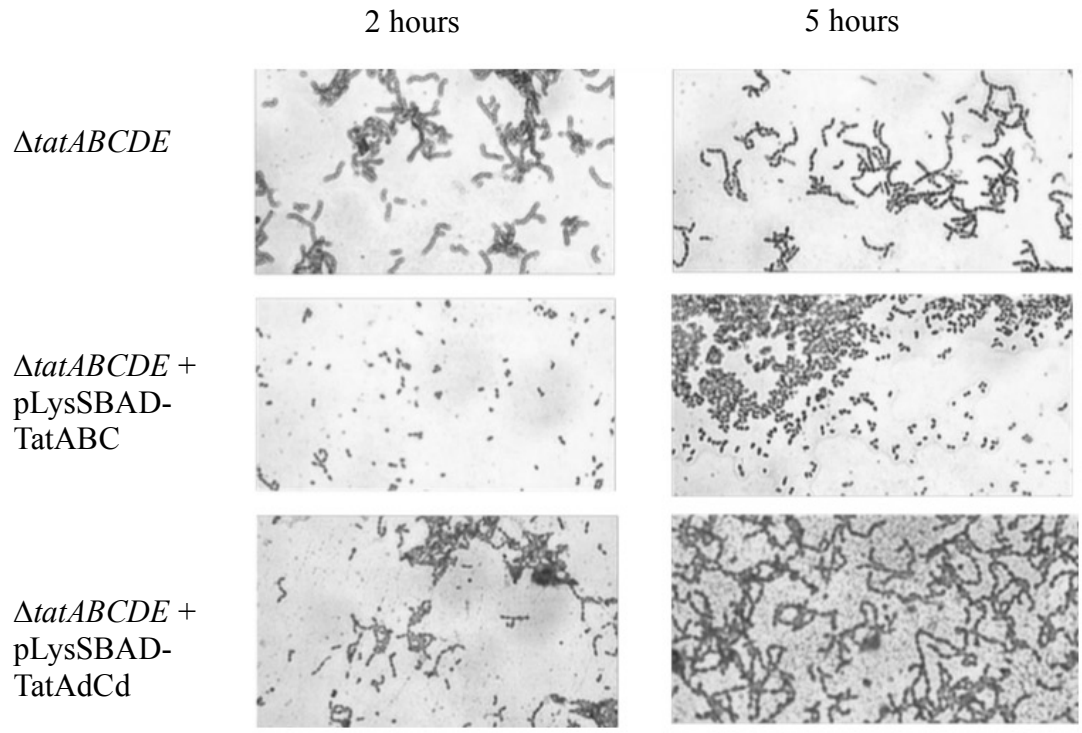


Figure 10. **Light microscope images of Δ*tat* cells expressing nothing, TatABC or TatAdCd**

Δtat cells were imaged either expressing nothing (top panels), TatABC (middle panels) or TatAdCd (bottom panels) from the pLysS-BAD plasmid after 2 or 5 hours of induction with arabinose (right and left images, respectively). The images show that TatABC complements the filamentous phenotype of *Δtat* cells, whereas TatAdCd does not.

3.2.2 TatAdCd fails to export two native *E. coli* Tat substrates

Before testing heterologous proteins we thought it important to check the ability of TatAdCd to export native *E. coli* Tat proteins. TatAdCd can export the *E. coli* Tat substrate TorA (Barnett et al. 2008)(Trimethylamine-N-oxide reductase) (Barnett et al. 2008), sized 86 kDa with a molybdenum cofactor, but what about other native Tat proteins? The protein NrfC (25kDa) has a role in nitrate reduction and had been shown to be exported by Tat (Matos et al. 2008). Figure 11 emulates these results as wild-type *E. coli* exports the protein

to the periplasm (P) where it is processed to the mature form, however TatAdCd rejects the protein and only the precursor form is present in the M (insoluble) fraction. FhuD (33 kDa), another native *E. coli* Tat substrate was synthesised for this study, is a periplasmic binding protein of ferric hydroxamate, an iron siderophore. Results (Fig. 11) show wild-type *E. coli* cells exporting FhuD to the periplasm (lane P), whilst TatAdCd did not export FhuD. The protein did express and is present in the M (insoluble fraction) like NrfC.

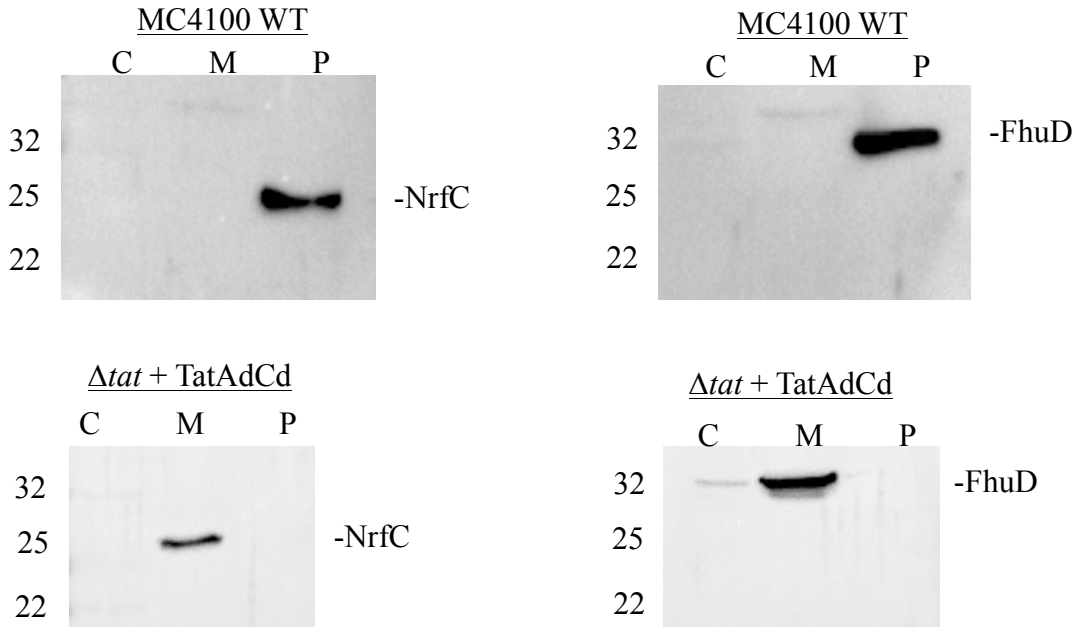


Figure 11. **NrfC and FhuD native *E. coli* Tat substrates are not exported by TatAdCd** The precursor forms of *E. coli* NrfC and FhuD were expressed in wild type (left panels) or Δ*tat* cells expressing TatAdCd from the pLysS-BAD plasmid (right panels). After 3 hours induction, cells were fractionated into cytoplasm, membrane or periplasm samples (C, M, P). The samples were immunoblotted against the C-terminal His tag on the target substrates. Wild type cells were able to export protein to the periplasm, cleaved to mature size whilst TatAdCd expressing cells fails to export the proteins.

3.2.3 TatAdCd fails to export a range of proteins that are exported by TatABC

The first heterologous protein tested in Δ*tat E. coli* expressing TatAdCd was TorA-GFP, this protein was exported to the periplasm and then secreted to the medium during batch-fermentation (Albiniak et al. 2013). In this study we investigated TatAdCd export with more complicated protein biotherapeutics during shake flask experiments. Proteins chosen, hGH, IFN α2β and a VH domain, had already been tested and successfully exported via TatABC with a TorA signal peptide. It's interesting to note that although some of these proteins contain disulphide bonds, none required their formation prior to export by Tat. CyDisCo was therefore not required as the proteins formed near native structures without disulphide bond formation (Alanen et al. 2015).

To test export, POI's were expressed from a pET-23 based plasmid alone in Δ*tat* cells (the negative control) or co-expressed with either TatABC (the positive control) or TatAdCd expressed from a pLys-BAD plasmid. After 3 hours of expression, the cells were fractionated into cytoplasm, membrane and periplasm to identify the localisation of the POI within the cell via immunoblots against the C-terminal His6 tag. Expression and membrane positioning of TatABC and TatAdCd was confirmed with an immunoblot against the C-terminal Strep II tag on TatC and TatCd- data not shown.

TorA-hGH (23.6 kDa), TorA-IFN (20.7 kDa) and a TorA-VH domain construct (9.5 kDa), as expected, are exported to the periplasm in TatABC expressing cells as mature, processed protein (without TorA signal peptide). The unexpected result was that TatAdCd expressing cells failed to export any of these proteins, no mature protein was detected in the periplasm- see Figure 12. The cytoplasm and membrane fractions also lack much protein; they were probably subject to proteolysis when they did not export, as seen before (Alanen et al. 2015). A media sample was also taken showed no protein release- data not shown.

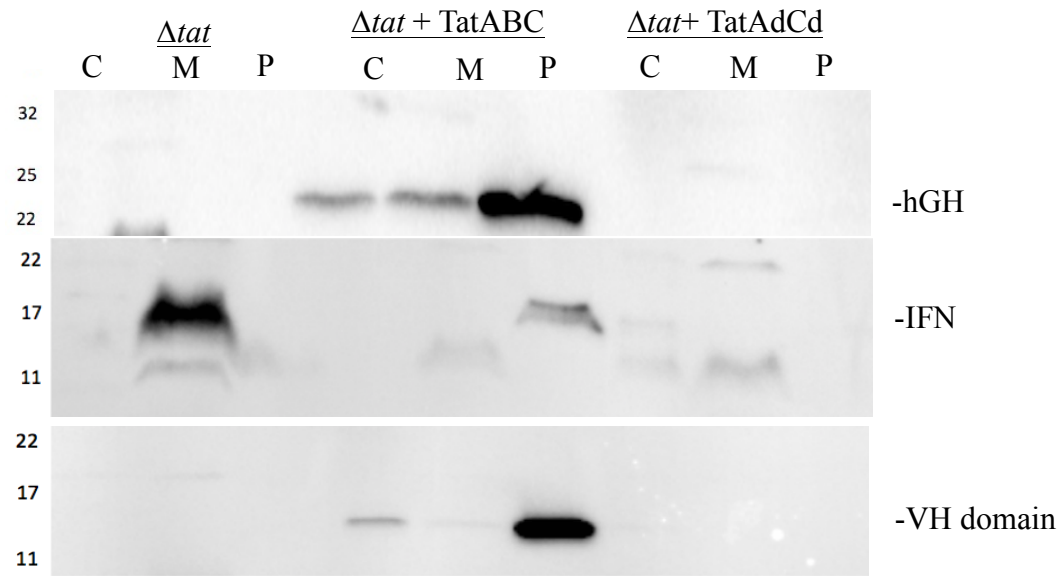


Figure 12. **TatAdCd fails to export hGH, IFN and VH domain protein constructs**
TorA-hGH, TorA-IFN and TorA-VH domain constructs were expressed in Δ*tat* cells and in Δ*tat* cells expressing either TatABC or TatAdCd. After 3 hours induction, cells were fractionated into cytoplasm, membrane or periplasm samples (C, M, P). The samples were subjected to immunoblotting using antibodies to the C-terminal His tag on the target heterologous proteins. The negative control on the left, Δ*tat* cells, do not export any protein to the periplasm. Δ*tat* cells expressing TatABC exports all three proteins, which are cleaved to mature size in the periplasm. The test on the right, Δ*tat* cells expressing TatAdCd fails to export the proteins to the periplasm.

3.2.4 TatAdCd fails to export a Maquette protein

While all the proteins tested previously are found in nature, the synthetically made Maquette protein is not. Based on the BT6 maquette developed (Farid et al. 2013), it's a small protein (17 kDa) consisting of four largely identical α-helices that form a water excluding cavity which accommodates cofactors. Its histidine residues within the cavity ligate two hemes, which make the tightly folded protein an ideal Tat substrate. Indeed TorA-BT6 is exported by TatABC and moreover, BT6 mutants that inhibit heme binding exhibit less tight folding which decreases Tat's export ability further displaying Tat's proofreading capability (G. Sutherland et al. 2018). Figure 13 shows Δ*tat* TatABC expressing cells export TorA-BT6 maquette whereas TatAdCd expressing cells do not.

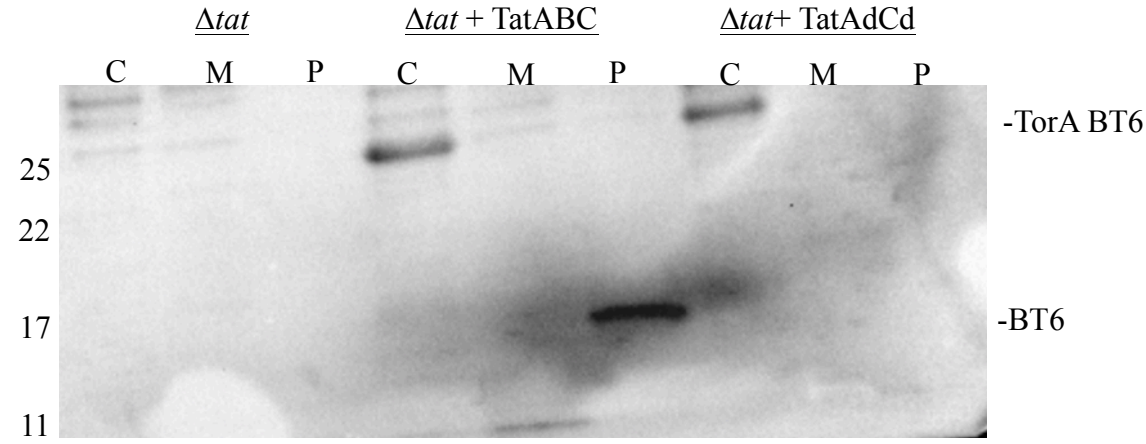


Figure 13. **The synthetically made Maquette protein is not exported by TatAdCd** TorA-Maquette (21 kDa) is expressed in Δ*tat* cells and in Δ*tat* cells expressing either TatABC or TatAdCd. After 3 hours induction, cells were fractionated into cytoplasm, membrane or periplasm samples (C, M, P). The samples were subjected to immunoblotting using antibodies to the C-terminal His tag on the target heterologous proteins. Δ*tat* cells do not export the maquette , it's degraded (17 kDa), TatABC expressing cells do, whilst TatAdCd expressing cells fail to export the protein to periplasm.

3.2.5 TatAdCd efficiently exports two different scFv proteins

We next tested the export of two Tor-A scFv constructs through TatAdCd in Δ*tat* cells. One scFv, scFvO was raised against the omega peptide of β-galactosidase, whilst the other, scFvM was raised against C-met. Both scFv's are translocated through wild-type TatABC. Of interest, svFvO does not require CyDisCo for export, whereas scFvM is much more dependent on the expression of CyDisCo for export (Jones et al. 2016)(Alanen et al. 2014). Figure 14 shows that both scFv's are exported by TatAdCd and moreover, similar to wild-type Tat, scFvO export is independent of CyDisCo for export whereas scFvM is largely dependent on CyDisCo expression. This suggests that at least for scFv's, both translocases exhibit similar preferences in terms of substrate structure.

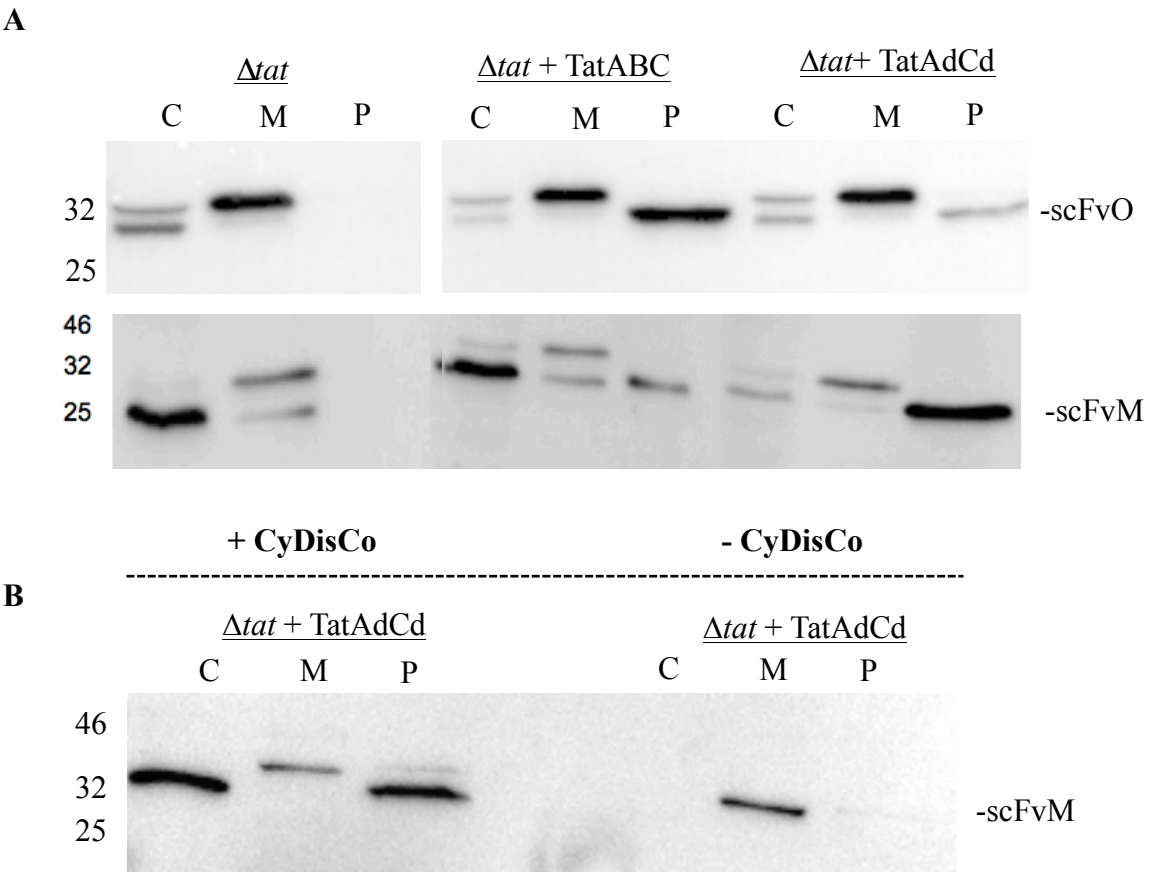


Figure 14. **TatAdCd efficiently exports two scFv's, scFvM requires CyDisCo.**

A. Two different scFv's termed scFvO and scFvM were expressed with an N-terminal TorA signal peptide in Δ*tat* cells and in Δ*tat* cells expressing either TatABC or TatAdCd. CyDisCo components Erv1p and PDI were also constitutively expressed with the proteins. After 3 hours induction, the cells were fractionated to yield cytoplasm, membrane or periplasm fractions (C, M, P) which were immunoblotted using antibodies to the C-terminal His tag. The negative control Δ*tat* cells do not export scFv's whilst both Δ*tat* cells expressing TatABC and Δ*tat* cells expressing TatAdCd export scFvO and scFvM to the periplasm, cleaved to mature size.
B. scFvM is only exported to the periplasm when CyDisCo is expressed.

3.2.6 Effects of scFvM surface mutations: TatAdCd is more selective than TatABC

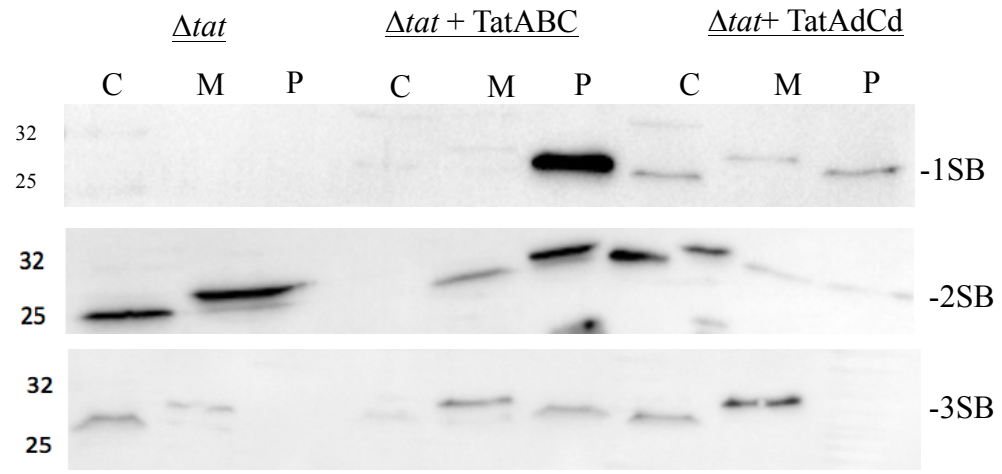
Up to this point, the only protein TatAdCd exported apart from GFP was scFv's. Therefore to investigate the proofreading capability of the Gram-positive Tat, structurally similar scFvM variants with surface mutations were tested via expression and export to the periplasm. These mutants had previously been tested with TatABC in *E. coli* and unexpectedly in general, the mutants had little effect on export efficiency compared to the unmodified scFv (Jones et al. 2016). Here, we tested whether TatAdCd was able to export the same range of variants, in order to test whether TatAdCd has different requirements to TatABC. Three mutants groups were generated in this study based on former literature of Tat's proofreading, the salt-bridge group, the charged group and the hydrophobic group. To note, in some experiments (1SB and 5Glu) it appears the Δ*tat* cells have not expressed the POI, however in our experience, these proteins have more than likely been degraded by proteases because they were not exported.

Figure 15A shows SB mutants (1 2 or 3 SB's), as shown previously TatABC exports these well but only 1SB and 2SB are exported reasonably via TatAdCd, 3SB is not exported at all.

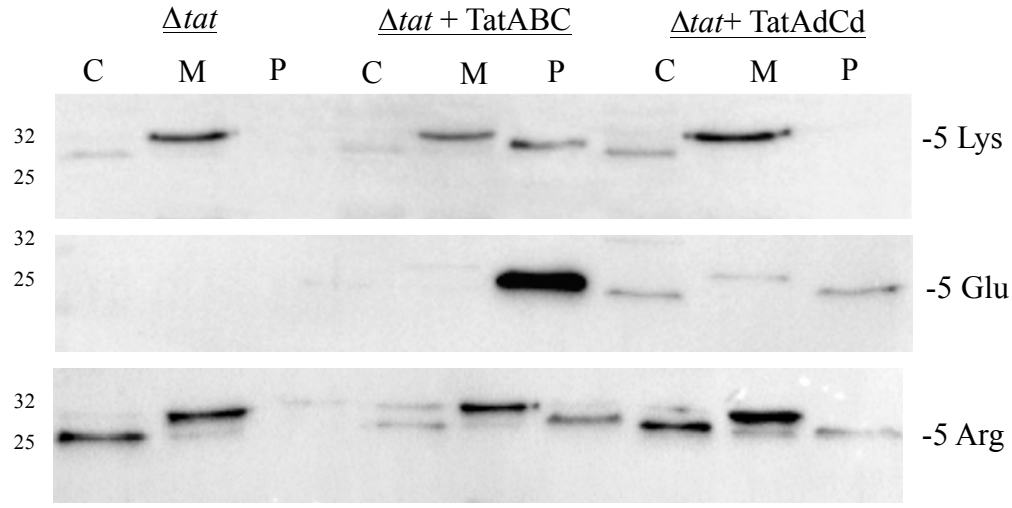
Figure 15B shows the charged group, which contain a more positively or more negatively charged areas (5 Lys, 5Glu or 5Arg). TatABC exports all the mutants, although the 5Glu variant is exported more efficiently than the 5Lys or 5Arg. Indeed in wild type cells, the 5Arg mutant did greatly decrease the export efficiency with wild-type TatABC (Jones et al. 2016). TatAdCd did export 5Glu and 5Arg, although not as efficiently as TatABC, but 5Lys was rejected. Given this result, to seek whether TatAdCd inherently discriminated against Lys residues, a mutant was constructed where 5Arg residues were mutated to 5Lys residues (5R>K). The results show TatABC exports 5Arg>K well and TatAdCd exports it lesser amount, but any export indicates TatAdCd does not have an inherent aversion to Lys residues.

Finally in Figure 15C, the hydrophobic group with an N-terminal hydrophobic patch of Leu residues were tested (4N-Leu, 5N-Leu, 6N-Leu). Once again, TatABC exported all three mutants well but TatAdCd only exported the 4N-Leu mutant weakly. Both the 5N-Leu and the 6N-Leu variants were rejected for export via TatAdCd. Overall, this data suggests the Gram- positive Tat exhibits a far more stringent level of substrate specificity compared to TatABC, especially when the substrates bear exposed hydrophobic domains.

A


Δtat
Δtat + TatABC
Δtat+ TatAdCd
C M P C M P C M P
32
25
-1SB
-2SB
-3SB

B


Δtat
Δtat + TatABC
Δtat+ TatAdCd
C M P C M P C M P
32
25
-5 Lys
-5 Glu
-5 Arg

C

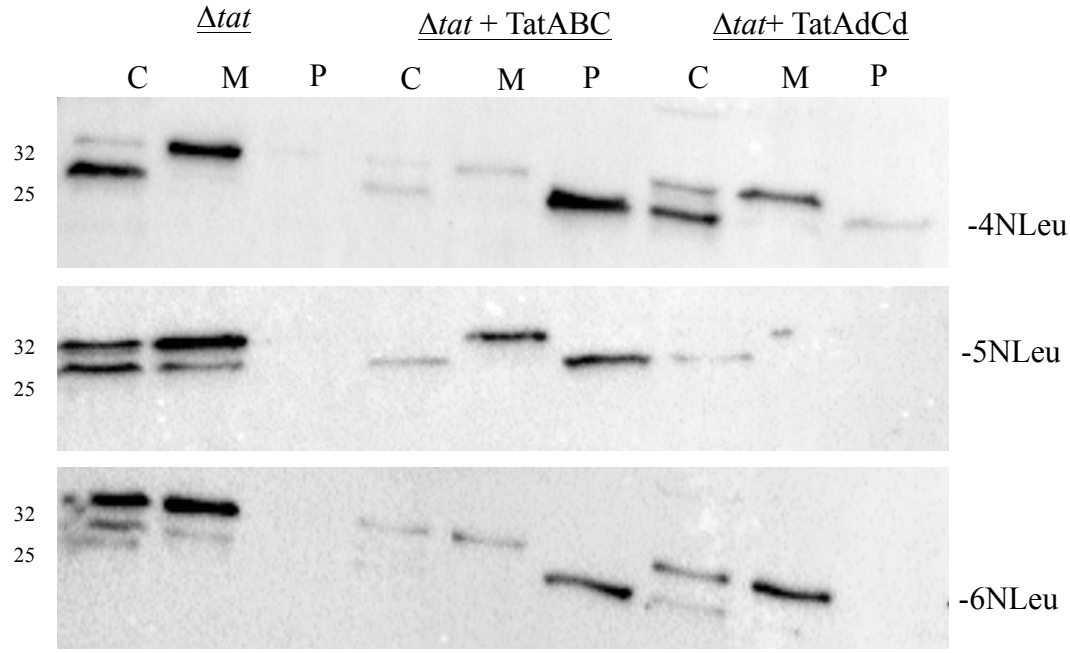

Δtat
Δtat + TatABC
Δtat+ TatAdCd
C M P C M P C M P
32
25
-4NLeu
-5NLeu
-6NLeu

Figure 15. **Differential export of mutant scFvM variants by TatABC and TatAdCd**

A. scFvM salt bridge variants SB1, SB2 and SB3 were expressed in Δ*tat* cells and in Δ*tat* cells expressing either TatABC or TatAdCd with CyDisCo. After 3 hours induction, the cells were fractionated to yield cytoplasm, membrane or periplasm fractions (C, M, P) which were immunoblotted using antibodies to the C-terminal His tag. All salt bridge variants were exported via TatABC whilst TatAdCd exported SB1 and SB2 to a lesser extent but did not export SB3 to the periplasm.

B. scFvM variants with either significant positive (Lys or Arg) or negative (Glu) charge in the same area were expressed in Δ*tat* cells and in Δ*tat* cells expressing either TatABC or TatAdCd with CyDisCo. After 3 hours induction, the cells were fractionated to yield cytoplasm, membrane or periplasm fractions (C, M, P) which were immunoblotted using antibodies to the C-terminal His tag. All variants were exported via TatABC whilst TatAdCd exported 5-Glu and 5-Arg to a lesser extent but did not export 5-Lys to the periplasm. TatAdCd could however weakly export scFvM 5R>K mutant.

C. scFvM variants with increased surface hydrophobicity 4N-Leu, 5N-Leu and 6N-Leu were expressed in Δ*tat* cells and in Δ*tat* cells expressing either TatABC or TatAdCd with CyDisCo. After 3 hours induction, the cells were fractionated to yield cytoplasm, membrane or periplasm fractions (C, M, P) which were immunoblotted using antibodies to the C-terminal His tag. All variants were exported via TatABC whilst TatAdCd exported only weakly exported the 4N-Leu mutant.

3.2.7 TatAdCd rejects misfolded variants of scFvM

Whilst the mutants point to a different level of substrate specificity, they do not address the question of whether TatAdCd rejects misfolded substrates. We therefore tested whether TatAdCd could export scFvM-26tail. This mutant has 26 residues appended to the C-terminus of scFvM which causes it to be rejected by TatABC (Jones et al. 2016). TatAdCd also rejected the same construct (Fig.16). Another mutant scFvM-P172S has a substitution between the two major scFvM domains, which we suspect causes destabilization to its overall structure. TatABC exported this variant (Jones et al. 2016), whereas TatAdCd rejected svFvM-P127S. This result suggests TatAdCd does have the ability to proofread misfolded proteins and is more sensitive to substrate conformation than TatABC.

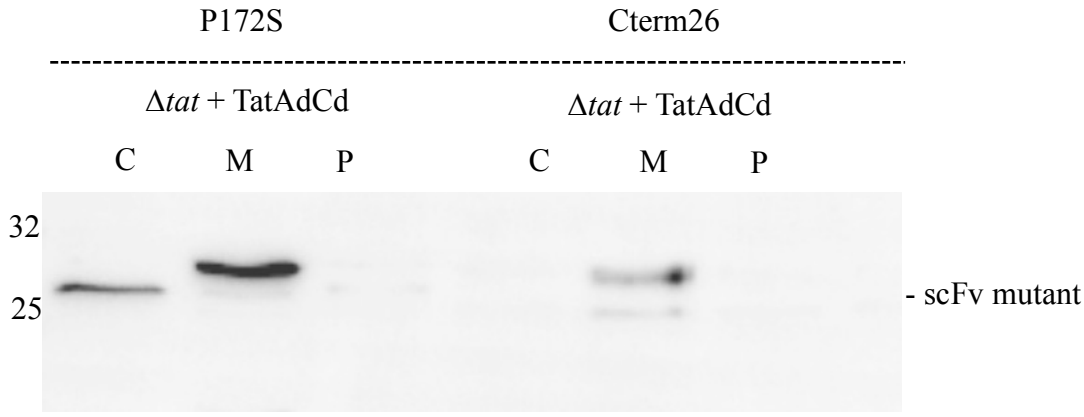


Figure 16. **TatAdCd rejects scFvM variants that are misfolded.**
TorA-scFvM variants P172S and 26tail (containing a 26- residue C-terminal extension) were expressed in TatAdCd-expressing CyDisCo cells, and the P172S mutant was also expressed in TatABC-expressing CyDisCo cells (top right). Samples were fractionated into cytoplasm, membrane or periplasm fractions (C, M, P) and immunoblotted using antibodies to the C-terminal His tag. TatABC can export the P172S mutant whilst TatAdCd rejects export along with the 26tail mutant.

3.3 Discussion

The primary aim of this study was to further investigate the capabilities of the *B. subtilis* TatAdCd, to compare substrate specificity and proofreading abilities with its *E. coli* TatABC counterpart. The TatAdCd system has been less well studied, perhaps because it only has one native substrate, PhoD (a 64 kDa phosphodieasterase). Under phosphate limited conditions, TatAdCd is co-regulated from the *phoD* operon (Pop et al. 2002). A second Gram-positive minimal Tat system exists, TatAyCy. It too has a single substrate YwbN (a 46 kDa heme containing peroxidase) although is constitutively expressed (Jongbloed et al. 2004). Both systems are able to partially complement Δ*tatABCDE E. coli* cells by forming an active translocon, but only TatAdCd exported one of the largest known Tat substrates, TorA (90 kDa) (Barnett et al. 2008). Moreover, TatAdCd was shown to have a more relaxed specifity, not only could TatAd functionally replace TatAy (TatAy cannot replace TatAd), but when overexpressed, TatAdCd can aid the translocation of YwbN (Eijlander, Jongbloed, et al. 2009). Thus Jongloed *et al*, concluded thefore that TatAdCd could be a good candidate for comercial production of heterologous proteins. To test this, the model protein GFP was chosen and fused to the TorA signal peptide, TorA-GFP was able to be exported via TatAdCd in *tat* null cells, of *E. coli* as well as DmsA-YFP (Mendel et al. 2008). Further studies upscaling the process from shake flask to Batch fermentation studies were undertaken with overexpression of both TorA-GFP and TatAdCd in *tat* null cells. During the fermenation mature GFP was excreted into the culture medium from the periplasm (Albiniak et al. 2013). This study and Fig.10 perhaps provide some insight into why this happens. TatAdCd does not fully complement *tat* null cells, the filamentous phenotype of Δ*tatABCDE E. coli* cells is still visible upom Gram-positive Tat expression. Cell wall amidases (AmiA/AmiC Tat substrates) involved in cytokinesis are thought to be

mislocalised due to the absence of TatABC, perhaps TatAdCd can not export one or both proteins, leading to a defective cell envelope (Ize et al. 2003). Although it's interesting that TatAdCd can export GFP fused to an AmiC signal peptide (Barnett et al. 2009). But the primary failure of TatAdCd to export the enzymes (one or both) that cleave the division septum, could render the outer membrane compromised and hence "leaky" once the periplasm is full of heterologous protein. Such a platform has the potential for biotechnological exploitation as the system could decrease downstream purification and shield the protein from proteases within the cell.

The native TatABC substrates NrfC and FhuD did not export via overexpressed TatAdCd in *tat* null cells, this result was not suprising as another native substrate, SufI (50 kDa), was also rejected by both TatAdCd and TatAyCy (Barnett et al. 2009). It's unclear whether it's the proteins native signal peptide which is not recognised by TatAdCd (although they all contain the RR motif) or the mature protein which is not accepted for export.

This study built upon previous studies with TorA signal peptide fused to alternative biotechnologically relevant heterologous proteins to exploring the substrate specifity of TatAdCd. When compared to the native *E. coli* Tat substrate data, overexpressed TatAdCd did not export a range of proteins which TatABC could. The proteins rejected for export through TatAdCd include a VH domain, IFN, hGH and a synthetically constructed maquette. It's unclear why the Gram- positive Tat is unable to export these proteins, especially when these proteins are all smaller than other proteins it can export, GFP, PhoD and TorA.

However, TatAdCd was able to export two different scFv's, scFvO (CyDisCo independent) and scFvM (CyDisCo dependent). Dependency upon CyDisCo emulates the TatABC export system (Alanen et al. 2015)(Jones et al. 2016) and interestingly, scFvM is folded and active in the cytoplasm in a reduced state (Edwardraja et al. 2010). We interpret this to mean CyDisCo 'locks' the two scFv domains in a more rigid confirmation and is therefore deemed more sufficiency folded for Tat export. Perhaps this tight conformation is exactly what TatAdCd will export, the fact it rejected the svFvM-P127S mutant is evidence for this. It's known GFP folds extremely fast and tightly in the cytoplasm (and hence it cannot be exported by Sec) and speculatively, scFv might too have a tighter confirmation than the other heterologous proteins tested in this study. scFvM's CyDisCo dependance might also be because reduced cysteine residues are more hydrophobic than their disulphide bonded counterparts (Nagano et al. 1999) and the hydrophobic scFvM mutants were either rejected or significantly reduced in export via TatAdCd.

Previously scFvM had been used to research the 'proofreading' phenomenon of TatABC (Jones et al. 2016) scFvM variants with surface mutations that are nonetheless structurally identical, were used to investigate Tat's ability to sense surface charge (Edwardraja et al. 2010). TatABC was able to export all the variants (although a couple to a lesser degree), however this study shows TatAdCd is much more selective. The presence of 1or 2 salt-bridges (SB) is tolerated by TatAdCd, but 3SB is completely rejected; it is however unclear if this is due to the position of the third SB or because of structural changes. The second group introduced significant either positive (Lys or Arg) or negative (Glu) charge in the same area. TatAdCd exported mutants' 5Lys and 5Glu, but 5Lys is completely rejected. Following this, 5Arg residues were mutated to 5Lys residues (5R>K) but TatAdCd weakly

exports this mutant, indicating TatAdCd does not have an inherent aversion to Lys residues. Further tests would be required to draw any conclusions regarding TatAdCd's preferences for substrate surface charge.

A clearer picture is drawn when analysing the export efficiency of variants of increasing hydrophobicity in their N-terminal domain with TatAdCd. Exposure of hydrophobicity is a hallmark of unfolded or incorrectly folded protein and it has been shown previously exposed hydrophobic domains block export for small, unstructured proteins (Richter et al. 2007). Although TatABC exported the hydrophobic scFvM variants, TatAdCd only exported 4N-Leu weakly and rejected 5N-Leu and 6N-Leu for export. This implies the TatAdCd proofreading mechanism rejects substrates above a certain hydrophobicity threshold, with the corresponding threshold for TatABC perhaps being set at a significantly higher level.

Whilst the misfolded scFv-26tail mutant is rejected for export by both TatABC and TatAdCd as one would expect given previous misfolded studies (Delisa et al. 2003) (Richter & Brüser 2005). The scFvM-P127S mutant that we suspect causes destabilization to its overall structure (due to its location between the two major scFv domains) is exported by TatABC and not by TatAdCd. This strongly suggests TatAdCd is more sensitive than TatABC in responding to changes in substrate confirmation and/or its flexibility.

Further investigations are clearly needed before a more definitive conclusion is made although it's evident TatABC and TatAdCd have different substrate selectivity mechanisms. Substrate hydrophobicity distinctly inhibited scFv mutant export via TatAdCd; alternative proteins (such as GFP) with hydrophobic patches could be

interesting. It would be interesting (although beyond the scope of this study) to more accurately identify both the heterologous proteins and the scFv variants structural flexibility. For example NMR would be a useful tool to link translocation efficiency and conformational flexibility. Finally TatAdCd is able to export a few proteins, there must be more it can translocate and will these proteins also be secreted to the culture medium during batch fermentation?

4. scFv format expression and export tests

4.1 Introduction

Biopharmaceuticals are largely therapeutic recombinant proteins produced from biological sources obtained by a biotechnological process. The production of the first therapeutic protein in the early 1980's generated a new industry, which has since thrived. 50% of the "best selling drugs" are currently biopharmaceuticals (Jozala et al. 2016) and the trend is still increasing as advances in science continue. Bacterial hosts were initially used and although other systems exist (mammalian, yeast, insect), bacterial hosts still account for 30% of recombinant pharmaceuticals on the market (Overton 2014) due to their rapid growth, high yield, cost-effectiveness and ease of scale up.

Monoclonal antibodies account for approximately half of the total sales of biopharmaceuticals (Ecker et al. 2015), however bacteria cannot fulfill eukaryotic post-translational modifications so are used for production of small antibody fragments instead. The antibody single-chain variable fragment (scFv) and antigen-binding fragment (Fab) became the most popular forms as the derivatives are easily manipulated and produced, but most importantly have specific antigen selectivity (Gupta & Shukla 2017). Indeed, scFv specificity, affinity and efficacy are integral and once designed they can be utilised for the preparation of immunotoxins, therapeutic gene delivery or as anti-cancer intrabodies (Ahmad et al. 2012).

Antibody formats are shown in Figure 15. A complete antibody contains four polypeptides, two identical heavy chains and two identical light chains connected by disulphide bonds and decorated with glycans. A Fab contains a variable and constant region from each chain

too, but only those that associate with the antigen. Whereas an scFv is simply composed of a variable heavy (V_H) and a variable light (V_L) region, joined by an additional flexible linker. A diabody exists as a non-covalent dimer of scFv, this structure has the capacity of multivalent binding, hence, there is an opportunity to design multifunctional antibody reagents.

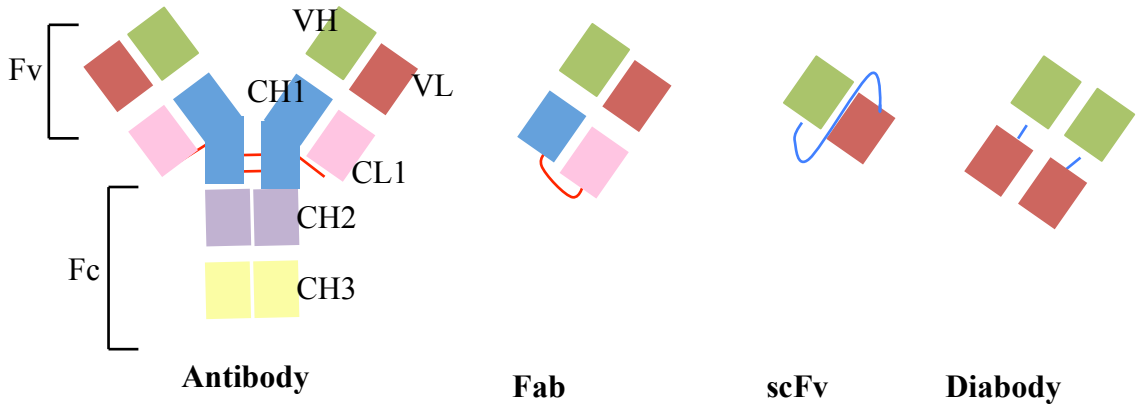


Figure 17. **Antibody formats.**

A whole antibody (IgG) has two identical light ($_L$) chains and two identical heavy chains ($_H$) bound by covalent and non-covalent forces. Each light chain is bound to a heavy chain by a disulphide bonds and the heavy chains are too linked by disulphide bonds (red lines). An antibody also contains complex carbohydrate structures (not shown). The upper variable regions (V) are responsible for specific antigen binding, whereas the lower constant regions (C) mediate biological activity. A Fab fragment contains two variable regions and one chain each from the constant region. Whilst an scFv only has two variable regions, usually linked by an additional linker (blue line) of at least 15 residues. A diabody has a much smaller linker (<11 residues), the protein cannot fold but non-covalent interactions can form a dimer structure.

Although many scFv's (just like in this study) originate from a phage display screen, their overall design is important for expression, stability and function. scFv construction should factor in V_H- V_L order, both domain orientations have been applied and it remains scFv

specific as to which order expresses and thereafter 'works' best (Völkel et al. 2001)(Weatherill et al. 2012).

Linker addition helps inherent instability issues (Bird et al. 1988) and its length is very important, the linker must span 3.5 nm between the C- terminus of the variable domain and the N-terminus of the other domain without affecting the ability of the domains to fold and form the binding site (Wörn & Plückthun 2001). A short linker (3-11 residues) is unable to fold into a functional Fv domain, but can form non-covalent dimers with other scFv's to form a diabody. The average linker is 15-20 residues which contain hydrophilic amino acids to avoid intercalation, often stretches of Gly and Ser are used for flexibility, $(GGGGS)_4$ is often used for example (Ahmad et al. 2012).

Finally, the stability of the scFv is imperative; it must at least be able to withstand physiological temperatures and pH. V_H and V_L association can be increased by the addition of intermolecular disulphide bonds. Disulphide stabilisation in one study showed pronounced improvement (>60 fold) of an scFv's half-life at 37^{o}C (Glockshuber et al. 1990). Another more recent study tested six disulphide bond locations and identified the best position (in terms of its biophysical properties). V_H40-V_L100 was then tested in 6 different scFv's to identify an interchain disulphide bond position that might improve the biophysical characteristic of scFv's universally (Weatherill et al. 2012). Although the disulphide stabilisation did not confer global protein stability (increased T_m for example), the variant did confer the benefit of permanently fixing monomer:dimer ratios.

Construct number	Design	Disulphide stabilised
001	V_H- linker-V_L	
002	V_H- linker- V_L	X
003	V_H-short linker-V_L	
004	V_H-short linker-V_L	X
005	V_L-linker-V_H	
006	V_L-linker-V_H	X
007	V_L-short linker-V_H	
008	V_L- short linker-V_H	X

Table 9. **CEA6 scFv construct designs.**

8 scFv constructs were chosen for this study; 001-004 have V_H-V_L orientations, while 005-008 have V_L-V_H orientations. 003/004/007/008 have a short linker and all even number constructs have an additional disulphide bond and are therefore "disulphide stabilised".

This study sought to test the ability of Tat to export 8 different variants of a widely used biopharmaceutical scFv specific for human carcinoembryonic antigen (CEA), commonly present on colon/rectal cancer cells (Osbourn et al. 1996). The scFv's were designed to be either in a V_H-V_L or V_L-V_H orientation, half were disulphide stabilised (with an additional disulphide, to have 3 disulphide bonds in total) and the variants had both long $(GGGGS)_4$ and short $(GGGGS)_1$ linkers. See Table 9. These designs were created to ascertain which variants expressed, folded and exported the best via Tat at shake flask and fermentation experiments. Information from our collaborators at MedImmune indicated that these 'new-format' scFv variants are difficult to express by traditional routes (cytoplasmic expression or export via Sec).

Each construct had a Tat specific TorA signal peptide and a C-terminal histidine tag for detection. For best results, the constructs were expressed (unless stated otherwise) in BL21 *TatExpress* cells, which have an engineered strong inducible bacterial promoter, ptac, upstream of the chromosomal *tatABCD* operon so that the Tat system is overexpressed

when IPTG is added. hGH export with a TorA signal peptide was increased 5 fold using *TatExpress* cells in comparison to wild-type (Browning et al. 2017). CyDisCo (hPDI and Erv1p) was also tested to aid disulphide bond formation in the scFv in the cytoplasm (Gaciarz et al. 2017). Optimisation of growth media was also undertaken, not least because medium composition can affect relative levels of soluble recombinant protein (Broedel et al. 2001).

4.2 Results

4.2.1 V_H-V_L format CEA6 scFv's express best in LB

All eight new format antibody structures (derived from CEA6) were expressed from a pET plasmid with an N-terminal TorA signal peptide for export to the periplasm and a C-terminal histidine tag for detection. The experiments used BL21 *TatExpress* cells which increase the number of Tat complexes in the membrane for best export yields (Browning et al. 2017). Previously two different scFv's have been exported via *E. coli* and *B. subtilis* Tat, one was CyDisCo dependent (scFvM) (Jones et al. 2016) and the other was not (Alanen et al., 2014). Therefore all constructs were co-expressed with CyDisCo on a different pBAD plasmid. The constructs were initially tested in LB media (like all other biotherapeutics tested in Chapter 3) and after 3 hours of expression, the cells were fractionated into cytoplasm, membrane and periplasm to identify where the variant scFv's had localised in the cells.

Only three of the eight formats were detectable when expressed in LB; 001, 31.5 kDa (V_H-V_L), 002, 31.5 kDa (V_H-V_L disulphide stabilised) and 003, 30.5 kDa (V_H-V_L short linker). However, they did not export and were mostly localised in the M, insoluble membrane fraction, Figure 18. This figure demonstrates the V_H-V_L orientation is best for protein expression in this system and that non-disulphide stabilised, long linker expresses the best according to band densitometry.

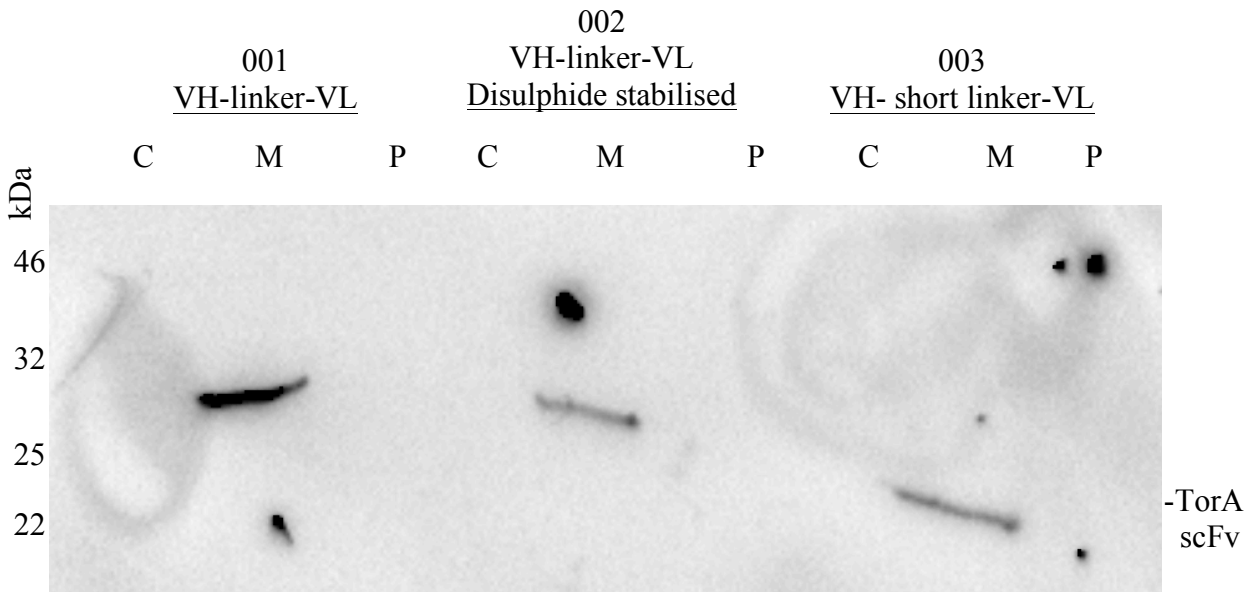


Figure 18. **Only V_H-V_L format scFv's are expressed in *TatExpress* cells using LB media**

Eight new format TorA-scFv constructs were tested in BL21 *TatExpress* cells, with CyDisCo, grown in LB media. After 3 hours induction, cells were fractionated into cytoplasmic, membrane or periplasm fractions (C, M ,P) and immunoblotted using antibodies to the C-terminal His tag. Only three constructs expressed and were localised to the M fraction, so were not exported by Tat. These expressed proteins were all in a V_H-V_L format; 001 expressed best with a long linker and only 2 disulphide bonds whereas 002 and 003 expressed less well, with an additional disulphide bond and a short linker, respectively.

4.2.2 LB and TB media experiments give similar results

The experiment (Section 4.2.1) was repeated in TB media to increase cell density. Teriffic Broth can also increase protein yields per cell and thus we hoped this would encourage translocation (Choi & Geletu 2018) of the scFv's to the periplasm. However, Figure 19 shows the expression levels and location of the three expressing scFv's are similar to the LB experiment, although the disulphide stabilised construct is now expressed less than the short linker variant.

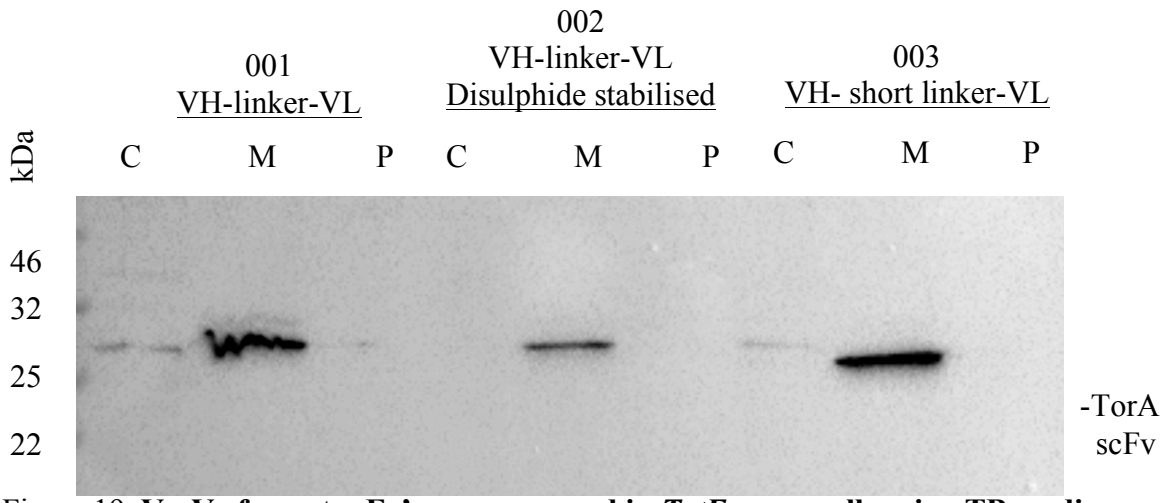


Figure 19. **V_H-V_L format scFv's are expressed in *TatExpress* cells using TB media**
New format TorA-scFv constructs were tested in BL21 *TatExpress* cells, with CyDisCo, grown in TB media. After 3 hours induction, cells were fractionated into cytoplasmic, membrane or periplasm fractions (C, M ,P) and immunoblotted using antibodies to the C-terminal His tag. Identical to expressions in LB media (Figure 18), the same three constructs expressed and were localised to the M fraction. However, the expression of the short linker variant (003) is better in TB than LB.

4.2.3 Auto Induction Media enables the export of 001/2/3 CEA6 scFv constructs

Auto Induction Media (AIM) was also tested with all 8 constructs. AIM was particularly suited this system as it is formulated to auto induce IPTG inducible (T7 promoters), and all constructs and TatExpress are under this promoter. Moreover, the media was developed for BL21 strains, so the TatExpress BL21 cells were optimal. The cells were pre-cultured in LB overnight and transferred to AIM for cell expression. At OD 0.4 the cells were induced with arabinose for pBAD CyDisCo production, but were not induced with IPTG. After both 3 and 21 hours (the latter time point was proven best for AIM expressions) (Studier 2005) the cells were fractionated. Figure 20A shows that after 3 hours, 001, 002 and 003 constructs were leaking into the periplasm. TorA was not cleaved from the scFv therefore this is not Tat export. However, after 21 hours (Figure 20B) the cells appeared to have

recovered as only TorA cleaved (mature) 001, 002 and 003 proteins were identified in the periplasmic fraction. Again, expression of disulphide stabilised 002 scFv construct is less than the non disulphide stabilised

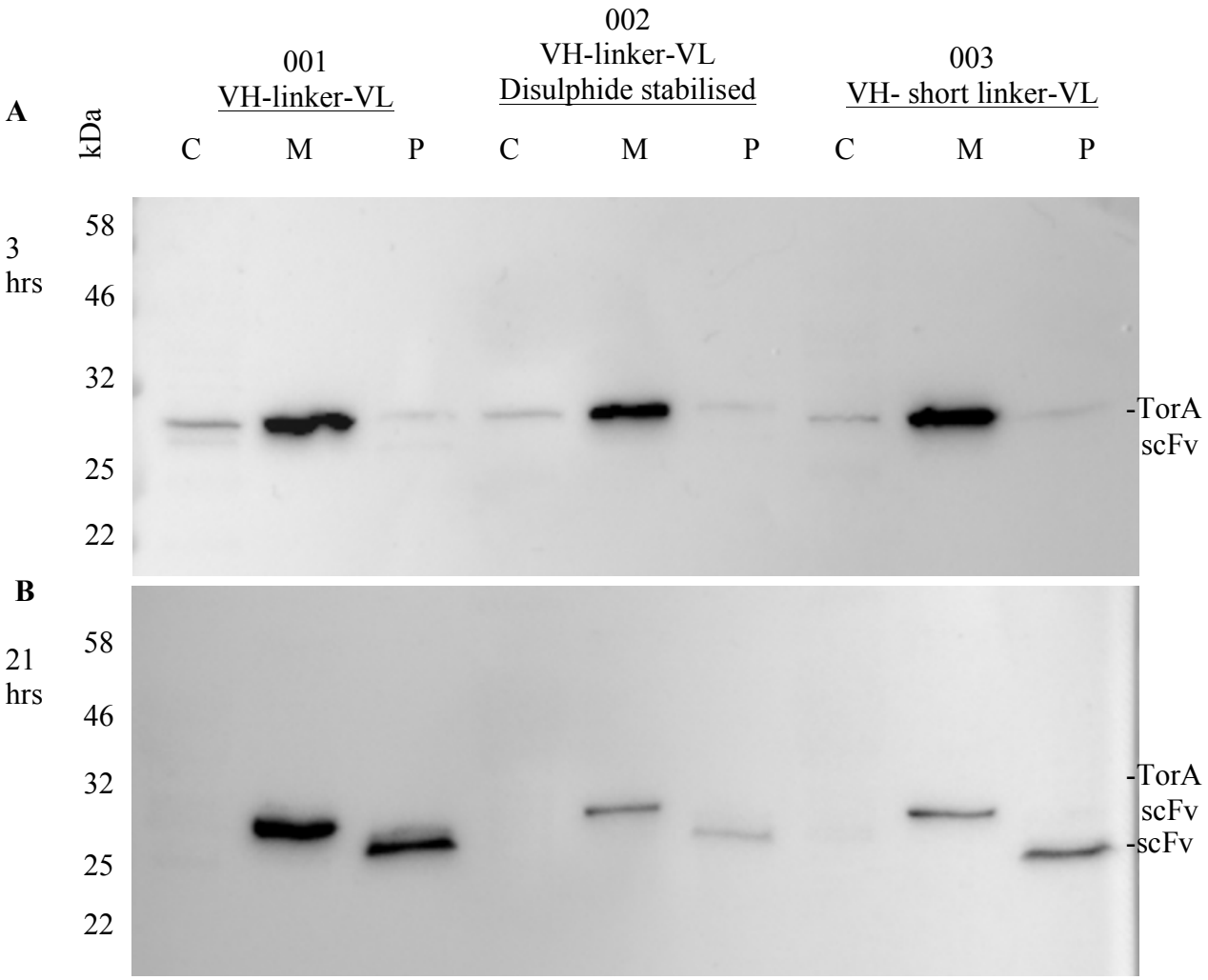


Figure 20. **scFv variants are exported to the periplasm using Auto Induction Media after 21 hours**

Eight new format TorA-scFv constructs were tested in BL21 *TatExpress* cells, with CyDisCo, grown in Auto Induction Media (AIM). Neither *TatExpress* nor scFv constructs were directly induced with addition of IPTG, but at OD 0.4 CyDisCo was induced with arabinose. **A.** After 3 hours of induction, cells were fractionated into cytoplasm, membrane and periplasm. Protein was mostly localised in the M fraction, but small amounts of precursor were visible in C and P too. **B.** After 21 hours of nduction cells were again fractionated into cytoplasm, membrane and periplasm. No cytoplasmic

protein is present but mature scFv has been exported to the periplasm. Best export is prevalent for the 001 construct, then 003 and finally the disulphide stabilised construct is less well expressed and exported.

4.2.4 Auto Induction Media hGH controls using *TatExpress*

As AIM had never been tested in our laboratory and the results were somewhat surprising, we used AIM with hGH control proteins to investigate the apparent lysing at 3 hours and export at 21 hours. Mature (without a signal peptide) and pre-protein (with a Tat specific TorA signal peptide) hGH were expressed in AIM using *TatExpress* cells. Without a signal peptide hGH will not export through Tat and will remain in the cytoplasmic fraction, but with TorA signal peptide, hGH can export through Tat to the periplasm.

The control results emulated the findings obtained in Section 4.2.3. After 3 hours and 21 hours 'post induction' (after OD 0.4 was reached), the cells were fractionated into cytoplasm, membrane and periplasm. The results show in Figure 21 that after 3 hours induction, both the mature and TorA- hGH proteins are present in the periplasm (at 22 kDa). It is likely that the mature hGH is present in the P fraction due to lysis of some of the cells. After 21 hours post induction, although the TorA-hGH construct has accumulated more as mature hGH in the periplasm, the second construct (mature hGH) is no longer present in the periplasm, suggesting that the cells are more intact at this point.

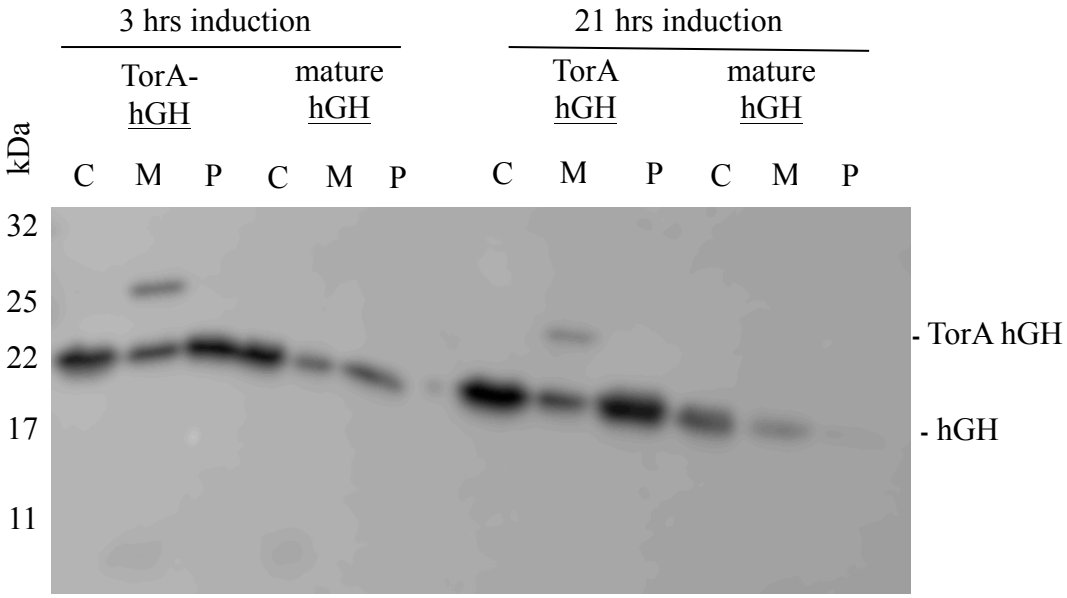


Figure 21. **Auto Induction Media TorA- and mature hGH controls.** TorA- and mature hGH were expressed in BL21 *TatExpress* cells using Auto Induction Media (AIM). After 3 and 21 hours of induction, the cells were fractionated into cytoplasm, membrane and periplasm and immunoblotted using antibodies to the C-terminal His tag. 3 hours post induction, both the TorA- and mature hGH are in all three fractions, although TorA-hGH also has the precursor protein in the M fraction. However after 21 hours induction, although more hGH is present in the TorA-hGH periplasm, no mature hGH is present in the mature hGH periplasm.

4.2.5 CyDisCo increases expression and export of CEA6 constructs

To investigate whether the proteins were dependent on the cytoplasmic expression of CyDisCo, the second plasmid (pBAD) CyDisCo was not co-transformed alongside the 8 constructs in *TatExpress*. The test proteins were instead solely expressed in AIM and the cells were fractionated at 3 and 21 hours post induction. There was no difference in cell growth without CyDisCo but protein expression and thereafter export differed significantly. Figure 20 shows two overexposed western blots, the blots had to be overexposed to identify protein presence. This alone indicates low protein expression or perhaps the protein was expressed as normal, but was degraded faster due to improper folding. After 3 hours post induction, 001, 002 and 003 are expressed in the cytoplasm/membrane fractions; no scFv is present in the periplasm. After 21 hours expression, 001 was exported (but to a lesser amount than with CyDisCo), 002 and 003 are not exported to the periplasm. Therefore, these disulphide stabilised and short linker variants require CyDisCo expression for Tat export.

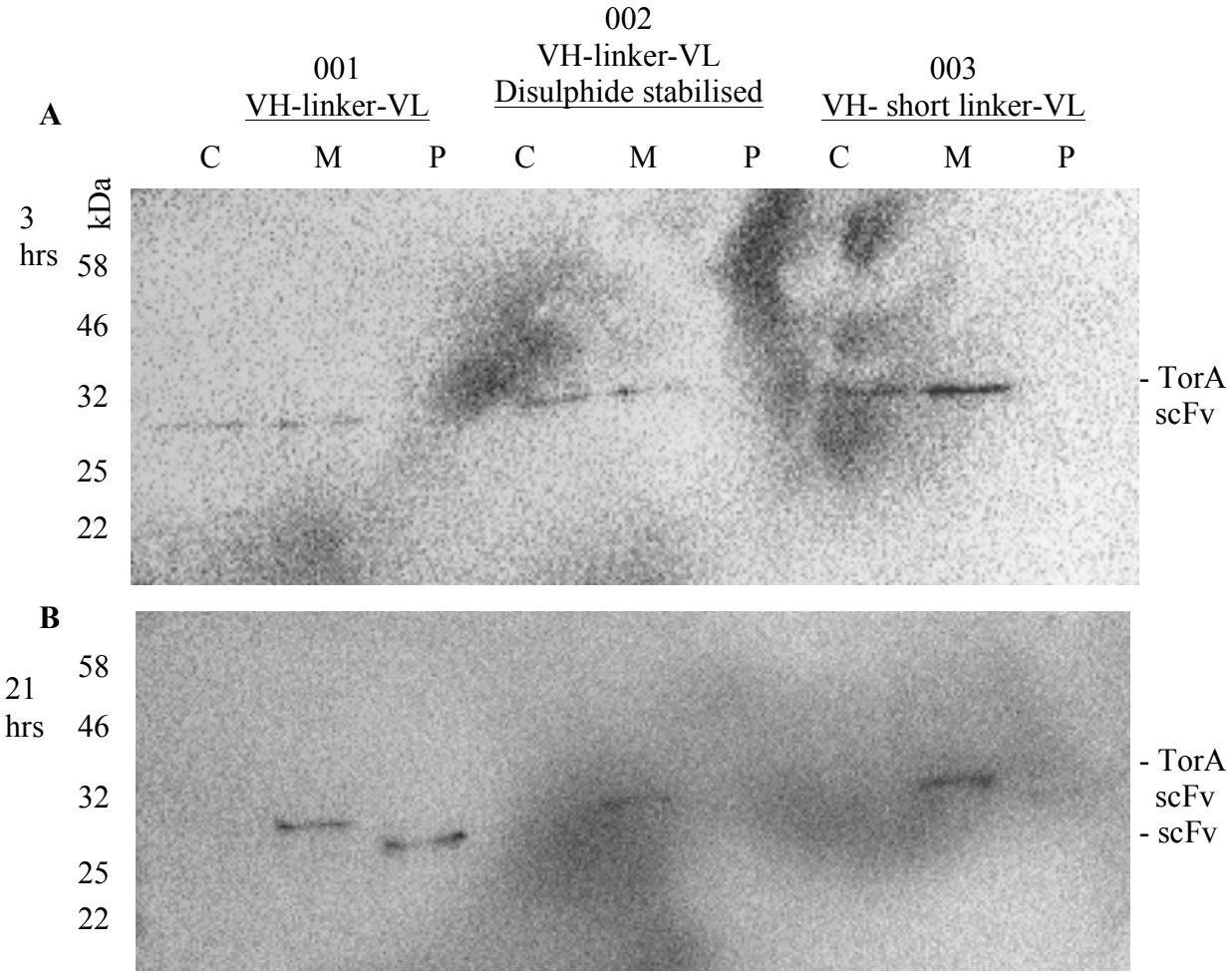


Figure 22. **Without CyDisCo, CEA6 scFv's express and export less well.** Eight new format TorA-scFv constructs were tested in BL21 *TatExpress* cells, without CyDisCo, grown in Auto Induction Media (AIM). After 3 and 21 hours of induction, the cells were fractionated into cytoplasm, membrane and periplasm and immunoblotted using antibodies to the C-terminal His tag. **A.** 3 hours post induction without CyDisCo, the protein expression levels were much lower, protein is present in the cytoplasm and membrane fraction. **B.** 21 hours post induction without CyDisCo, protein expression levels are still very low and only 001 construct is exported (to a small degree) while 002 and 003 remain primarily in the membrane fraction.

4.2.6 TatAdCd fails to export the CEA6 scFv constructs without CyDisCo

TatAdCd was previously able to export two different scFv's, Section 3.2.3 (Frain et al., 2017). To test whether it would also export other scFv's, delta *tat* cells were transformed with pBAD TatAdCd and the 8 scFv variants. As TatAdCd was expressed from a pBAD plasmid, pBAD CyDisCo could not be expressed, so these experiments were undertaken without the presence of CyDisCo. At 0D 0.4, *tatAdCd* was induced with arabinose while the constructs were left to auto-induce in AIM. 21 hours post induction, delta *tat* cells were fractionated into cytoplasm, membrane and periplasm. Although all 8 scFv variants did not export by TatAdCd, this was expected as no CyDisCo was expressed and Section 4.2.5 demonstrated the importance of CyDisCo. However, the unexpected result was that (for the first time) more than three constructs expressed, see Figure 23. As usual, 001 and 003 were detected in the M/insoluble fraction, while 002 (DS) was hardly detectable. 004, 006, 007 and 008 CEA scFv were also visualised in the M/insoluble fraction. After this result, CyDisCo was cloned polycistronically behind each scFv to test the ability of TatAdCd to export these proteins with CyDisCo and to streamline the auto-induction process.

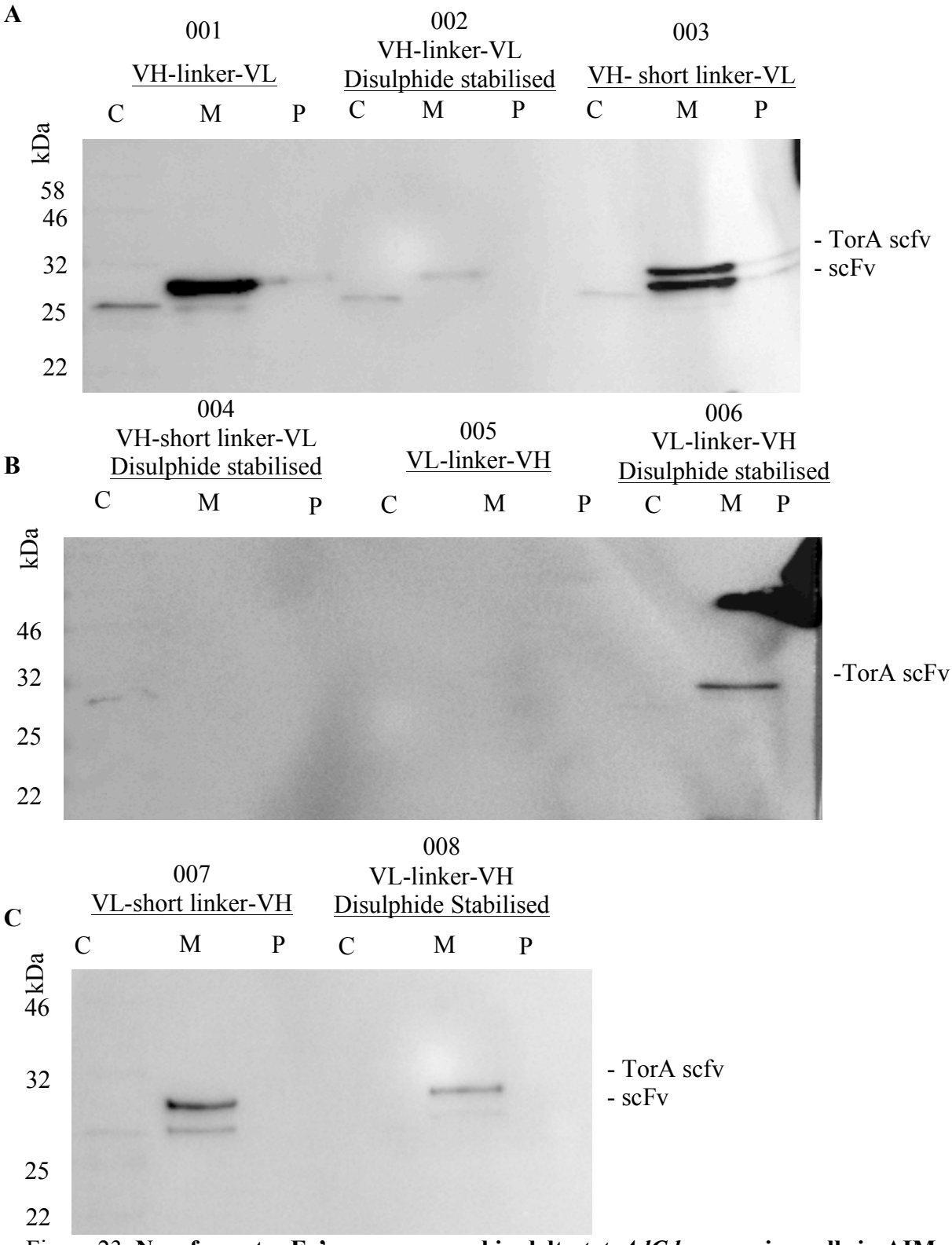


Figure 23. **New format scFv's are expressed in delta *tat*, *AdCd* expressing cells in AIM**

TorA-scFv constructs were tested in delta *tat* cells expressing TatAdCd without CyDisCo, grown in Auto Induction Media (AIM). After 21 hours induction, the cells were fractionated into cytoplasm, membrane and periplasm and immunoblotted using antibodies to the C-terminal His tag. Almost all constructs expressed protein (to varying degrees) and protein was mostly localised in the M fraction, no export was identified

A. Precursor scFv 001 expressed best, whereas 002 is weakly expressed and 003 localises in the membrane fraction as precursor and mature scFv.

B. 004, disulphide stabilised short linker scFv is just identifiable in the cytoplasmic fraction, 005 V_H-V_L is not expressed but its disulphide stabilised partner 006 is expressed and found in the M fraction.

C. Both V_H-V_L orientations with short linkers are expressed, but the non-disulphide stabilised variant expresses slightly better, albeit they're both localised in the M fraction.

4.2.7. Expression of scFv formats is not successful in SM6 media at shake flask

Three proteins in the V_H-V_L domain orientation were exported to the periplasm when grown in AIM with CyDisCo, Section 4.2.3. In two more standard media we did not observe scFv export (Section 4.2.1 and 4.2.2), but it was important to test the fermentation media, SM6, at shake flask too. Unlike other media, a pre-inoculant is grown for 6 hours in PY media before 1 mL is transferred to SM6 media. In SM6 at 0D 2, cells are induced with IPTG for 24 hours and then fractionated into cytoplasm, membrane and periplasm. Figure 22 shows both protein expression and export using SM6 was largely unsuccessful. Protein variants 001, 002 and 003 (with CyDisCo cloned on the same plasmid) were not expressed well, hence the oversaturated blot and all protein was localised in the membrane/insoluble fraction.

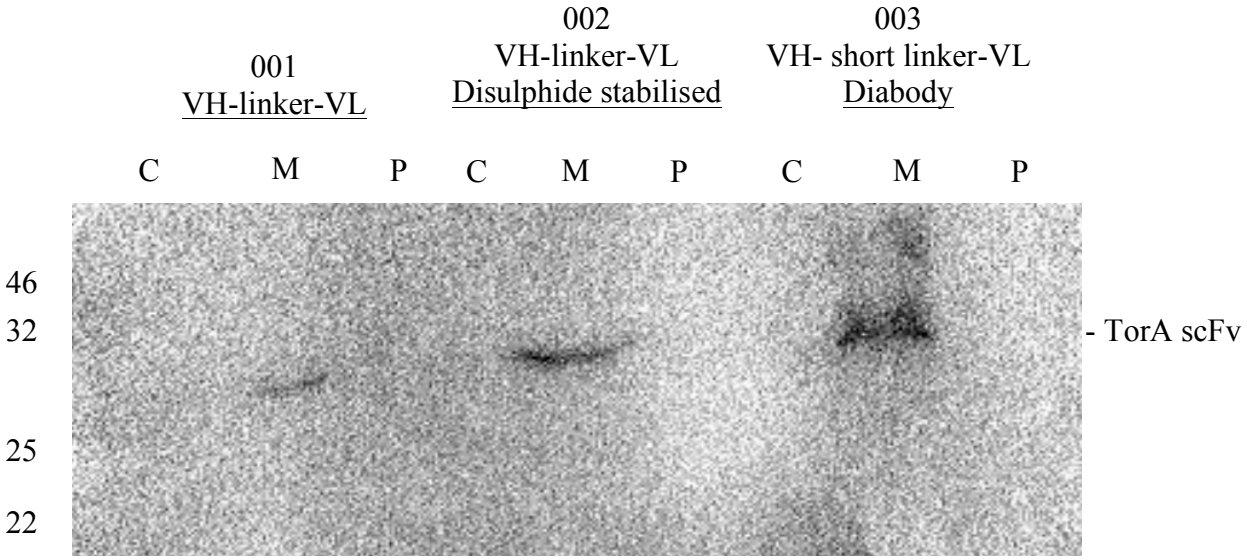
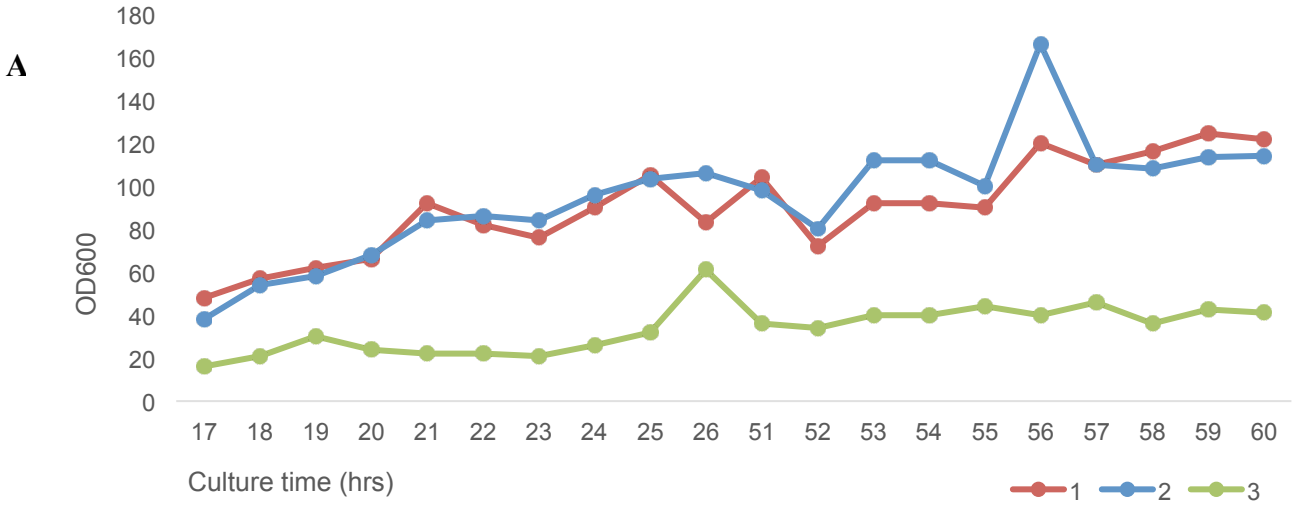


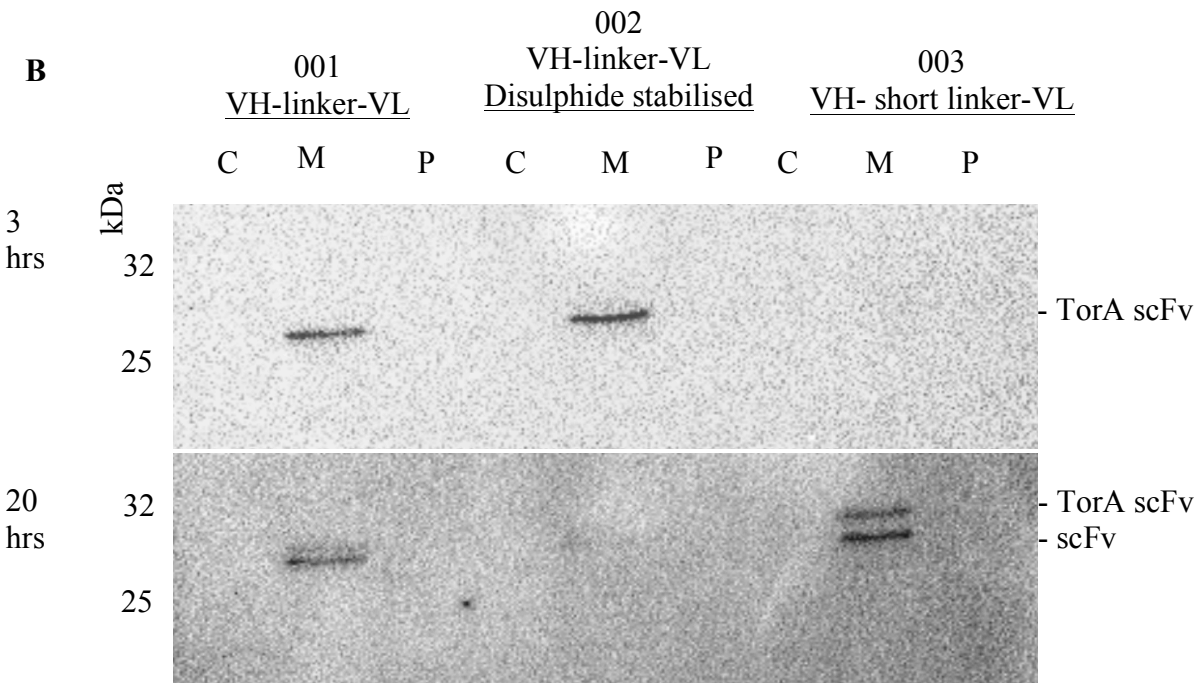
Figure 24.**V_H-V_L format scFv's are expressed in *TatExpress* cells using SM6 media**
Eight new format TorA-scFv constructs were tested in BL21 *TatExpress* cells, with polycistronically expressed CyDisCo, grown in SM6 media. After 21 hours induction, cells were fractionated into cytoplasmic, membrane or periplasm fractions (C, M ,P) and immunoblotted using antibodies to the C-terminal His tag. 001, 002 and 003 constructs expressed poorly and were localised to the M fraction, so were not exported by Tat.

4.2.8 Fed-batch fermentation of scFv constructs is successful but scFv's are not exported

Although at shake flask, SM6 media did not express or export the scFv constructs well, we thought the different conditions during fed-batch fermentation might be more beneficial. The three most successful constructs with CyDisCo were tested in parallel in Minifors 2 using eve® software. Figure 25A shows the OD 600 growth curve of all 3 constructs from when culture temperature was dropped to 25°C at 17 hours culture time to 60 hours culture time. Constructs 001 and 002 grew best and reached OD's of 120 and 115, respectively. 003 scFv did not grow as well as it only reached OD40 but all three constructs were induced at the same time point with IPTG. The cells were fractionated into cytoplasm,

membrane and periplasm pre-induction and 3, 20 and 29 hours post-induction. Pre-induction the blots were both blank (not shown). Figure 25B shows anti-his blots for 3 and 20 hours post scFv induction. At 3 hours, 001 and 002 are expressed but localised in the M insoluble fraction, 003 is not seen. Whereas 20 hours post induction, 002 is now mostly degraded and 003 is expressing in the M insoluble fraction. The general protein amounts are however very low. Figure 25C shows anti-flag blots for PDI on CyDisCo (57 kDa), CyDisCo is expressed well in the cytoplasm of both 001 and 002 constructs, but not so well for 003 construct.





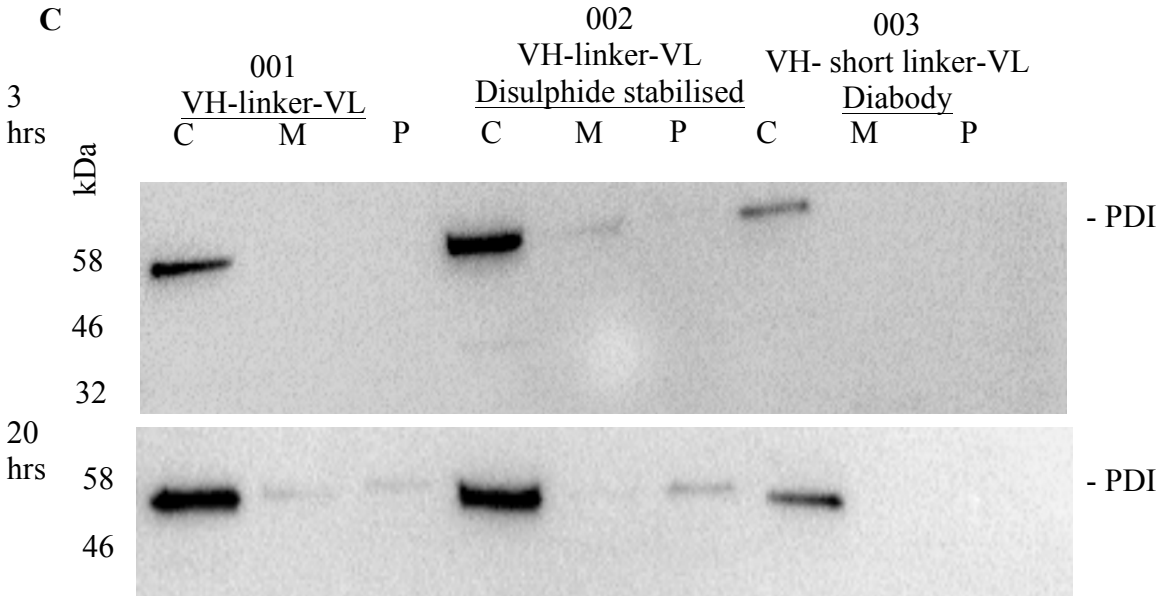


Figure 25. **Growth curves, scFv and CyDisCo blot of 001, 002 and 003 V_H-V_L constructs during fed-batch fermentation**

The three most promising scFv variants were tested with CyDisCo in fed batch fermentation using SM6 media. Cells were collected and fractionated pre-induction and 3, 20 and 29 hours post-induction with IPTG. Cytoplasm, membrane and periplasm fractions (C, M ,P) were immunoblotted using antibodies to the C-terminal His tag and the N-terminal PDI CyDisCo flag tag.

A. Cell growth was measured by taking OD600 every hour from 17 to 60 hours culture time during the day. Constructs 001 and 002 grew well and reached OD's of 120 and 115, respectively. 003 scFv did not grow as well as it only reached OD40. **B.** At 3 hours post induction, protein expression of constructs 002 and 003 were very low and protein was localised in the M fraction. 003 is not detected. Whereas at 20 hours post induction (lower panel) 002 is no longer detected, but 003 is now expressing localised to the M fraction. **C.** CyDisCo expression from 3 hours and 20 hours post induction is similar, PDI in CyDisCo is mostly present in the cytoplasmic fraction. PDI (in CyDisCo) polycistronic expression in 003 construct is much weaker than the other two constructs.

4.3 Discussion

The eight CEA scFv variants were created to test Tat's ability to export different protein orientations, disulphide stabilisation states and novel bispecifics, diabodies. It was clear from the beginning using LB media, that the V_H-V_L orientation was favored as only these proteins expressed at all; an expression preference due to variable region order has been noted previously (Tsumoto et al. 1994), but is scFv dependent. The fact that this orientation is preferred is not surprising as this scFv was selected in the V_H-V_L format from a phage screen (Osbourn et al. 1996); moreover, this orientation leaves the N-terminal end of the heavy chain variable region free, which contributes most to antigen interactions.

All variants had an N-terminal, Tat specific TorA signal peptide for export to the periplasm. In several of the constructs tested previously, the TorA signal peptide includes four additional residues of mature TorA protein to ensure authentic and efficient Tat export (Fisher et al. 2008). However proteolytic cleavage after ATA (Figure 3) means the additional residues remain on the mature biotherapeutic protein, producing an unfinished product, which is not ideal. Therefore, these constructs were the first to be tested without the four residues, which might contribute to the initial lack of export. This hypothesis might be far stretched, but a very recent study demonstrates the importance of the early mature part of Tat precursor proteins in TatBC binding (Ulfig & Freudl 2018). It is also interesting to note that our group has identified an alternative signal peptide, PhoD, which particularly exports scFv's well (unpublished) but the signal peptide requires altering for Tat specificity. Perhaps this could be tested in the future.

Disulphide stabilisation within the construct was a key component of this investigation as some literature has shown that this is beneficial (Section 4.1). Disulphide-stabilised constructs have never before been used with the Tat system, but it was considerd likely that the more folded the protein, the less conformational flexibility it has and the more likely the Tat system would export the protein as fully folded. The requirement for CyDisCo (Section 4.2.5) supports that disulphide bond formation is necessary for better scFv expression and export, which has been seen previously (Jones et al. 2016). However, the addition of one more disulphide bond to allow disulphide stabilisation did not aid expression or export, if anything CEA 002 expression was less than that of its non-disulphide stabilised partner. Although this might sound contradictory, additional disulphide bonds have been known to destabilise the protein in some cases (Pecher & Arnold 2009), and disulphide bond positioning is very important. It is also possible that the addition is unfavorable because the extra disulphide bond is another folding process for the protein. A Tat study, using an scFv protein suggests Tat export competence is related to the protein folding rate (Ribnicky et al, 2007). In my study, the additional disulphide bond would have slowed down protein folding and perhaps therefore hampered export.

A surprising result from this study was that scFv export was only ever seen when Auto Induction Media (AIM) was used, although we also tested three other media. AIM was a method first described by Studier, based on the different metabolic states on each individual cell, expression is induced automatically (Studier 2005). The method is not only very simple (no need to measure OD), but the system has been shown to increase protein yield and solubility (Grabski et al. 2005)(Lebendiker & Danieli 2014). Tat will not export aggregated, insoluble protein and so perhaps AIM offers just the right conditions for export.

I have not found in the literature an example of export incompetence being reversed using AIM, and this study is therefore novel.

The ability of scFv 003 to form a diabody either in the cytoplasm (pre-transport) or in the periplasm could be tested by running a native gel. Diabody formation could account for the export results and could demonstrate a novel form of Tat-exported recombinant protein. Other future perspectives include testing the ability of TatAdCd to export the scFvs with CyDisCo. One can hypothesise that they would export (like the other 2 scFv's) but these experiments were not carried out due to lack of time. Downstream, other approaches could also be investigated to aid Tat protein export in the future such as codon harmonisation (Claassens et al. 2017) and harmonisation of the target gene expression with Tat (Baumgarten et al., 2018).

5. TatAyCy structural study utilising SMA

5.1 Introduction

The Gram- negative and Gram- positive Tat systems have clear functional overlap, but their differing sizes and compositions allude to alternative, interesting export mechanisms. The data presented in this thesis further highlight their distinctions as Gram- positive TatAdCd has greater substrate selectivity and a tighter proof reading mechanism than TatABC in *E. coli* (Frain et al. 2017). To investigate the structure of *B. subtilis* TatAyCy, a novel, detergent free Styrene Maleic Acid (SMA) approach was undertaken to study the translocase structure.

Reminiscent of the situation in *E. coli*, TatAyCy is composed of two membrane protein complexes: TatAyCy and TatAy, which (as judged by gel filtration) form complexes of ~200 kDa (Barnett et al. 2009). TatAdCd is similarly sized at ~230 kDa hence both minimal TatAC systems are significantly smaller than *E. coli* TatABC (Barnett et al. 2008) which exists in a complex of ~370 kDa (Oates et al. 2005).

All Tat systems can export large substrates, which argues against the 'simple' channel forming model (like the SEC pathway) as the diameters of Gram- positive discrete TatA complexes are too small to export these proteins. Instead a translocation mechanism which involves the coalescence of relatively homogenous TatA complexes, with the formation of a more flexible pore (whose diameter adjusts according substrate size) has been suggested (Beck et al. 2013) (Baglieri et al. 2012). But recent doubt has also been cast on this model in *E. coli* TatA due to the arrangement of its APH (Koch et al. 2012)(Rodriguez et al. 2013) and a lack of flexibility in its glycine hinge (Walther et al. 2010). In which case the

alternative export mechanism “membrane destabilisation” is favoured (Brüser & Sanders 2003) and too supported by *E. coli* TatA data (Hou et al. 2018). A lack of knowledge still surrounds the structural properties of TatAyCy and TatAdCd; better structural insight could assist in explaining the Gram-positive Tat mechanism.

Published structural data largely focuses on *E. coli* individual constituents: TatA (Gohlke et al. 2005)(Zhang, Hu, et al. 2014), TatB (Zhang, Wang, et al. 2014), TatC (Ramasamy, Abrol, Suloway, Clemons, et al. 2013)(Rollauer et al. 2013) and TatE (Baglieri et al. 2012). However individual Gram- positive TatAd and TatCd structural data exists too (Walther et al. 2010)(Beck et al. 2013)(Nolandt et al. 2009). But there’s a lack of structural information on entire Tat complexes because pursuit of 3D Tat reconstruction is limited by several factors. Firstly, the interaction of Tat and its substrates in the inner membrane is transient and therefore hard to capture. Its heterogeneity is also a problem, as thousands of particles would have to be gathered to obtain enough data to represent the entire size range of the complex. And finally, Tat’s relatively small size makes it difficult to obtain electron microscopy of sufficiently high contrast data for analysis. However, due to the particularly small sizes of TatAyCy and TatAdCd these complexes have the advantage of being small enough to fit into a SMALP.

Membrane proteins ideally need to be kept in their native environment, the lipid bilayer, to maintain their structural integrity. The process of membrane protein purification therefore usually relies on detergent as their inherent amphipathic nature mimics the lipid bilayer. Regardless of the fact detergent screening can be time costing and expensive, the scientific basis behind the approach is too somewhat flawed. The detergent micelle created only

provides a rough approximation of the natural environment and can also lead to miscalculation of actual protein size. Moreover, recent studies have shown that the lipid bilayer does not simply act as a scaffold as it can too have profound effects on protein structures and/or function (Lundbaek et al. 2010). The novel approach of using Styrene Maleic Acid (SMA), which has alternating hydrophobic styrene and hydrophilic maleic acid moieties, allows membrane protein extraction by encapsulating the protein within its native lipid bilayer, see Figure 26 (Postis et al. 2015). This method results in maintenance of protein function and is highly amenable to single-particle cryo-EM (Gulati et al. 2014). Hence in this study, TatAyCy was purified with a SMALP scaffold and characterised using single-particle Cryo- EM.

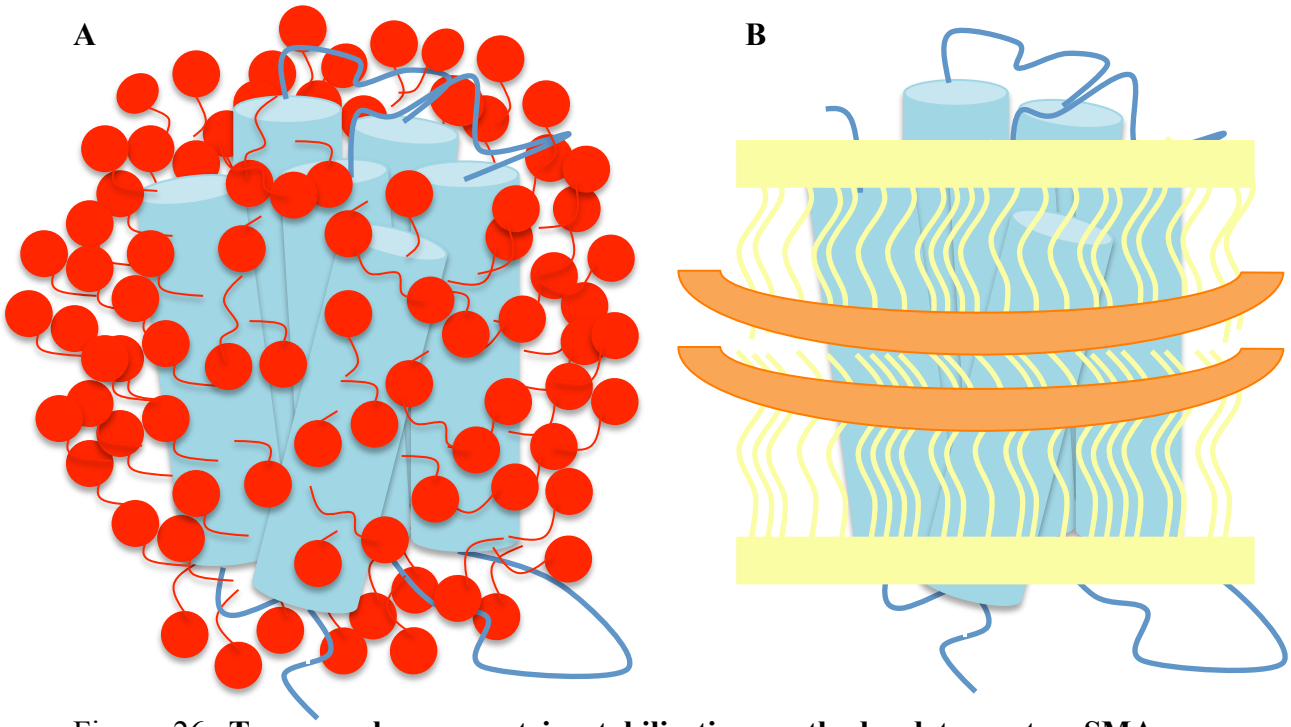


Figure 26. **Two membrane protein stabilisation methods, detergent v SMA.** Membrane protein is shown in blue. **A.** It is common for proteins to be solubilised with detergent (red) to create a spherical micelle. **B.** Using SMA, the native lipid bilayer is kept intact (yellow), amphipathic SMA copolymers form discoidal particles, nanodsics, which wrap around the protein complex and plasma membrane.

5.2 Results

5.2.1 Affinity chromatography of TatAyCy

Membranes containing TatAyCy were isolated as described in Materials and Methods and SMA was used to perturb the native membrane to create SMALPs. To achieve a highly pure sample, two purification steps were used. Initially affinity chromatography was performed using a nickel charged IMAC Sepharose^{TM} column, which would specifically bind the 10X His tag on the C-terminus of TatCy. Unspecific proteins were either captured as flow through or washed off the column with an increasing imidazole gradient. Tightly bound proteins (His tagged) were eluted with a maximum 200mM Imidazole concentration. Each column step was analysed both by SDS-PAGE stained with Coomassie and blotting against the C-terminal His tag. The results are shown in Figure 25.

Figure 27A is the Coomassie stained SDS-PAGE gel. Lane 1 is post SMA addition to the membrane sample and lane 2 is column flow through, they both contain hundreds of proteins that need to be removed. Lanes 3 to 5 show an increase from 5mM to 60mM imidazole in the wash buffer. Each wash was 3 column volumes and with each increase in imidazole, the sample becomes purer. There is some TatAyCy in these samples, suggesting not all bound to the beads, but the majority of the protein was eluted. Lanes 6 to 10 are the elution fractions; the most prominent protein bands are TatAy (5 kDa) and TatCy (28 kDa) so the purification was successful and there was co-elution of the complex. However other contaminating proteins are still present.

Figure 27B shows the corresponding blot, His antibody was targeted to the C-terminal 10X his tag attached to the TatCy. Intense bands in lanes 5 to 10 validate the presence of TatCy at 28 kDa. Perhaps a small amount of protein degradation is also present at about 20 kDa but this is minor in comparison to the purified TatCy protein.

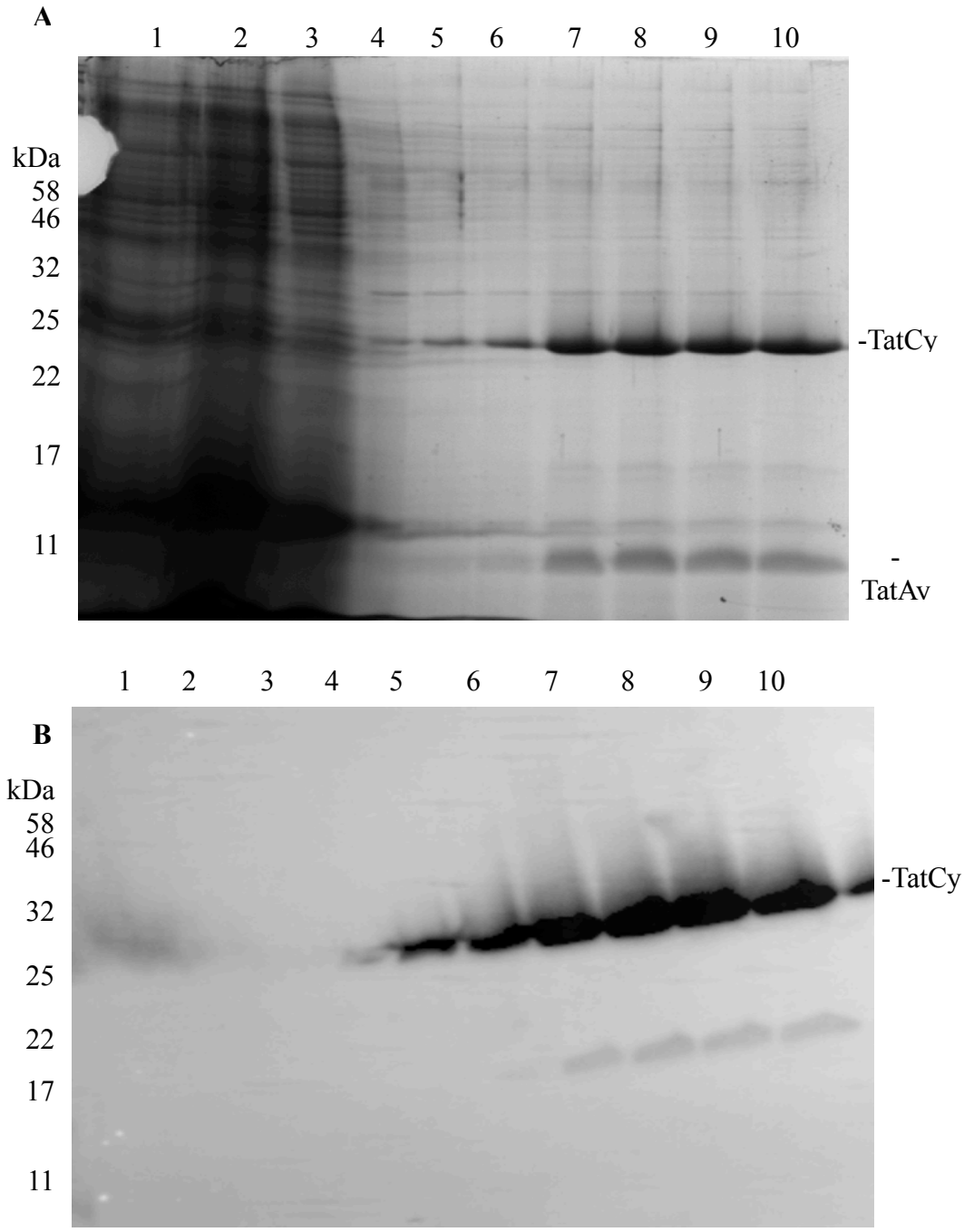


Figure 27. **Purification of TatAyCy by Affinity Chromatography.**

E. coli membranes were isolated from Δ*tat* cells expressing TatAyCy from a pBAD plasmid and subjected to SMA. SMALPs were then applied to a nickel charged IMAC Sepharose™ column for affinity chromatography via the 10X His tag on TatCy. **A**:SDS-Polyacrylamide Coomassie stained gel visualised (in order) SMA addition to isolated membranes, flow through, wash fractions and elutions in lanes 6 to 10. TatAy and TatCy are highlighted on the right. **B**: The corresponding gel was immunoblotted using antibodies to the C-terminal His tag on TatCy. TatCy is present in fractions 5 to 10.

5.2.2 Gel filtration of TatAyCy complex

The second polishing purification step involved using a Superdex200 10/300 gel filtration column on an ÄKTA™ FPLC system. Affinity purified TatAyCy, specifically fractions 7 and 8 were pooled and subjected to size elution chromatography for further purification to obtain a homogeneous complex for Cryo EM analysis. 0.5mL was applied to the column and fractions were eluted in 0.5 mL. Peak fractions, 18 to 22, in the spectra observed when absorbance was measured at 280 nM (Figure 28) and fractions either side were analysed further.

Figure 29A shows the SDS-PAGE stained Coomassie of the fractions; TatAyCy is now very pure with either a few or no contaminating proteins. Confirmation of TatAyCy purification is seen in Figure 29B, a blot against the N-terminal 10X his tag attached to the TatCy. The sensitivity of the blot in comparison to a Coomassie stain means TatCy is identified in more fractions than the Coomassie stain. It is also important to note that in comparison to affinity chromatography (Figure 27), the band is less intense perhaps signifying the protein is now less concentrated.

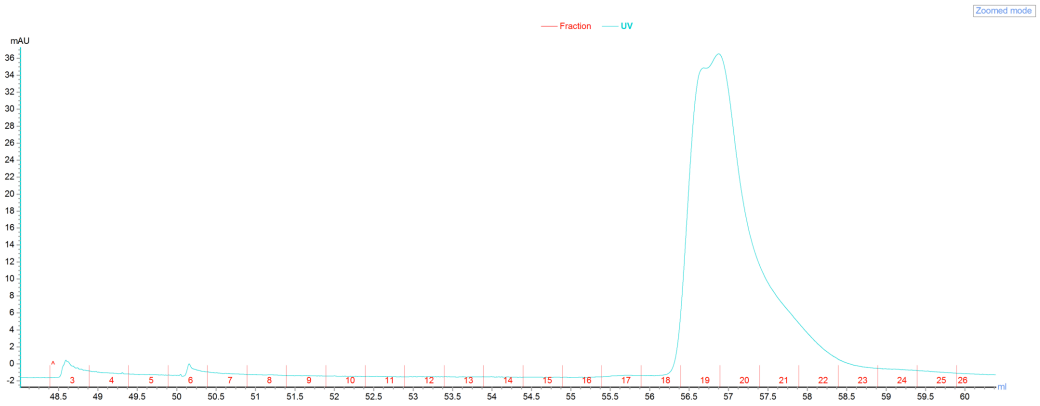


Figure 28. **Chromatogram of TatAyCy by Size Exclusion Chromatography.**
Affinity purified TatAyCy was applied to a Superdex200 10/300 gel filtration column and eluted in 0.5 mL fractions. Absorbance was measured at 280 nm, peak fractions 18-25 were analysed further.

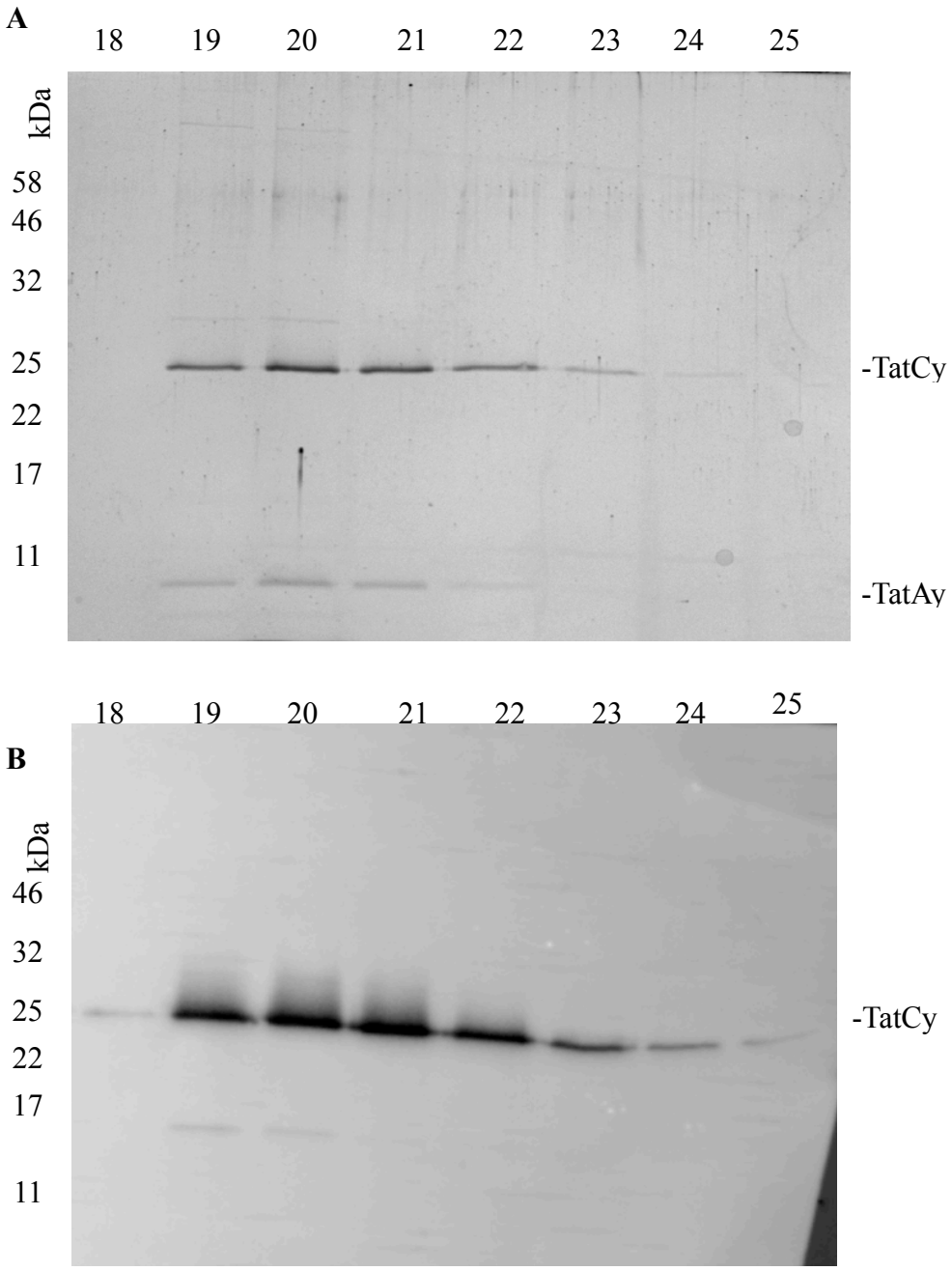


Figure 29. **Purification of TatAyCy by Size Exclusion Chromatography. A.** SDS-Polyacrylamide Coomassie stained gel visualised fractions 17 to 25 which showed the complex to be co-eluted and very pure. **B.** The corresponding gel was immunoblotted using antibodies to the C-terminal His tag on TatCy to confirm the proteins presence . TatCy was identified in all fractions but more protein is present in fractions 19 to 22.

5.2.3 Negative stain electron micrographs of TatAyCy

Negative staining is a routine sample preparatory step to increase contrast under the electron microscope. The method is rapid and useful for the detection of small biological macromolecules, so the step is vital for the initial characterisation of Tat protein complexes. TatAyCy samples eluted from affinity and size exclusion chromatography were coated in a thin layer of dried heavy salt method (Section 2.8.1) before imaging at the University of Warwick by Dr Sarah Smith. The amount of protein purified was problematic at this stage, so I optimised the purification. Either the sample was too dilute and therefore could not be visualised, or it was too concentrated and aggregated. For optimal results (although not ideal) Dr Smith used the nickel column fractions for both negative stain and cyro transmission electron microscopy (EM). Figure 30A shows an example of the relatively heterogeneous population of particles present. The diameters of 100 randomly selected particles from different images were measured by myself. The mean average diameter was 12.8 nm and diameters ranged between 9 and 16 nm, these values are consistent with previous unpublished TatAyCy work. Data was plotted on a size distribution graph, 30B.

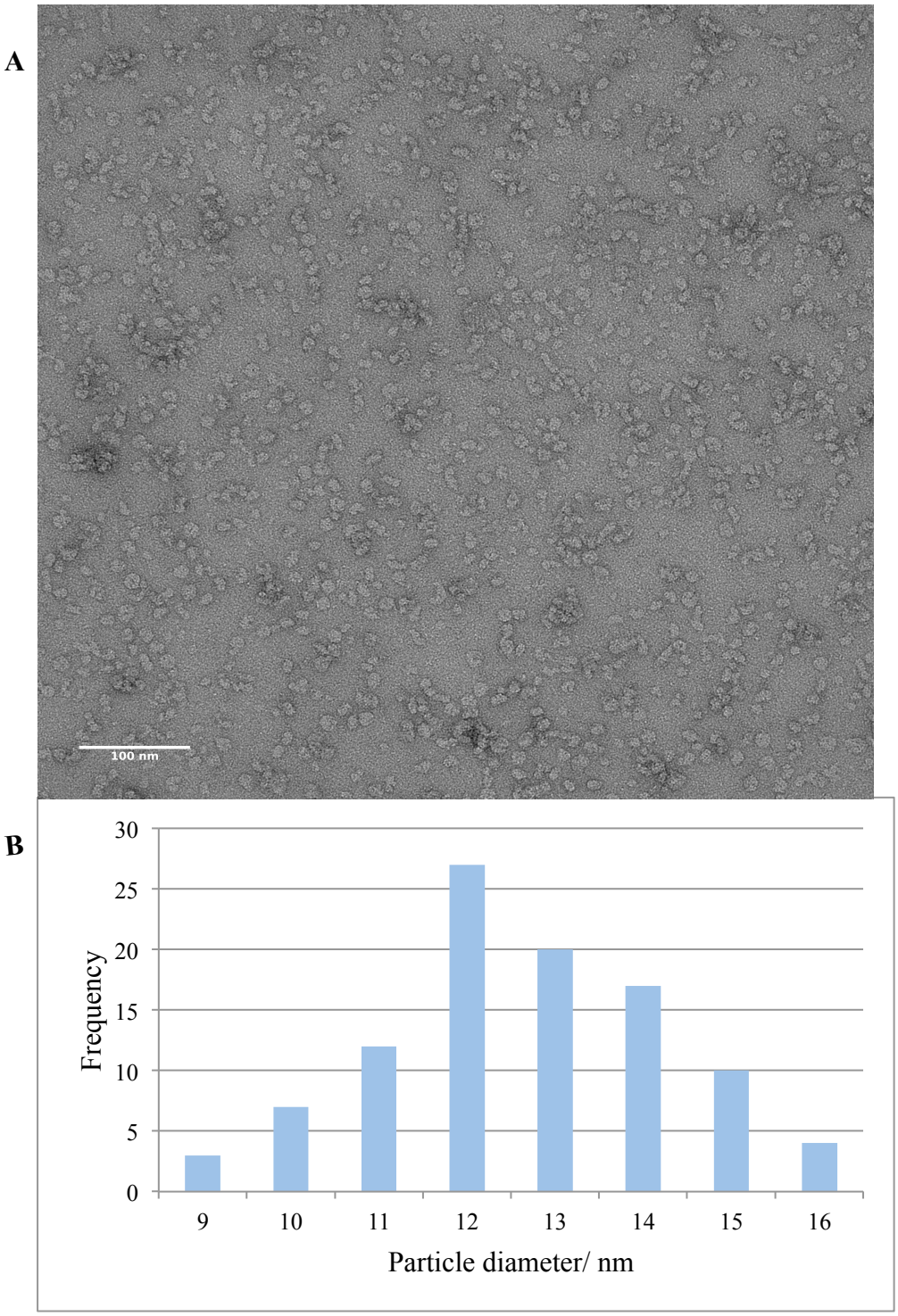


Figure 30. **Electron micrograph and size distribution analysis of TatAyCy.**

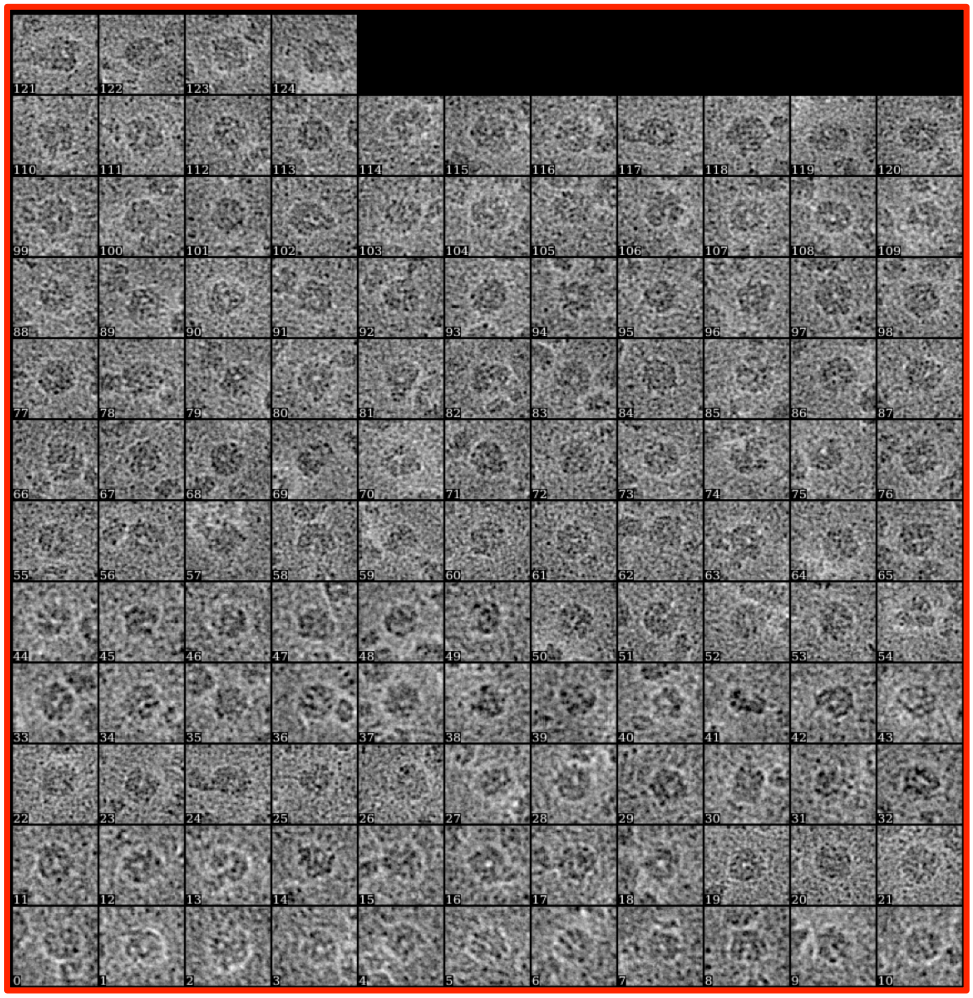
A. Negative stain electron micrograph of SMA and affinity purified TatAyCy sample.

B. Using ImageJ, the diameter of 100 random particles were measured and plotted onto a size distribution graph. Particles ranged in size from 9 to 16 nm, whilst the average particle size was 12.8nm.

5.2.4 Negative stain single particle analysis

Negative stain EM images were collected and analysed by Sarah Smith at the University of Warwick. Dr Smith also undertook data processing using Electron Micrograph Analysis (EMAN) (Ludtke et al. 1999). Individual images can often be noisy and hard to interpret, so post classification, similar images with constant features are aligned (Figure 31A) and titled (Figure 31B) to create the average TatAyCy particle. The protein complex generated has a doughnut like shape with a hole in the center.

A



B

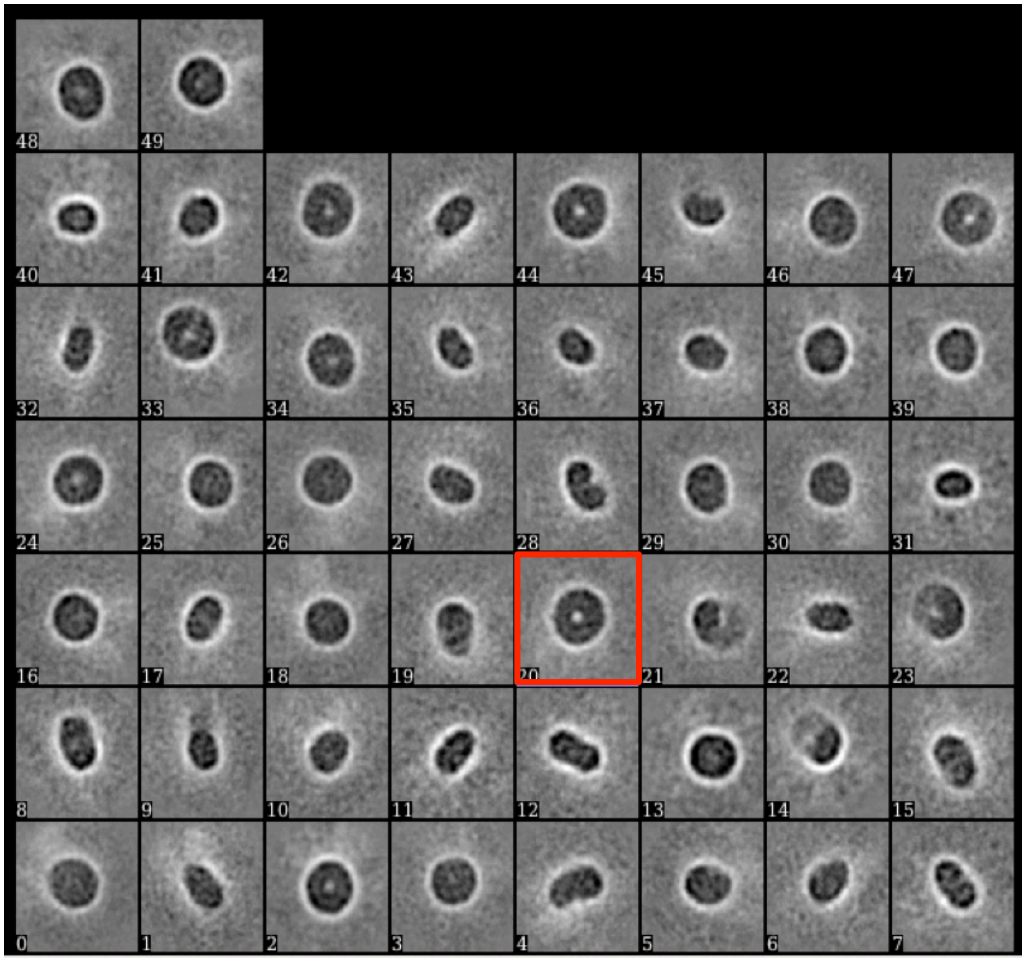


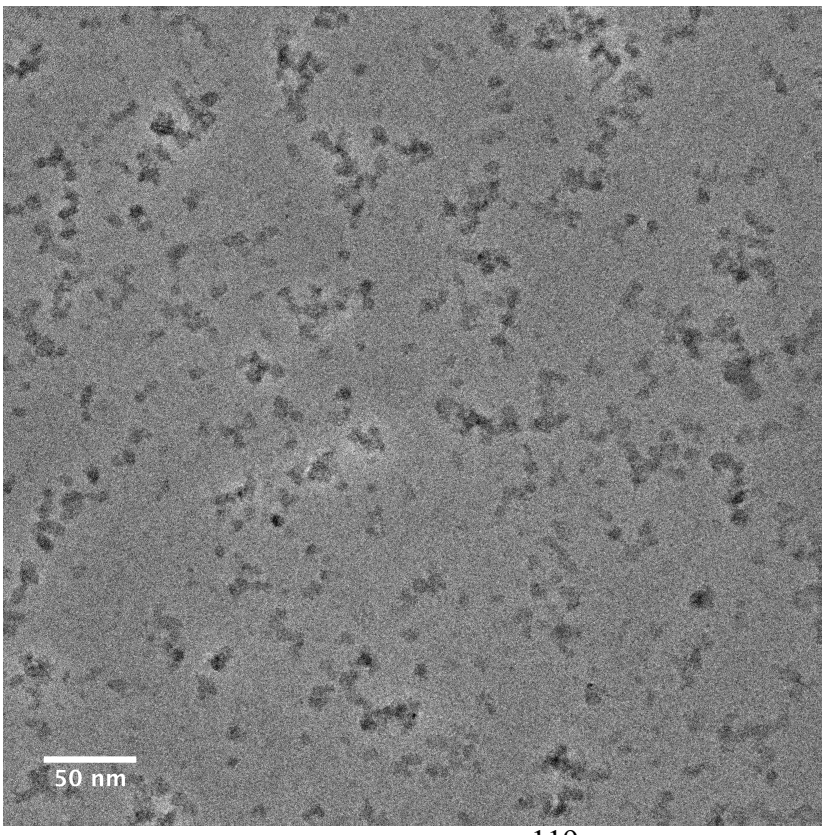
Figure 31. **Single particle reconstruction of TatAyCy from negative stain EM.**

A. Using EMAN, particles (purified with SMA and affinity chromatography) were collected and classified into similar orientations. An example grouping is displayed on p124. **B.** Reconstructions from different orientations were grouped to show the TatAyCy particle from different angles. The red square is the reconstruction from (A) above.

5.2.5 Cryo-EM images of TatAyCy

Given the success of TEM using negative staining, the next step towards gaining a more detailed structure of TatAyCy involved utilising cryogenic electron microscopy. Initially Sarah Smith commenced the Cryo-EM technique at the University of Warwick, but even after optomisation, results were poor. Optomisation methods for the small protein complex included making the ice particles very thin and freezing at 4^{o}C with >90% humidity. For better quality images, grid samples of TatAyCy were taken to the Krios center in Leicester by Dr Corrine Smith. The Titan Cryo-EM has 300 kV compared to 200 kV at the University of Warwick, the higher the voltage, the more energy the electrons have and the further they go into the sample to produce a higher resolution image. However, the quality of the data was still disappointing (Figure 32), the 2D classes only are on par with negative staining, but the data should be much darker. Moreover, there is a lot of heterogeneity in the data set and some background. The background could be imidazole

A



B

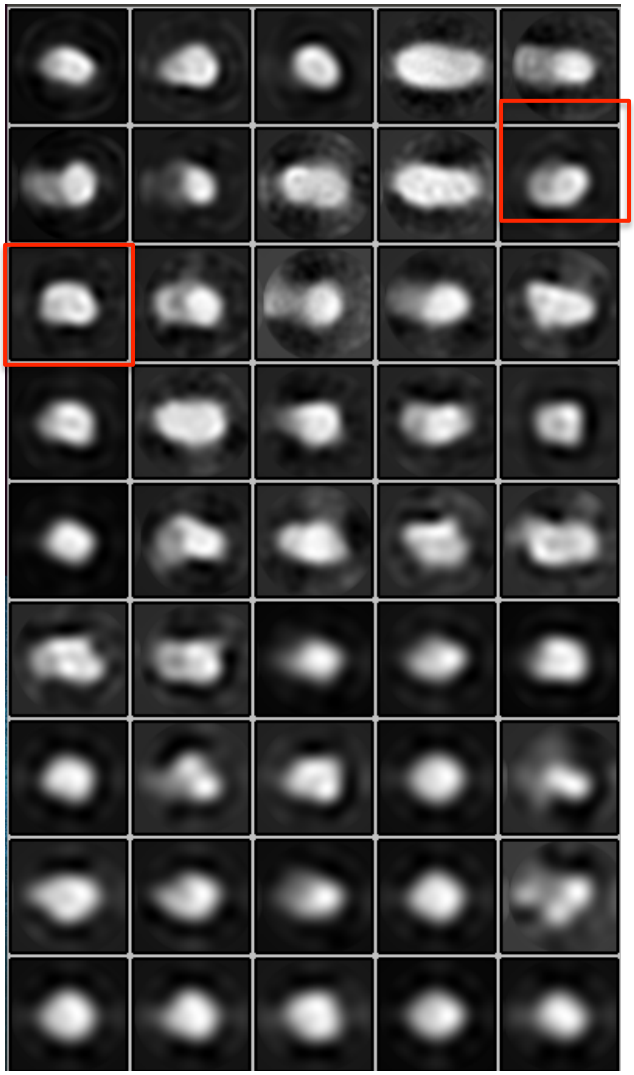


Figure 32. **Electron micrograph and 2D classifications of TatAyCy from Cryo-EM**.

A. Cryo Electron micrograph of of SMA and affinity purified TatAyCy sample.

B. TatAyCy sample 2D classifications from Cryo-EM which show a lot of heterogeneity is present in the data set. 5 classes (highlighted in red) were taken further for initial 3D model generation.

5.2.6 Initial TatAyCy 3D model generation

Using 5 classes from the heterogenous data set, Dr Smith created 3D reconstructions of TatAyCy. Particle occupancy highlighted a definite preferred particle orientation on the grid (top views), the sides of the protein are not visible. Figure 33A indicates angle coverage (more red is more coverage), blue dots should cover the sphere but there are blank sides. Figure 33B graph shows how many particles belong to each of the 5 classes, the plateauing line towards the end of the run demonstrates that the algorithm/projection-matching is “happy” with the particle placing for each 3D classification below. Class 2 was used as a reference in the final 3D model generation using 7,164 particles in Figure 31C.

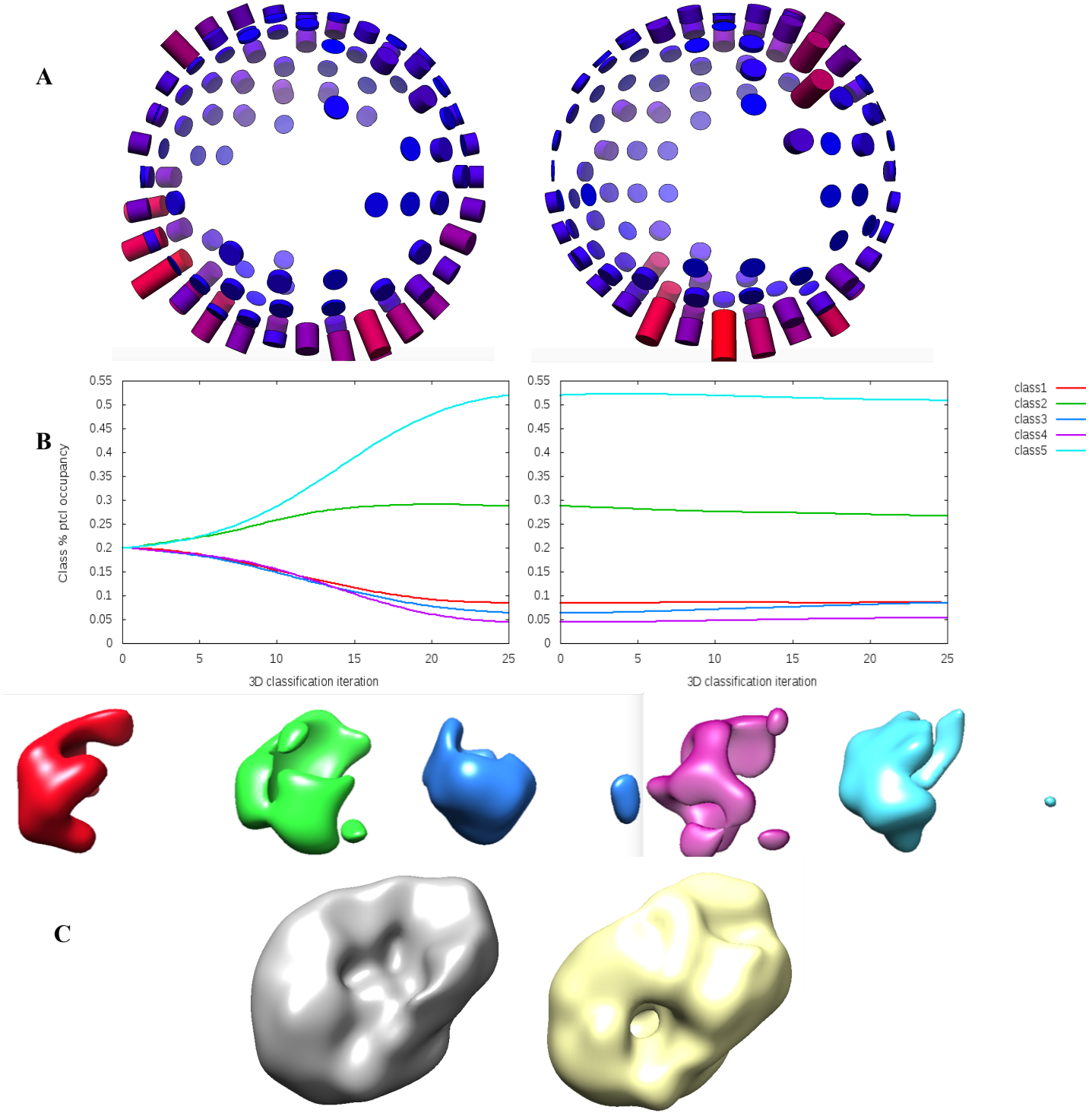
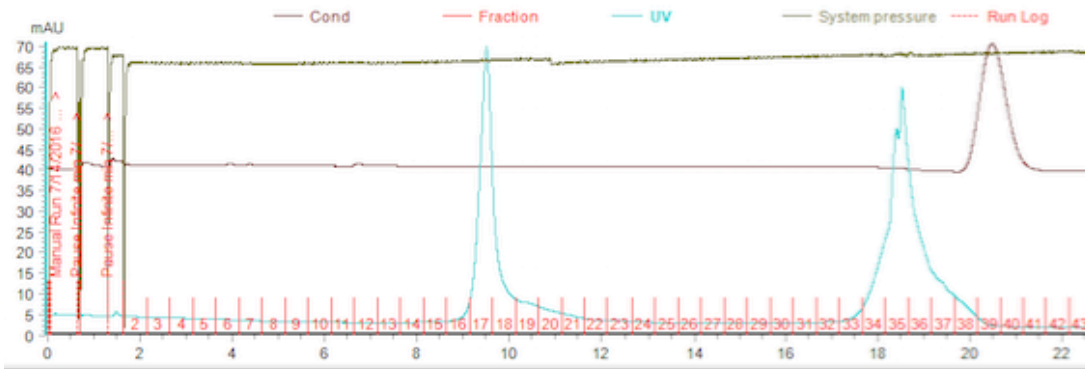


Figure 33.**Cryo-EM angle coverage, particle occupancy and 3D classification.**

A. Angle coverage indicates a lack of coverage at the sides of TatAyCy, blue dots should cover the sphere and red cylinders indicate best coverage, a preferred particle orientation is present. **B.** Class percentage particle occupancy for the 5 particle classes chosen and their corresponding 3D reconstructions. **C.** Class 2 was used as a reference in the final 3D model generation using 7,164 particles.

5.2.7 Evidence other Tat subunits can be purified by SMA

The optimisation of TatAyCy for cryo-EM analysis was a time consuming process, so we sought to only attempt the structural characterization of this complex during this study. Initially however other Tat proteins were cloned, expressed and purified using SMA for future analysis. Little is known about TatE (Section 1.3.2.2) but it shares 57% similarity to TatA, so this small protein (7 kDa) is interesting. Figure 32 shows TatE can also be purified by using SMA, affinity chromatography and size exclusion chromatography. In fact, about double the protein concentration of TatE was purified after SEC (Figure 34) which could result in better (less dilute) microscopy data in the future.



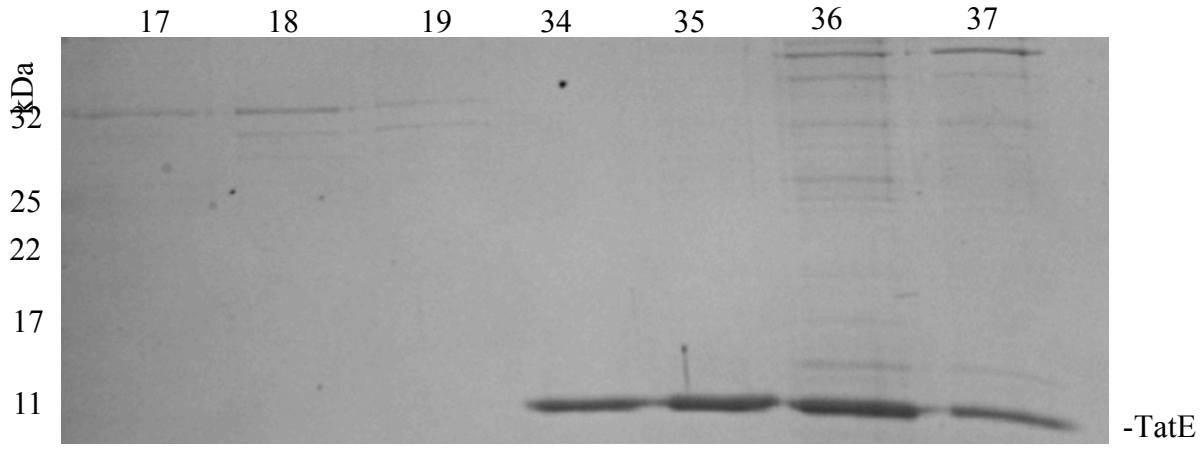


Figure 34. **Chromatogram and purification of SMAlp-purified TatE by SEC.** After SMA and affinity chromatography, TatE sample was loaded onto a Size Exclusion Column. Peak fractions, 17- 19 and 34-36 were loaded onto a SDS-gel for verification. Fractions 34-37 on the coomassie above were identified as TatE by western blot to the N-terminal 10X his tag (not shown)- fractions 34 and 35 are particularly pure.

5.2 Discussion

Most membrane protein complexes and all Tat complexes so far have been stabilised and purified using detergents. Previously, particular focus has fallen on the bigger Gram-negative TatABC complex (Gohlke et al. 2005; Ramasamy et al. 2013) so little is known about the structure of or function of the minimal Gram- positive TatAC complexes. The detergents DDM and digitonin have thus far been used to study the complex size of TatAyCy, and the non-ionic detergents form large micelles of 70 kDa. Not only does the micelle contribute to complex size heterogeneity and ambiguity, but the solubilisation also strips away any native phospholipid bilayer that might be otherwise supporting the protein *in vivo.*

In this study, we utilised a relatively new tool for membrane protein purification, styrene maleic acid (SMA) (Lee et al. 2016). The detergent free method exploits the hydrophilic and hydrophobic properties of SMA's copolymers to form a discoidal particle, which isolates proteins with their lipid bilayer to form SMA-lipid particles (SMALPs). The discoidal shape is size limited to about 15 nm in diameter, therefore the smaller Tat complexes, Gram- positive (~200 kDa) (Barnett et al. 2009) are more suited to this experiment. SMALP formation is not selective, but a 10X his tag on the N-terminal of each protein protrudes from the particle for ease of affinity chromatography.

Recent advances in Cryo-EM technology mean the visualisation of small protein complexes in high resolution (<500 kDa) is now more attainable (Rapisarda et al. 2018). Therefore we sought to combine the two methods to further characterise TatAyCy. The ideal concentration of TatAyCy complex was difficult to achieve for Cryo-EM. To purify the

sample further, which is clearly required as the heterogeneity observed in Figure 30 is not ideal, size exclusion chromatography is necessary. The experiment was continued using only affinity chromatography purified samples. In hindsight, perhaps protein dialysis would have been useful to remove the imidazole for less background noise.

Nonetheless, the preliminary data for SMALP-purified Tat is encouraging and is effective for both Tat complexes (TatAyCy) and individual subunits (TatE). The data presented in this study has reinforced the differences and similarities between Gram-positive and Gram-negative Tat systems. TatAyCy is much smaller than TatABC but both translocases have inherent heterogeneity, which was observed before in Sarah Smith's thesis. Structural insight could help discover Tat's export mechansim. Does TatAy form the translocase pore (Gohlke et al. 2005), weaken the phospholipid bilayer (Hou et al. 2018) or have another role? Like TatAd, the hole or pore seen in Figure 31 is simply too small to export the big proteins it can export (Beck et al. 2013). For example, one native substrate for TatAyCy, YwbN is 45 kDa but it can also export the recombinant protein GFP, which is 27 kDa.

Further optimisations are required to better understand the structure of TatAyCy. The next step would be to quantify the complex's composition, how many subunits are within translocase? Other than the low resolution data presented from cryo-EM, another popular problem arose, preferred particle preference (Figure 33) so the sides of TatAyCy were not visualised. Perhaps carbon coating the grid could increase the number of views during Cryo-EM and method paper's exist for just this problem (Zi Tan et al. 2017). It's also beneficial to understand that SMAlping is not the only non detergent protein purification method available, given more time, other tools such as nanodiscs and amphipols could be investigated too (Bayburt & Sligar 2010).

6.0 Final discussion

The transport of protein across a membrane is an essential cellular function of almost all organisms. Nearly 30 years go, the Twin Arginine Translocase pathway was discovered, since then homologs to the translocase have been identified in many archaea, bacteria, chloroplasts and mitochondria (Yen et al. 2002). Every protein within a cell partakes in specific roles; the Tat protein complex is unique in that it transports fully folded protein (often with cofactors) across the plasma membrane to their site of function.

Most knowledge regarding the Tat translocase originates from the Gram- negative *E. coli* system, which has 3 essential proteins (Berks 2015). TatA, TatB and TatC have all been studied in terms of both their structure and mode of function (reviewed in Chapter 1). Two main transportation mechanisms have been theorised and supported with scientific evidence. Initially, the pore forming mechanism was favored whereby the substrate binding dock Tat(A)BC (370 kDa) recruited separate TatA and this TatA complex formed a pore whose heterogenous size (100- 500 kDa) was dependent on substrate size (Oates et al. 2005). This theory was reinforced by low resolution EM images of TatA which showed ring shaped structures of varying diameter (9 - 13 nm) with a pore wide enough to accommodate one of the largest Tat substrates, TorA (90 kDa) (Gohlke et al. 2005). However, in light of new evidence, including that from Gram- positive Tat (Beck et al. 2013) a different hypothesis has prevailed (Hou et al. 2018), where upon substrate binding, TatA protomers cause membrane destabilization (Brüser & Sanders 2003) which allows transport of the protein substrate.

Tat's unqiue mechanism is not its only distinction as the translocase also has a quality control system that ensures only correctly folded and assmbled proteins are translocated (Robinson et al. 2011). Several studies mostly using *E. coli* Tat have probed this ability from both the substrate and translocase points of view, but the specifics of this process still remain unclear. Nevertheless, this capability is very advantagous to the biotechnology industry as recombinant proteins can be screened prior to periplasmic export which alleviates downstream purification requirements.

In collaboration with Medimmune (AstraZeneca), the aims of my project comprised of investgating the Gram- positive Tat translocase for biopharmaceutical exploitation, including its quality control feature. Given the results from this study (Chapter 3) I went on to explore one specific biotherapeutic, an scFv, in different orientations and conditions in conjunction with TatExpress and TatAdCd (Chapter 4). Clearly, similarities and differences existed between the Gram-negative and Gram-positive translocases and so my final research tested a new membrane protein purification technique (SMA) with state of the art Cryo-EM technology to structually characterise TatAyCy (Chapter 5).

The *Bacillus subtilis* TatAdCd system exhibits an extreme level of substrate specificity

B. subtilis Tat differs from *E. coli* by lacking TatB; a minimal TatAC system transports proteins across the plasma membrane. One such TatAC system, TatAdCd with only one native substrate, PhoD (Pop et al. 2002) was shown previously in an *E. coli tat* null background to export native TorA and TorA-GFP (Albiniak et al. 2013). Here, I examined its ability to export a range of biotherapeutics including hGH, IFN, VH domain and 2 different scFvs all with a TorA signal peptide. To our surprise, TatAdCd did not export the

majority of the proteins in this study including other native *E. coli* substrates but the minimal translocase did export both scFvs tested.

We therefore took this opportunity to examine the proof reading ability of a Gram-positive Tat translocase using scFv variants. To get a clearer view of what Tat 'sensed' as a mis-folded protein to reject, scFv variants were designed with additional salt bridges, altered charge or further hydrophobicity, without changing the structural state of the substrate. The *E. coli* TatABC tolerated and exported all variants to the periplasm with the aid of CyDisCo (Jones et al. 2016). In contrast, TatAdCd completely rejects many of the variants and preliminary suggests it cannot tolerate substrates above a certain hydrophobic threshold. This combined data shows the two systems have different substrate specificities and selectivity.

Auto Induction Media (AIM) aids Tat mediated scFv export to the periplasm

Given the success of scFv export via both TatABC and TatAdCd, the next study (Chapter 4) sought to scale up scFv production by Tat. The CEA scFv was different from the two in the previous study and was provided by MedImmune (Osbourn et al. 1996) in 8 different orientations to trial for expression and export with a Tat specific TorA signal peptide. In all media (LB, TB, SM6 and AIM) tested, only the V_H-V_L orientation expressed and the disulphide stabilised variant either made no difference or decreased protein expression.

The two scFvs in Chapter 3 had different CyDisCo requirements; scFvO did not require CyDisCo for export, but scFvM was CyDisCo dependent. No-CyDisCo, CyDisCo on an alterative plasmid and polycistronic CyDisCo expression were all tested with CEA scFv,

and the results showed CyDisCo expression increased expression and hence, export quantities.

The biggest surprise of this investigation was that only when grown in AIM did CEA scFv export to the periplasm. In the literature, both protein solubility and expression levels of proteins have risen using AIM, but I cannot find any literature demonstrating the ability of AIM to promote export of a protein that was otherwise not exporting to the periplasm. One can only assume the combination of customised induction (when each cell reached the right metabolic state) and Tat's unique mechanism are very compatible for export.

Detergent free, SMAlp purification and visualisation of TatAyCy

In pursuit of exploiting Tat for Biotechnology, we confirmed Gram-negative and Gram-positive Tat systems are different. Therefore, to structurally characterise another *B. subtilis* Tat translocase, TatAyCy, we used a novel method of membrane protein purification, styrene maleic acid lipid particles in conjunction with the latest Cryo-EM technologies. The purification of TatAyCy and TatE using SMA was very successful but a higher yield after SEC could have improved Cryo-EM data quality.

The data obtained from negative stain and cryo-EM showed TatAyCy is a much smaller, homogenous complex than TatABC. Like TatAdCd, TatAyCy has a 'pore' like structure, which is too small to transport the big proteins it's capable of exporting like TorA (Beck at al. 2013). This preliminary data therefore does not support the TatA pore forming mechanism but for more conclusive results (like how many subunits fit into a complex) higher resolution data without preferred particle orientation is required.

6.1 Future perspectives

Given the different substrate selectivities and specificities of TatAdCd and TatABC, it would be interesting to investigate alternative Tat systems. Would exploiting Agrobacterium, Rhizobia or Cyanobacterial Tat in *E. coli* produce a substrate selectivity pattern? How would the new Tat systems respond to recombinant protein production and export? Given that the work in this study has used a dual plasmid system (with TatAdCd on pBAD), I think the way forward would be cloning the different Tat systems onto the *E. coli* chromosome, as has been done with TatExpress cells.

As regards to the scFv variants, the next experiment I would like to test is whether TatAdCd can export the scFvs with CyDisCo on the same plasmid. If so, this would further add to the catalogue of scFvs it can export. It goes without saying that finding out why TatAdCd preferentially exports scFvs over smaller and less complicated biotherapeutics would be interesting. Testing export of similar format biotherapeutics for T cell receptors, scTv's (Aggen et al. 2012) could provide some answers. Checking for diabody formation in the periplasm for CEA 003 construct using a blue native gel would also confirm whether Tat can export diabodies.

Structural characterisation supports the process of fundamentally distinguishing the Gram-positive and Gram- negative Tat systems. This study has shown that non-detergent methods (SMA) can isolate the Tat system effectively. I believe this process, with optimisation, higher yields post SEC and during Cryo-EM could at least verify the number of subunits in TatAyCy complex. The golden ticket would be to trap a substrate within the protein complex or produce a functional assay to prove TatAyCy is active within the SMAlp.

This study has shown both Gram-negative and Gram- positive Tat have the potential to be used within the biotechnology industry. However, more research is required to understand how its quality control feature works (particularly in *B. subtilis*) and it would be interesting to understand how its export mechanism functioned. Recent advances in science, such as Cryo-EM, mean we're closer than ever to finding out the intricacies of individual cell building blocks.

7. References

Aggen, D.H., Chervin, A.S., Insaidoo, F.K., Piepenbrink, K.H., Baker., B.M., & Kranz, D.M. (2010). Identification and enginerring of human variable regoins that allow expression of stable single-chain T cell receptors. *Protein Engineering, Design & Selection, 24(4),* 361-372.

Ahmad, Z. A., Yeap, S. K., Ali, A. M., Ho, W. Y., Alitheen, N. B. M., & Hamid, M. (2012). ScFv antibody: Principles and clinical application. *Clinical and Developmental Immunology*.

Alami, M., Lüke, I., Deitermann, S., Eisner, G., Koch, H. G., Brunner, J., & Müller, M. (2003). Differential interactions between a twin-arginine signal peptide and its translocase in *Escherichia coli*. *Molecular Cell*, *12*(4), 937–946.

Alanen, H. I., Walker, K. L., Lourdes Velez Suberbie, M., Matos, C. F. R. O., Bönisch, S., Freedman, R. B., Keshavarz-Moore, E., Ruddock, L., & Robinson, C. (2014). Efficient export of human growth hormone, interferon α2b and antibody fragments to the periplasm by the *Escherichia coli* Tat pathway in the absence of prior disulfide bond formation. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1853*(3), 756–763.

Albiniak, A. M., Matos, C. F. R. O., 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. *The FEBS Journal*, *280*(16), 3810–3821.

Alcock, F., Baker, M. A. B., 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. *Proceedings of the National Academy of Sciences*, *110*(38)

Alcock, F., Stansfeld, P. J., Basit, H., Habersetzer, J., Baker, M. A., Palmer, T., Walace, M., & Berks, B. C. (2016). Assembling the tat protein translocase. *ELife*, *5,* 1-18.

Alder, N. N., & Theg, S. M. (2003). Energetics of protein transport across biological membranes: A study of the thylakoid ΔpH-dependent/cpTat pathway. *Cell*, *112*(2), 231–242. Aldridge, C., Ma, X., Gerard, F., & Cline, K. (2014). Substrate-gated docking of pore subunit Tha4 in the TatC cavity initiates Tat translocase assembly. *Journal of Cell Biology*, *205*(1), 51–65.

Aldridge, C., Storm, A., Cline, K., & Dabney-Smith, C. (2012). The chloroplast twin arginine transport (Tat) component, Tha4, undergoes conformational changes leading to Tat protein transport. *Journal of Biological Chemistry*, *287*(41),

Allen, W. J., Corey, R. A., Oatley, P., Sessions, R. B., Baldwin, S. A., Radford, S. E., Tuma, R., & Collinson, I. (2016). Two-way communication between SecY and SecA suggests a brownian ratchet mechanism for protein translocation. *ELife*, *5.*

Auclair, S. M., Bhanu, M. K., & Kendall, D. A. (2012). Signal peptidase I: Cleaving the way to mature proteins. *Protein Science*. *21* (1), 13-25.

Austerberry, J. I., Dajani, R., Panova, S., Roberts, D., Golovanov, A. P., Pluen, A., … Curtis, R. (2017). The effect of charge mutations on the stability and aggregation of a human single chain Fv fragment. *European Journal of Pharmaceutics and Biopharmaceutics*, *115*, 18–30.

Bageshwar, U. K., & Musser, S. M. (2007). Two electrical potential-dependent steps are required for transport by the *Escherichia coli* Tat machinery. *Journal of Cell Biology*, *179*(1), 87–99.

Baglieri, J., Beck, D., Vasisht, N., Smith, C. J., & Robinson, C. (2012). Structure of TatA paralog, TatE, suggests a structurally homogeneous form of Tat protein translocase that transports folded proteins of differing diameter. *Journal of Biological Chemistry*, *287*(10), 7335–7344.

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. *Journal of Biological Chemistry*, *283*(5), 2534–2542.

Barnett, J. P., Lawrence, J., Mendel, S., & Robinson, C. (2011). Expression of the bifunctional *Bacillus subtilis* TatAd protein in *Escherichia coli* reveals distinct TatA/B-family and TatB-specific domains. *Archives of Microbiology*, *193*(8), 583–594.

Barnett, J. P., Van Der Ploeg, R., Eijlander, R. T., Nenninger, A., Mendel, S., Rozeboom, R., Kuipers, O.P., van Dijl, JM., & Robinson, C. (2009). The twin-arginine translocation (Tat) systems from *Bacillus subtilis* display a conserved mode of complex organization and similar substrate recognition requirements. *FEBS Journal*, *276*(1), 232–243.

Barrett, C. M. L., Freudl, R., & Robinson, C. (2007). Twin arginine translocation (Tat)-dependent export in the apparent absence of TatABC or TatA complexes using modified *Escherichia coli* TatA subunits that substitute for TatB. *Journal of Biological Chemistry*, *282*(50), 36206–36213.

Barrett, C. M. L., Mathers, J. E., & Robinson, C. (2003). Identification of key regions within the *Escherichia coli* TatAB subunits. *FEBS Letters*, *537*(1–3), 42–46.

Barrett, C. M. L., Ray, N., Thomas, J. D., Robinson, C., & Bolhuis, A. (2003). Quantitative export of a reporter protein, GFP, by the twin-arginine translocation pathway in

Escherichia coli. *Biochemical and Biophysical Research Communications*, *304*(2), 279–284.

Bauer, B. W., Shemesh, T., Chen, Y., & Rapoport, T. A. (2014). A "push and slide" mechanism allows sequence-insensitive translocation of secretory proteins by the SecA ATPase. *Cell*, *157*(6), 1416–1429.

Baumgarten, T., Ytterberg, A. J., Zubarev, R. A., & De Gier, J.-W. (2018). Optimizing recombinant protein production in the *E. coli* periplasm alleviates stress 2 3, 1–12.

Bayhurt, T. & Sligar, SG. (2010). Membrane protein assembly into Nanodiscs. *FEBS Lett*, *584*(9), 1721-7.

Bechtluft, P., van Leeuwen, R. G. H., Tyreman, M., Tomkiewicz, D., Nouwen, N., Tepper, H. L., Driessen, A.J.M., & Tans, S. J. (2007). Direct observation of chaperone-induced changes in a protein folding pathway. *Science (New York, N.Y.)*, *318*(5855), 1458–1461.

Beck, D., Vasisht, N., Baglieri, J., Monteferrante, C. G., van Dijl, J. M., Robinson, C., & Smith, C. J. (2013). Ultrastructural characterisation of *Bacillus subtilis* TatA complexes suggests they are too small to form homooligomeric translocation pores. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1833*(8), 1811–1819.

Beck, K., Wu, L. F., Brunner, J., & Müller, M. (2000). Discrimination between SRP- and SecA/SecB-dependent substrates involves selective recognition of nascent chains by SRP and trigger factor. *The EMBO Journal*, *19*(1), 134–143.

Behrendt, J., & Brüser, T. (2014). The TatBC complex of the tat protein translocase in escherichia coli and its transition to the substrate-bound TatABC complex. *Biochemistry*, *53*(14), 2344–2354.

Behrendt, J., Standar, K., Lindenstrauß, U., & Brüser, T. (2004). Topological studies on the

twin-arginine translocase component TatC. *FEMS Microbiology Letters*, *234*(2), 303–308.

Berks, B. C. (1996). A common export pathway for proteins binding complex redox cofactors? *Molecular Microbiology*, *22*(3), 393–404.

Berks, B. C., Lea, S. M., & Stansfeld, P. J. (2014). Structural biology of tat protein transport. *Current Opinion in Structural Biology*, *27*(1), 32–37.

Berks, B. C., Palmer, T., & Sargent, F. (2003). The Tat protein translocation pathway and its role in microbial physiology. *Advances in Microbial Physiology*.

Berks, B. C., Sargent, F., & Palmer, T. (2000). The Tat protein export pathway. *Mol Microbiol.*, *35*(2), 260–274.

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*(5), 1171–1182.

Bernstein, H. D., Poritz, M. A., Strub, K., Hoben, P. J., Brenner, S., & Walter, P. (1989). Model for signal sequence recognition from amino-acid sequence of 54K subunit of signal recognition particle. *Nature*, *340*(6233), 482–486.

Bird, R. E., Hardman, K. D., Jacobson, J. W., Johnson, S., Kaufman, B. M., Lee, S. M., … Whitlow, M. (1988). Single-chain antigen-binding proteins. *Science*, *242*(4877), 423–426.

Blaudeck, N., Kreutzenbeck, P., Müller, M., Sprenger, G. A., & Freudl, R. (2005). Isolation and characterization of bifunctional *Escherichia coli* TatA mutant proteins that allow efficient Tat-dependent protein translocation in the absence of TatB. *Journal of Biological Chemistry*, *280*(5), 3426–3432.

Blümmel, A. S., Drepper, F., Knapp, B., Eimer, E., Warscheid, B., Müller, M., & Fröbel, J.

(2017). Structural features of the TatC membrane protein that determine docking and insertion of a twin-arginine signal peptide. *Journal of Biological Chemistry*, *292*(52), 21320–21329.

Blümmel, A. S., Haag, L. A., Eimer, E., Müller, M., & Fröbel, J. (2015). Initial assembly steps of a translocase for folded proteins. *Nature Communications*, *6*.

Bogsch, E., Brink, S., & Robinson, C. (1997). Pathway specificity for a ΔpH-dependent precursor thylakoid lumen protein is governed by a "Sec-avoidance" motif in the transfer peptide and a "Sec-incompatible" mature protein. *EMBO Journal*, *16*(13), 3851–3859.

Bogsch, E. G., Sargent, F., Stanley, N. R., Berks, B. C., Robinson, C., & Palmer, T. (1998). An essential component of a novel bacterial protein export system with homologues in plastids and mitochondria. *Journal of Biological Chemistry*, *273*(29), 18003–18006.

Bolhuis, A., Mathers, J. E., Thomas, J. D., Claire, M., Barrett, L., & Robinson, C. (2001). TatB and TatC Form a Functional and Structural Unit of the Twin-arginine Translocase from *Escherichia coli*. *Journal of Biological Chemistry*, *276*(23), 20213–20219.

Braun, N. A., Davis, A. W., & Theg, S. M. (2007). The chloroplast tat pathway utilizes the transmembrane electric potential as an energy source. *Biophysical Journal*, *93*(6), 1993–1998.

Broedel, S. E., Papciak, S. M., & Jones, W. R. (2001). The Selection of Optimum Media Formulations for Improved Expression of Recombinant Proteins In *E . coli*. *Athena Enzyme SystemsTM Group*, *2*, 8.

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. *Biotechnology and Bioengineering*, *114*(12), 2828–2836.

Bruch, M. D., McKnight, C. J., & Gierasch, L. M. (1989). Helix Formation and Stability in a Signal Sequence. *Biochemistry*, *28*(21), 8554–8561.

Brüser, T., & Sanders, C. (2003). An alternative model of the twin arginine translocation system. *Microbiological Research*, *158*(1), 7–17.

Buchanan, G., Sargent, F., Berks, B. C., & Palmer, T. (2002). A genetic screen for suppressors of *Escherichia coli* Tat signal peptide mutations establishes a critical role for the second arginine within the twin-arginine motif. *Archives of Microbiology*, *177*(1), 107–112.

Cao, T. B., & Saier, M. H. (2003). The general protein secretory pathway: Phylogenetic analyses leading to evolutionary conclusions. *Biochimica et Biophysica Acta - Biomembranes*, *1609*(1), 115–125.

Celedon, J. M., & Cline, K. (2012). Stoichiometry for binding and transport by the twin Arginine translocation system. *Journal of Cell Biology*, *197*(4), 523–534.

Chaddock, A. M., Mant, a, Karnauchov, I., Brink, S., Herrmann, R. G., Klösgen, R. B., & Robinson, C. (1995). A new type of signal peptide: central role of a twin-arginine motif in transfer signals for the delta pH-dependent thylakoidal protein translocase. *The EMBO Journal*, *14*(12), 2715–2722.

Chan, C. S., Chang, L., Rommens, K. L., & Turner, R. J. (2009). Differential interactions between tat-specific redox enzyme peptides and their chaperones. *Journal of Bacteriology*, *191*(7), 2091–2101.

Chan, C. S., Zlomislic, M. R., Tieleman, D. P., & Turner, R. J. (2007). The TatA subunit of *Escherichia coli* twin-arginine translocase has an N-in topology. *Biochemistry*, *46*(25),

7396–7404.

Clark, S. A., & Theg, S. M. (1997). A folded protein can be transported across the chloroplast envelope and thylakoid membranes. *Molecular Biology of the Cell*, *8*(5), 923–934.

Claassens, N. J., Siliakus, M. F., Spaans, S. K., Creutzburg, S. C. A., Nijsse, B., Schaap, P. J., … Van Der Oost, J. (2017). Improving heterologous membrane protein production in *Escherichia coli* by combining transcriptional tuning and codon usage algorithms. *PLoS ONE*, *12*(9)

Cline, K. (2015). Mechanistic aspects of folded protein transport by the twin arginine translocase (Tat). *Journal of Biological Chemistry*, *290*(27), 16530–16538.

Cline, K., & McCaffery, M. (2007). Evidence for a dynamic and transient pathway through the TAT protein transport machinery. *EMBO Journal*, *26*(13), 3039–3049.

Cline, K., & Mori, H. (2001). Thylakoid ΔpH-dependent precursor proteins bind to a cpTatC-Hcf106 complex before Tha4-dependent transport. *Journal of Cell Biology*, *154*(4), 719–729.

Cristóbal, S., De Gier, J. W., Nielsen, H., & Von Heijne, G. (1999). Competition between Sec- and TAT-dependent protein translocation in *Escherichia coli. EMBO Journal*, *18*(11), 2982–2990.

Dabney-Smith, C., Mori, H., & Cline, K. (2006). Oligomers of Tha4 organize at the thylakoid Tat translocase during protein transport. *Journal of Biological Chemistry*, *281*(9), 5476–5483.

DeLisa, M. P., Lee, P., Palmer, T., & Georgiou, G. (2004). Phage Shock Protein PspA of *Escherichia coli* Relieves Saturation of Protein Export via the Tat Pathway. *Journal of Bacteriology*, *186*(2), 366–373.

DeLisa, M. P., Samuelson, P., Palmer, T., & Georgiou, G. (2002). Genetic analysis of the twin arginine translocator secretion pathway in bacteria. *The Journal of Biological Chemistry*, *277*(33), 29825–29831.

Delisa, M. P., Tullman, D., & Georgiou, G. (2003). Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *PNAS*, *100*(10), 6115–6120.

Ding, Z., & Christie, P. J. (2003). *Agrobacterium tumefaciens* twin-arginine-dependent translocation is important for virulence, flagellation, and chemotaxis but not type IV secretion. *Journal of Bacteriology*, *185*(3), 760–771.

Drulyte, I., Johnson, R.M., Hesketh, E.L., Hurdiss, D.L., Scarff, C.A., Porav, S.A., Ranson, N.A., Muench., & Thompson, R.F. (2018). Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Cryst.*D74, 560-571.

Ecker, D. M., Jones, S. D., & Levine, H. L. (2015). The therapeutic monoclonal antibody market. *MAbs*. 7(1): 9-14

Edwardraja, S., Neelamegam, R., Ramadoss, V., Venkatesan, S., & Lee, S. G. (2010). Redesigning of anti-c-Met single chain Fv antibody for the cytoplasmic folding and its structural analysis. *Biotechnology and Bioengineering*, *106*(3), 367–375.

Egea, P. F., & Stroud, R. M. (2010). Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. *Proceedings of the National Academy of Sciences*, *107*(40), 17182–17187.

Eijlander, R. T., Jongbloed, J. D. H., & Kuipers, O. P. (2009). Relaxed specificity of the *Bacillus subtilis* TatAdCd translocase in Tat-dependent protein secretion. *Journal of Bacteriology*, *91*(1), 196–202.

Eijlander, R. T., Kolbusz, M. A., Berendsen, E. M., & Kuipers, O. P. (2009). Effects of

altered TatC proteins on protein secretion efficiency via the twin-arginine translocation pathway of *Bacillus subtilis*. *Microbiology*, *155*(6), 1776–1785.

Eimer, E., Fröbel, J., Blümmel, A. S., & Müller, M. (2015). TatE as a regular constituent of bacterial twin-arginine protein translocases. *Journal of Biological Chemistry*, *290*(49), 29281–29289.

Eimer, E., Kao, W. C., Fröbel, J., Blümmel, A. S., Hunte, C., & Müller, M. (2018). Unanticipated functional diversity among the TatA-type components of the Tat protein translocase. *Scientific Reports*, *8*(1), 1–12.

Farid, T. A., Kodali, G., Solomon, L. A., Lichtenstein, B. R., Sheehan, M. M., Fry, B. A., … Dutton, P. L. (2013). Elementary tetrahelical protein design for diverse oxidoreductase functions. *Nature Chemical Biology*, *9*(12), 826–833.

Finazzi, G., Chasen, C., Wollman, F. A., & De Vitry, C. (2003). Thylakoid targeting of Tat passenger proteins shows no ΔpH dependence in vivo. *EMBO Journal*, *22*(4), 807–815.

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*. *Microbial Biotechnology*, *1*, 403–415

Frain, K. M., Jones, A. S., Schoner, R., Walker, K. L., & Robinson, C. (2017). The *Bacillus subtilis* TatAdCd system exhibits an extreme level of substrate selectivity. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1864*(1), 202–208.

French, C., Keshavarz-Moore, E., & Ward, J. M. (1996). Development of a simple method for the recovery of recombinant proteins from the *Escherichia coli* periplasm. *Enzyme and Microbial Technology*, *19*(5), 332–338.

Fröbel, J., Rose, P., Lausberg, F., Blümmel, A.-S., Freudl, R., & Müller, M. (2012). Transmembrane insertion of twin-arginine signal peptides is driven by TatC and regulated by TatB. *Nature Communications*, *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 col*i in fed-batch fermentations on chemically defined minimal media. *Microbial Cell Factories*, *16(108)*

Gérard, F., & Cline, K. (2006). Efficient twin arginine translocation (Tat) pathway transport of a precursor protein covalently anchored to its initial cpTatC binding site. *Journal of Biological Chemistry*, *281*(10), 6130–6135.

Gérard, F., & Cline, K. (2007). The thylakoid proton gradient promotes an advanced stage of signal peptide binding deep within the Tat pathway receptor complex. *Journal of Biological Chemistry*, *282*(8), 5263–5272.

Glockshuber, R., Malia, M., Pfitzinger, I., & Plückthun, A. (1990). A Comparison of Strategies To Stabilize Immunoglobulin Fv-Fragments. *Biochemistry*, *29*(6), 1362–1367

Gohlke, U., Pullan, L., McDevitt, C. A., Porcelli, I., de Leeuw, E., Palmer, T., … Berks, B. C. (2005). The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(30), 10482–10486.

Goosens, V. J., De-San-Eustaquio-Campillo, A., Carballido-López, R., & van Dijl, J. M. (2015). A Tat ménage à trois - The role of *Bacillus subtilis* TatAc in twin-arginine protein translocation. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1853*(10), 2745–2753.

Goosens, V. J., Monteferrante, C. G., & Van Dijl, J. M. (2014). Co-factor insertion and disulfide bond requirements for twin-Arginine translocase-dependent export of the *Bacillus subtilis* Rieske protein QcrA. *Journal of Biological Chemistry*, *289*(19), 13124–13131.

Goosens, V. J., Monteferrante, C. G., & Van Dijl, J. M. (2014). The Tat system of Gram-positive bacteria. *Biochimica et Biophysica Acta - Molecular Cell Research*.

Gouffi, K., Gérard, F., Santini, C. L., & Wu, L. F. (2004). Dual Topology of the *Escherichia coli* TatA Protein. *Journal of Biological Chemistry*, *279*(12), 11608–11615.

Gouridis, G., Karamanou, S., Gelis, I., Kalodimos, C. G., & Economou, A. (2009). Signal peptides are allosteric activators of the protein translocase. *Nature*, *462*(7271), 363–367.

Grabski, A., Mehler, M., & Drott, D. (2005). The Overnight Express Autoinduction System: High-density cell growth and protein expression while you sleep. *Nature methods* 2, 233-235.

Greene, N. P., Porcelli, I., Buchanan, G., Hicks, M. G., Schermann, S. M., Palmer, T., & Berks, B. C. (2007). Cysteine scanning mutagenesis and disulfide mapping studies of the TatA component of the bacterial twin arginine translocase. *Journal of Biological Chemistry*, *282*(33), 23937–23945.

Gulati, S., Jamshad, M., Knowles, T. J., Morrison, K. A., Downing, R., Cant, N., Collins,R., Koenderink, J.B., Ford, R.C.,Overduin, M., Kerr, I.D., Dafforn, T.R., & Rothnie, A. J. (2014). Detergent-free purification of ABC (ATP-binding-cassette) transporters. *Biochemical Journal*, *461*(2), 269–278.

Gupta, S. K., & Shukla, P. (2017). Microbial platform technology for recombinant antibody

fragment production: A review. *Critical Reviews in Microbiology*, *43*(1), 31–42.

Habersetzer, J., Moore, K., Cherry, J., Buchanan, G., Stansfeld, P. J., & Palmer, T. (2017). Substrate-triggered position switching of TatA and TatB during Tat transport in *Escherichia coli*. *Open Biology*, *7*(8).

Halbig, D., Hou, B., Freudl, R., Sprenger, G. A., & Klösgen, R. B. (1999). Bacterial proteins carrying twin-R signal peptides are specifically targeted by the ΔpH-dependent transport machinery of the thylakoid membrane system. *FEBS Letters*, *447*(1), 95–98.

Halbig, D., Wiegert, T., Blaudeck, N., Freudl, R., & Sprenger, G. A. (1999). The efficient export of NADP-containing glucose-fructose oxidoreductase to the periplasm of *Zymomonas mobilis* depends both on an intact twin- arginine motif in the signal peptide and on the generation of a structural export signal induced by cofactor bind. *European Journal of Biochemistry*, *263*(2), 543–551.

Hanada, M., Nishiyama, K., & Tokuda, H. (1996). SecG plays a critical role in protein translocation in the absence of the proton motive force as well as at low temperature. *FEBS Letters*, *381*(1–2), 25–28.

Harris, C. R., & Silhavy, T. J. (1999). Mapping an interface of SecY (PrlA) and SecE (PrlG) by using synthetic phenotypes and in vivo cross-linking. *Journal of Bacteriology*, *181*(11), 3438–3444.

Harrison, J. J., Ceri, H., Badry, E. a, Roper, N. J., Tomlin, K. L., & Turner, R. J. (2005). Effects of the twin-arginine translocase on the structure and antimicrobial susceptibility of *Escherichia coli* biofilms. *Canadian Journal of Microbiology*, *51*, 671–683.

Hartl, F. U., Lecker, S., Schiebel, E., Hendrick, J. P., & Wickner, W. (1990). The binding

cascade of SecB to SecA to SecY E mediates preprotein targeting to the *E. coli* plasma membrane. *Cell*, *63*(2), 269–279.

Hatahet, F., Nguyen, V. D., Salo, K. E. H., & Ruddock, L. W. (2010). Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of E. coli. *Microbial Cell Factories*, *9*, 67.

Hauer, R. S., Schlesier, R., Heilmann, K., Dittmar, J., Jakob, M., & Klösgen, R. B. (2013). Enough is enough: TatA demand during Tat-dependent protein transport. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1833*(5), 957–965.

Hicks, M. G., De Leeuw, E., Porcelli, I., Buchanan, G., Berks, B. C., & Palmer, T. (2003). The *Escherichia coli* twin-arginine translocase: Conserved residues of TatA and TatB family components involved in protein transport. *FEBS Letters*, *539*(1–3), 61–67.

Hinsley, A. P., Stanley, N. R., Palmer, T., & Berks, B. C. (2001). A naturally occurring bacterial Tat signal peptide lacking one of the "invariant" arginine residues of the consensus targeting motif. *FEBS Letters*, *497*(1), 45–49.

Holland, I. B. (2004). Translocation of bacterial proteins - An overview. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1694*(1-3), 5-16.

Holzapfel, E., Eisner, G., Alami, M., Barrett, C. M. L., Buchanan, G., Lüke, I., … Müller, M. (2007). The entire N-terminal half of TatC is involved in twin-arginine precursor binding. *Biochemistry*, *46*(10), 2892–2898.

Hou, B., Heidrich, E. S., Mehner-Breitfeld, D., & Brüser, T. (2018). The TatA component of the twin-arginine translocation system locally weakens the cytoplasmic membrane of *Escherichia coli* upon protein substrate binding. *Journal of Biological Chemistry*, (17), 293, 7592-7605.

Hu, Y., Zhao, E., Li, H., Xia, B., & Jin, C. (2010). Solution NMR structure of the TatA component of the twin-arginine protein transport system from gram-positive bacterium *Bacillus subtilis*. *Journal of the American Chemical Society*, *132*(45), 15942–15944.

Hynds, P. J., Robinson, D., & Robinson, C. (1998). The Sec-independent twin-arginine translocation system can transport both tightly folded and malfolded proteins across the thylakoid membrane. *Journal of Biological Chemistry*, *273*(52), 34868–34874.

Ize, B., Stanley, N. R., Buchanan, G., & Palmer, T. (2003a). Role of the *Escherichia coli* Tat pathway in outer membrane integrity. *Molecular Microbiology*, *48*(5), 1183–1193.

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. *Journal of Bacteriology*, *183*(5), 1801–1804.

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. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1863*(12), 3116–3124.

Jones, A.S. (2017). Proofreading of substrate by the Escherichia coli Twin Arginine Translocase. Doctor of Philosophy (PhD) thesis, University of Kent.

Jongbloed, J. D. H., Grieger, U., Antelmann, H., Hecker, M., Nijland, R., Bron, S., & Van Dijl, J. M. (2004). Two minimal Tat translocases in *Bacillus*. *Molecular Microbiology*, *54*(5), 1319–1325.

Jongbloed, J. D. H., Martin, U., Antelmann, H., Hecker, M., Tjalsma, H., Venema, G., Bron, S., Van Dijl, J. M ., & Müllere, J. (2000). TatC is a specificity determinant for

protein secretion via the twin-arginine translocation pathway. *Journal of Biological Chemistry*, *275*(52), 41350–41357.

Jozala, A. F., Geraldes, D. C., Tundisi, L. L., Feitosa, V. de A., Breyer, C. A., Cardoso, S. L., … Pessoa, A. (2016). Biopharmaceuticals from microorganisms: from production to purification. *Brazilian Journal of Microbiology*, *47*, 51–63.

Kihara, a, Akiyama, Y., & Ito, K. (1995). FtsH is required for proteolytic elimination of uncomplexed forms of SecY, an essential protein translocase subunit. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(May), 4532–4536.

Kim, J. Y., Kim, Y.-G. G., & Lee, G. M. (2012). CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Applied Microbiology and Biotechnology*, *93*(3), 917–930.

Kipping, M., Lilie, H., Lindenstrauß, U., Andreesen, J. R., Griesinger, C., Carlomagno, T., & Brüser, T. (2003). Structural studies on a twin-arginine signal sequence. *FEBS Letters*, *550*(1–3), 18–22.

Klösgen, R. B., Brock, I. W., Herrmann, R. G., & Robinson, C. (1992). Proton gradient-driven import of the 16 kDa oxygen-evolving complex protein as the full precursor protein by isolated thylakoids. *Plant Molecular Biology*, *18*(5), 1031–1034.

Kneuper, H., Maldonado, B., Jäger, F., Krehenbrink, M., Buchanan, G., Keller, R., Muller, M., Berks., B.C., & Palmer, T. (2012). Molecular dissection of TatC defines critical regions essential for protein transport and a TatB-TatC contact site. *Molecular Microbiology*, *85*(5), 945–961.

Koch, S., Fritsch, M. J., Buchanan, G., & Palmer, T. (2012). *Escherichia coli* TatA and TatB proteins have N-out, C-in topology in intact cells. *Journal of Biological*

Chemistry, *287*(18), 14420–14431.

Kreutzenbeck, P., Kröger, C., Lausberg, F., Blaudeck, N., Sprenger, G. A., & Freudl, R. (2007). *Escherichia coli* twin arginine (Tat) mutant translocases possessing relaxed signal peptide recognition specificities. *Journal of Biological Chemistry*, *282*(11), 7903–7911.

Kuhn, P., Weiche, B., Sturm, L., Sommer, E., Drepper, F., Warscheid, B., Sourjik., V., & Koch, H. G. (2011). The Bacterial SRP Receptor, SecA and the Ribosome Use Overlapping Binding Sites on the SecY Translocon. *Traffic*, *12*(5), 563–578.

Kumamoto, C. A. (1989). *Escherichia coli* SecB protein associates with exported protein precursors in vivo. *Proceedings of the National Academy of Sciences*, *86*(14), 5320–5324.

Lange, C., Müller, S. D., Walther, T. H., Bürck, J., & Ulrich, A. S. (2007). Structure analysis of the protein translocating channel TatA in membranes using a multi-construct approach. *Biochimica et Biophysica Acta - Biomembranes*, *1768*(10), 2627–2634.

Lebendiker, M., & Danieli, T. (2014). Production of prone-to-aggregate proteins. *FEBS Letters*. 588(2): 236-46.

Lee, P. A., Buchanan, G., Stanley, N. R., Berks, B. C., & Palmer, T. (2002). Truncation analysis of TatA and TatB defines the minimal functional units required for protein translocation. *Journal of Bacteriology*, *184*(21), 5871–5879.

Lee, P. A., Orriss, G. L., Buchanan, G., Greene, N. P., Bond, P. J., Punginelli, C., Jack, R.L., Sansom, M.S.P., Berks, B.C., & Palmer, T. (2006). Cysteine-scanning mutagenesis and disulfide mapping studies of the conserved domain of the twin-arginine translocase TatB component. *Journal of Biological Chemistry*, *281*(45),

34072–34085.

Lee, S. C., Knowles, T. J., Postis, V. L. G., Jamshad, M., Parslow, R. A., Lin, Y. P., Goldman, A., Sridhar, P., Overduin, M., Muench, S.P., & Dafforn, T. R. (2016). A method for detergent-free isolation of membrane proteins in their local lipid environment. *Nature Protocols*, *11*(7), 1149–1162.

Lill, R., Cunningham, K., Brundage, L. A., Ito, K., Oliver, D., & Wickner, W. (1989). SecA protein hydrolyzes ATP and is an essential component of the protein translocation ATPase of *Escherichia coli*. *The EMBO Journal*, *8*(3), 961–966.

Lill, R., Dowhan, W., & Wickner, W. (1990). The ATPase activity of secA is regulated by acidic phospholipids, secY, and the leader and mature domains of precursor proteins. *Cell*, *60*(2), 271–280.

Lindenstrauß, U., & Brüser, T. (2009). Tat transport of linker-containing proteins in *Escherichia coli. FEMS Microbiology Letters*, *295*(1), 135–140.

Ludtke, S., Baldwin, P., Chui, W. EMAN: semiautomated sofware forhigh-resolution-single-particle-reconstructions. J. Sruct. Biol, 128, 82-97.

Lüke, I., Handford, J. I., Palmer, T., & Sargent, F. (2009). Proteolytic processing of *Escherichia coli* twin-arginine signal peptides by LepB. *Archives of Microbiology*, *191*(12), 919–925.

Lundbaek, J. A., Koeppe, R. E., & Andersen, O. S. (2010). Amphiphile regulation of ion channel function by changes in the bilayer spring constant. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(35), 15427–15430.

Ma, X., & Cline, K. (2010). Multiple precursor proteins bind individual Tat receptor complexes and are collectively transported. *EMBO Journal*, *29*(9), 1477–1488.

Maillard, A., Chen, K., & Duong, F. (2005). Preprotein translocation through the Sec translocon in bacteria. *Protein Movement Across Membranes, Eichler J (Ed)*, 1–13.

Matos, C. F. R. O., 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 Reports*, *10*(5), 474–479.

Matos, C. F. R. O., Robinson, C., Alanen, H. I., Prus, P., Uchida, Y., Ruddock, L. W., Freedman., R., & Keshavarz-Moore, E. (2014). Efficient export of prefolded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in *Escherichia coli* CyDisCo strains. *Biotechnology Progress*, *30*(2), 281–290.

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*(15), 2055–2063.

Matsumoto, G., Mori, H., & Ito, K. (1998). Roles of SecG in ATP- and SecA-dependent protein translocation. *Proc Natl Acad Sci U S A*, *95*(23), 13567–13572.

Mattanovich, D., Branduardi, P., Dato, L., Gasser, B., Sauer, M., & Porro, D. (2012). Recombinant Protein Production in Yeasts. In *Recombinant Gene Expression*, 824 329–358.

Maurer, C., Panahandeh, S., Jungkamp, A.-C., Moser, M., & Müller, M. (2010). TatB functions as an oligomeric binding site for folded Tat precursor proteins. *Molecular Biology of the Cell*, *21*(23), 4151–4161.

Mendel, S., McCarthy, A., Barnett, J. P., Eijlander, R. T., Nenninger, A., Kuipers, O. P., & Robinson, C. (2008). The *Escherichia coli* TatABC System and a *Bacillus subtilis* TatAC-type System Recognise Three Distinct Targeting Determinants in Twin-arginine Signal Peptides. *Journal of Molecular Biology*, *375*(3), 661–672.

Monteferrante, C. G., Baglieri, J., Robinson, C., & van Dijl, J. M. (2012). TatAc, the third TatA subunit of *Bacillus subtilis,* can form active twin-arginine translocases with the TatCd and TatCy subunits. *Applied and Environmental Microbiology*, *78*(14), 4999–5001.

Monteferrante, C. G., Miethke, M., Van Der Ploeg, R., Glasner, C., & Van Dijl, J. M. (2012). Specific targeting of the metallophosphoesterase YkuE to the *Bacillus* cell wall requires the twin-arginine translocation system. *Journal of Biological Chemistry*, *287*(35), 29789–29800.

Mori, H., & Cline, K. (2002). A twin arginine signal peptide and the pH gradient trigger reversible assembly of the thylakoid ΔpH/Tat translocase. *Journal of Cell Biology*, *157*(2), 205–210.

Mori, H., Summer, E. J., Ma, X., & Cline, K. (1999). Component specificity for the thylakoidal Sec and ΔpH-dependent protein transport pathways. *Journal of Cell Biology*, *146*(1), 45–55.

Mould, R. M., & Robinson, C. (1991). A proton gradient is required for the transport of two lumenal oxygen-evolving proteins across the thylakoid membrane. *Journal of Biological Chemistry*, *266*(19), 12189–12193.

Mould, R. M., Shackleton, J. B., & Robinson, C. (1991). Transport of proteins into chloroplasts. Requirements for the efficient import of two lumenal oxygen-evolving complex proteins into isolated thylakoids. *Journal of Biological Chemistry*, *266*(26), 17286–17289.

Müller, S. D., De Angelis, A. A., Walther, T. H., Grage, S. L., Lange, C., Opella, S. J., & Ulrich, A. S. (2007). Structural characterization of the pore forming protein TatAd of the twin-arginine translocase in membranes by solid-state 15N-NMR. *Biochimica et*

Biophysica Acta - Biomembranes, *1768*(12), 3071–3079.

Murphy, C. K., & Beckwith, J. (1994). Residues essential for the function of SecE, a membrane component of the *Escherichia coli* secretion apparatus, are located in a conserved cytoplasmic region. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(7), 2557–2561.

Musser, S. M., & Theg, S. M. (2000). Proton transfer limits protein translocation rate by the thylakoid ΔpH/Tat machinery. *Biochemistry*, *39*(28), 8228–8233.

Nagano, N., Ota, M., & Nishikawa, K. (1999). Strong hydrophobic nature of cysteine residues in proteins. *FEBS Letters*, *458*(1), 69–71.

Nguyen, V. D., Hatahet, F., Salo, K. E. H., 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*. *Microbial Cell Factories*, *10*.

Nolandt, O. V., Walther, T. H., Roth, S., Bürck, J., & Ulrich, A. S. (2009). Structure analysis of the membrane protein TatCd from the Tat system of *B. subtilis* by circular dichroism. *Biochimica et Biophysica Acta - Biomembranes*, *1788*(10), 2238–2244.

Oates, J., Barrett, C. M. L., Barnett, J. P., Byrne, K. G., Bolhuis, A., & Robinson, C. (2005). The *Escherichia coli* twin-arginine translocation apparatus incorporates a distinct form of TatABC complex, spectrum of modular TatA complexes and minor TatAB complex. *Journal of Molecular Biology*, *346*(1), 295–305.

Orfanoudaki, G., & Economou, A. (2014). Proteome-wide Subcellular Topologies of *E. coli* Polypeptides Database (STEPdb). *Molecular & Cellular Proteomics*, *13*(12), 3674–3687.

Orriss, G. L., Tarry, M. J., Ize, B., Sargent, F., Lea, S. M., Palmer, T., & Berks, B. C.

(2007). TatBC, TatB, and TatC form structurally autonomous units within the twin arginine protein transport system of *Escherichia coli*. *FEBS Letters*, *581*(21), 4091–4097.

Osbourn, J. K., Field, A., Wilton, J., Derbyshire, E., Earnshaw, J. C., Jones, P. T., … McCafferty, J. (1996). Generation of a panel of related human scFv antibodies with high affinities for human CEA. *Immunotechnology*, *2*(3), 181–196.

Overton, T. W. (2014). Recombinant protein production in bacterial hosts. *Drug Discovery Today*, *19*(5), 590–601.

Pal, D., Fite, K., & Dabney-Smith, C. (2013). Direct Interaction between a Precursor Mature Domain and Transport Component Tha4 during Twin Arginine Transport of Chloroplasts. *Plant Physiology*, *161*(2), 990–1001.

Palmer, T., Sargent, F., & Berks, B. C. (2005). Export of complex cofactor-containing proteins by the bacterial Tat pathway. *Trends in Microbiology*, *13*(4), 175–180.

Palmer, T., Sargent, F., & Berks, B. C. (2011). The Tat protein export pathway. In EcoSal-*Escherichia coli and Salmonella: Cellular and Molecular Biology.* (D. Eds Böck, A., Curtiss III, R., Kaper, J.B., Karp, P.D., Neidhardt, F.C., Nyström, T., Slauch, J.M., Squires, C.L. and Ussery, Ed.). ASM Press.

Panahandeh, S., Maurer, C., Moser, M., DeLisa, M. P., & Müller, M. (2008). Following the path of a twin-arginine precursor along the TatABC translocase of *Escherichia coli*. *Journal of Biological Chemistry*, *283*(48), 33267–33275.

Park, E., & Rapoport, T. A. (2012). Mechanisms of Sec61/SecY-Mediated Protein Translocation Across Membranes. *Annual Review of Biophysics*, *41*(1), 21–40.

Patel, R., Vasilev, C., Beck, D., Monteferrante, C. G., Van Dijl, J. M., Hunter, C. N., Smith, C., & Robinson, C. (2014). A mutation leading to super-assembly of twin-

arginine translocase (Tat) protein complexes. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1843*(9), 1978–1986.

Pecher, P., & Arnold, U. (2009). The effect of additional disulfide bonds on the stability and folding of ribonuclease A. *Biophysical Chemistry*, *141*(1), 21–28.

Pooley, H. M., Merchante, R., & Karamata, D. (1996). Overall Protein Content and Induced Enzyme Components of the Periplasm of *Bacillus subtilis*. *Microbial Drug Resistance*, *2*(1), 9–15.

Pop, O., Martin, U., Abel, C., & Müller, J. P. (2002). The twin-arginine signal peptide of PhoD and the TatAd/Cd proteins of *Bacillus subtilis* form an autonomous tat translocation system. *Journal of Biological Chemistry*, *277*(5), 3268–3273.

Porcelli, I., De Leeuw, E., Wallis, R., Van den Brink-van der Laan, E., De Kruijff, B., Wallace, B. A., Palmer, T., & Berks, B. C. (2002). Characterization and membrane assembly of the TatA component of the *Escherichia coli* twin-arginine protein transport system. *Biochemistry*, *41*(46), 13690–13697.

Postis, V., Rawson, S., Mitchell, J. K., Lee, S. C., Parslow, R. A., Dafforn, T. R., Bauldwin, S.A., & Muench, S. P. (2015). The use of SMALPs as a novel membrane protein scaffold for structure study by negative stain electron microscopy. *Biochimica et Biophysica Acta - Biomembranes*, *1848*(2), 496–501.

Ramasamy, S., Abrol, R., Suloway, C. J. M., & Clemons, W. M. (2013). The glove-like structure of the conserved membrane protein tatc provides insight into signal sequence recognition in twin-arginine translocation. *Structure*, *21*(5), 777–788.

Randall, L. & Hardy, S. (1986). Correlation of competence for export with lack of tertiary structure of the mature species: a study in vivo of maltose-binding protein in *E. coli*. *Cell*, *46*, 921–928.

Rapisarda, C., Tassinari, M., Gubellini, F., & Fronzes, R (2018). Using Cyro-EM to investigate bacteria secretion systems. *Annual review of microbiology, 73.*

Richter, S., & Brüser, T. (2005). Targeting of unfolded PhoA to the TAT translocon of *Escherichia coli*. *Journal of Biological Chemistry*, *280*(52), 42723–42730.

Richter, S., Lindenstrauss, U., Lücke, C., Bayliss, R., & Brüser, T. (2007). Functional tat transport of unstructured, small, hydrophilic proteins. *Journal of Biological Chemistry*, *282*(46), 33257–33264.

Ridder, A.N.J.A., de Jong, E.M., Jongbloed, J.D.H., & Kuipers, OP (2009). Subcellular localization of TatAd of *Bacillus subtilis* depends on the presence of TatCd or TatCy. *Journal of bacteriology*, *191*(13). 4410-4418.

Ribnicky, B., Van Blarcome, T., & Georgiou, G. (2007). A scFv Antibody Mutant Isolated in a Genetic Screen for Improved Export via the Twin Arginine Transporter Pathway Exhibits Faster Folding. *J Mol Biol*, *369*(3), 631–639.

Robinson, C., Matos., C.F., Beck, D., Ren, C., Lawrence, J., Vasisht, N., & Mendel, S. (2011). Transport and proofreading of proteins by the twin-arginine translocation (Tat) pathway in bacteria. *Biochim Biophys Acta, 1808 (3)*, 876-84.

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*(33), 13392–13397.

Rodriguez, F., Rouse, S. L., Tait, C. E., Harmer, J., De Riso, A., Timmel, C. R., Sansom, M., 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*(12), E1092-101.

Rollauer, S. E., Tarry, M. J., Graham, J. E., Jääskeläinen, M., Jäger, 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., Palmer, T., Högbom, M., Berks, B.C., & Lea, S. M. (2012). Structure of the TatC core of the twin-arginine protein transport system. *Nature*, *492*(7428), 210–214.

Rose, P., Fröbel, J., Graumann, P. L., & Müller, M. (2013). Substrate-Dependent Assembly of the Tat Translocase as Observed in Live *Escherichia coli* Cells. *PLoS ONE*, *8*(8), 1–17.

Saaf, A., Monne, M., deGier, J. W., & von Heijne, G. (1998). Membrane topology of the 60-kDa Oxalp homologue from *Escherichia coli*. *Journal of Biological Chemistry*, *273*(46), 30415–30418.

Sachelaru, I., Petriman, N. A., Kudva, R., Kuhn, P., Welte, T., Knapp, B., Drepper, F., Warscheid, B., & Koch, H. G. (2013). YidC occupies the lateral gate of the SecYEG translocon and is sequentially displaced by a nascent membrane protein. *Journal of Biological Chemistry*, *288*(23), 16295–16307.

Sakaguchi, M., Tomiyoshi, R., Kuroiwa, T., Mihara, K., & Omura, T. (1992). Functions of signal and signal-anchor sequences are determined by the balance between the hydrophobic segment and the N-terminal charge. *Proceedings of the National Academy of Sciences*, *89*(1), 16–19.

Sanders, C., Wethkamp, N., & Lill, H. (2001). Transport of cytochrome c derivatives by the bacterial Tat protein translocation system. *Molecular Microbiology*, *41*(1), 241–246.

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. *Journal of Biological Chemistry*, *276*(11), 8159–8164.

Sargent, F., Berks, B. C., & Palmer, T. (2006). Pathfinders and trailblazers: A prokaryotic targeting system for transport of folded proteins. *FEMS Microbiology Letters*, *254*(2), 198-207.

Sargent, F., Bogsch, E. G., Stanley, N. R., Wexler, M., Robinson, C., Berks, B. C., & Palmer, T. (1998). Overlapping functions of components of a bacterial Sec-independent protein export pathway. *EMBO Journal*, *17*(13), 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. *European Journal of Biochemistry*, *268*(12), 3361–3367.

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. *Journal of Biological Chemistry*, *274*(51), 36073–36082.

Schaerlaekens, K., Schierová, M., Lammertyn, E., Geukens, N., Anné, J., & Van Mellaert, L. (2001). Twin-arginine translocation pathway in *Streptomyces lividans*. *Journal of Bacteriology*, *183*(23), 6727–6732.

Schatz, P. J., Riggs, P. D., Jacq, A., Fath, M. J., & Beckwith, J. (1989). The secE gene encodes an integral membrane protein required for protein export in . *Genes & Development*, *3*(7), 1035–1044.

Schiebel, E., Driessen, A. J. M., Hartl, F. U., & Wickner, W. (1991). ΔμH+ and ATP function at different steps of the catalytic cycle of preprotein translocase. *Cell*, *64*(5), 927–939.

Schreiber, S., Stengel, R., Westermann, M., Volkmer-Engert, R., Pop, O. I., & Müller, J. P.

(2006). Affinity of TatCd for TatAd elucidates its receptor function in the *Bacillus subtilis* twin arginine translocation (Tat) translocase system. *Journal of Biological Chemistry*, *281*(29), 19977–19984.

Settles, A. M., Yonetani, A., Baron, A., Bush, D. R., Cline, K., & Martienssen, R. (1997). Sec-independent protein translocation by the maize Hcf106 protein. *Science*, *278*(5342), 1467–1470.

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*(11).

Stanley, N. R., Findlay, K., Berks, B. C., & Palmer, T. (2001). *Escherichia coli* strains blocked in Tat-dependent protein export exhibit pleiotropic defects in the cell envelope. *Journal of Bacteriology*, *183*(1), 139–144.

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*. *Journal of Biological Chemistry*, *275*(16), 11591–11596.

Stolle, P., Hou, B., & Brüser, T. (2016). The Tat substrate CueO is transported in an incomplete folding state. *Journal of Biological Chemistry*, *291*(26), 13520–13528.

Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification*, *41*(1), 207–234.

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. *Journal of Biological Chemistry*,

293,

Tarry, M. J., Schäfer, E., Chen, S., Buchanan, G., Greene, N. P., Lea, S. M., … 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 of the United States of America*, *106*(32), 13284–13289.

Taubert, J., & Brüser, T. (2014). Twin-Arginine translocation-Arresting protein regions contact TatA and TatB. *Biological Chemistry*, *395*(7–8), 827–836.

Taubert, J., Hou, B., Risselada, H. J., Mehner, D., Lünsdorf, H., Grubmüller, H., & Brüser, T. (2015). TatBC-Independent TatA/Tat substrate interactions contribute to transport efficiency. *PLoS ONE*, *10*(3), 1–24.

Tranier, S., Mortier-Barrière, I., Ilbert, M., Birck, C., Iobbi-Nivol, C., Méjean, V., & Samama, J.-P. (2002). Characterization and multiple molecular forms of TorD from *Shewanella massilia*, the putative chaperone of the molybdoenzyme TorA. *Protein Science : A Publication of the Protein Society*, *11*(9), 2148–2157.

Tsirigotaki, A., De Geyter, J., Šoštarić, N., Economou, A., & Karamanou, S. (2017). Protein export through the bacterial Sec pathway. *Nature Reviews Microbiology*, *15*(1), 21–36.

Tsumoto, K., Nakaoki, Y., Ueda, Y., Ogasahara, K., Yutani, K., Watanabe, K., & Kumagai, I. (1994). Effect of the order of antibody variable regions on the expression of the single-chain hyhel10 fv fragment in *Escherichia coli* and the thermodynamic analysis of its antigen-binding properties. *Biochemical and Biophysical Research Communications* 201(2): 546-51.

Tsukazaki, T., Mori, H., Fukai, S., Ishitani, R., Mori, T., Dohmae, N., … Nureki, O. (2008). Conformational transition of Sec machinery inferred from bacterial SecYE

structures. *Nature*, *455*(7215), 988–991.

Tullman-ercek, D., Delisa, M. P., Kawarasaki, Y., Iranpour, P., Ribnicky, B., Palmer, T., & Georgiou, G. (2009). NIH Public Access, *282*(11), 8309–8316.

Turner, R. J., Papish, A. L., & Sargent, F. (2004). Sequence analysis of bacterial redox enzyme maturation proteins (REMPs). *Canadian Journal of Microbiology*, *50*(4), 225–238.

Ulfig, A., & Freudl, R. (2018). The early mature part of bacterial twin-arginine translocation (Tat) precursor proteins contributes to TatBC receptor binding. *Journal of Biological Chemistry*, (10)

Valent, Q. A., Scotti, P. A., High, S., De Gier, J. W. L., Von Heijne, G., Lentzen, G., Wintermeyer, W., Oudega, B., & Luirink, J. (1998). The *Escherichia coli* SRP and SecB targeting pathways converge at the translocon. *EMBO Journal*, *17*(9), 2504–2512.

Van Den Berg, B., Clemons, W. M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C., & Rapoport, T. A. (2004). X-ray structure of a protein-conducting channel. *Nature*, *427*(6969), 36–44.

Van Der Ploeg, R., Barnett, J. P., Vasisht, N., Goosens, V. J., Pöther, D. C., Robinson, C., & Van Dijl, J. M. (2011). Salt sensitivity of minimal twin arginine translocases. *Journal of Biological Chemistry*, *286*(51), 43759–43770.

van der Ploeg, R., Mäder, U., Homuth, G., Schaffer, M., Denham, E. L., Monteferrante, C. G., Miethke, M.A., Harwood, C.R., Winter, T., Hecker, M., Antelmann, H., & van Dijl, J. M. (2011). Environmental salinity determines the specificity and need for tat-dependent secretion of the YwbN protein in *Bacillus subtilis*. *PLoS ONE*, *6*(3) e 18140.

Völkel, T., Korn, T., Bach, M., Müller, R., & Kontermann, R. E. (2001). Optimized linker sequences for the expression of monomeric and dimeric bispecific single-chain diabodies. *Protein Engineering*, *14*(10), 815–823.

von Heijne, G. (1986). The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *The EMBO Journal*, *5*(11), 3021–3027.

von Heijne, G. (1990). The Signal peptide. *J Membr Biol*, *115*, 195–201.

Walther, T. H., Gottselig, C., Grage, S. L., Wolf, M., Vargiu, A. V., Klein, M. J., … Ulrich, A. S. (2013). Folding and self-assembly of the TatA translocation pore based on a charge zipper mechanism. *Cell*, *152*(1–2), 316–326.

Walther, T. H., Grage, S. L., Roth, N., & Ulrich, A. S. (2010). Membrane Alignment of the Pore-Forming Component TatAd of the Twin-Arginine Translocase from *Bacillus subtilis* Resolved by Solid-State NMR Spectroscopy. *Journal of the American Chemical Society*, *132*(45), 15942–15944.

Weatherill, E. E., Cain, K. L., Heywood, S. P., Compson, J. E., Heads, J. T., Adams, R., & Humphreys, D. P. (2012). Towards a universal disulphide stabilised single chain Fv format: Importance of interchain disulphide bond location and vLvH orientation. *Protein Engineering, Design and Selection*, *25*(7), 321–329.

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 (1), 93-101.

Welte, T., Kudva, R., Kuhn, P., Sturm, L., Braig, D., Muller, M., Warscheid, B., Drepper, F., & Koch, H.-G. (2012). Promiscuous targeting of polytopic membrane proteins to SecYEG or YidC by the *Escherichia coli* signal recognition particle. *Molecular*

Biology of the Cell, *23*(3), 464–479.

Wexler, M., Sargent, F., Jack, R. L., Stanley, N. R., Bogsch, E. G., Robinson, C., … Palmer, T. (2000). TatD is a cytoplasmic protein with DNase activity. No requirement for TatD family proteins in Sec-Independent protein export. *Journal of Biological Chemistry*, *275*, 16717–16722.

Wexler, M., Sargent, F., Jack, R. L., Stanley, N. R., Bogsch, E. G., Robinson, C., Berks., & Palmer, T. (2000). TatD is a cytoplasmic protein with DNase activity. *The Journal of Biological Chemistry*, *275*, 16717–16722.

White, G. F., Schermann, S. M., Bradley, J., Roberts, A., Greene, N. P., Berks, B. C., & Thomson, A. J. (2010). Subunit organization in the TatA complex of the twin arginine protein translocase: A site-directed EPR spin labeling study. *Journal of Biological Chemistry*, *285*(4), 2294–2301.

Wickner, W., & Leonard, M. R. (1996). *Escherichia coli* preprotein translocase. *Journal of Biological Chemistry*.

Wörn, A., & Plückthun, A. (2001). Stability engineering of antibody single-chain Fv fragments. *Journal of Molecular Biology*, *305*(5), 989–1010.

Yen, M. R., Tseng, Y. H., Nguyen, E. H., Wu, L. F., & Saier, M. H. (2002). Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system. *Archives of Microbiology*.

Zhang, Y., Hu, Y., Li, H., & Jin, C. (2014). Structural basis for TatA oligomerization: An NMR study of *Escherichia coli* TatA dimeric structure. *PLoS ONE*, *9*(8).

Zhang, Y., Wang, L., Hu, Y., & Jin, C. (2014). Solution structure of the TatB component of the twin-arginine translocation system. *Biochimica et Biophysica Acta - Biomembranes*, *1838*(7), 1881–1888.

Zheng, N., & Gierasch, L. M. (1997). Domain interactions in E. coli SRP: Stabilization of M domain by RNA is required for effective signal sequence modulation of NG domain. *Molecular Cell*, *1*(1), 79–87.

Zoufaly, S., Fröbel, J., Rose, P., Flecken, T., Maurer, C., Moser, M., & Müller, M. (2012). Mapping precursor-binding site on TatC subunit of twin arginine-specific protein translocase by site-specific photo cross-linking. *Journal of Biological Chemistry*, *287*(16), 13430–13441.