Proofreading of substrate by the *Escherichia coli* Twin Arginine Translocase.

Alexander Stephen Jones

A thesis submitted for the degree of Doctor of Philosophy

University of Kent

Department of Biosciences

August 2017

Contents

List of Figures iv
List of Tables vi
Acknowledgements viii
Declaration ix
Summary x
List of Abbreviations xi
1 Introduction 1
1.1 Targeting: the Signal Peptide 4
1.2 General Secretory Pathway, Sec 7
1.2.1 Components of the Sec machinery 8
1.2.2 Co-translational translocation 13
1.2.3 Post-translational translocation 15
1.2.4 Membrane insertion and YidC 18
1.3 The Twin Arginine Translocation pathway, Tat 18
1.3.1 Components of the Tat pathway 20
1.3.2 Tat systems in other organisms 25
1.3.3 Structure of the TatABC complex 26
1.3.4 Mechanism of translocation 29
1.3.5 Energy requirement of the Tat system 33
1.3.6 Proofreading and Quality control 34
1.4 Aims of this project 41
2 Materials and Methods 46
2.1 DNA Techniques 46
2.1.1 Preparation of plasmid DNA 46
2.1.2 Amplification Polymerase Chain Reaction (PCR) 46
2.1.3 Amplification (PCR) Primers 47
2.1.4 Agarose gel electrophoresis 47
2.1.5 Purification of DNA from agarose gels 49
2.1.6 Restriction Digests of DNA 49
2.1.7 Ligation of DNA fragments into plasmid vector backbone 50
2.1.8 Ligation of DNA fragments 50
2.1.9 Site-specific DNA mutagenesis 50

2.1.10 Site-specific mutagenesis Primers....51
2.1.11 Sequencing of plasmid DNA....53
2.1.12 Constructs generated in this study....53
2.2 Growth and Maintenance of E. coli cultures....56
2.2.1 Glycerol stocks....56
2.2.2 Preparation of competent cells....56
2.2.3 Media....56
2.2.4 Transformation of competent E. coli cells....57
2.3 Protein production....57
2.3.1 Culture of E. coli and Plasmid induction....57
2.3.2 Preparation of spheroplasts....58
2.3.3 Cytoplasm and Membrane separation....58
2.3.4 Separation of Insoluble fraction....59
2.3.5 Time course assay....59
2.4 Protein purification....59
2.4.1 Immobilised Metal Affinity Chromatography (IMAC), Nickel....59
2.4.2 Protein quantification assay....61
2.4.3 De-salting IMAC peak fractions....61
2.4.4 Concentrating IMAC peak fractions....61
2.5 Protein separation....61
2.5.1 SDS poly-acrylamide gel electrophoresis (SDS-PAGE)....61
2.5.2 Detection of proteins with Coomassie....62
2.6 Protein imaging....62
2.6.1 Western-blotting....62
2.6.2 Detection of proteins by immunoblotting....62
Testing the proofreading capacity of the Twin-Arginine Transolcase through export of an scFv with altered surface charge.....66
Testing the proofreading capacity of the Tat system with a wholly synthetic, heme-binding BT6 Maquette.....96
Comparison of quality of human Growth Hormone (hGH) exported to the periplasm via the General Secretory (Sec) or Twin Arginine Translocase (Tat) pathways.....115
Truncation analysis of Tat components effect on proofreading capacity.....128
Final Discussion.....143
Future Perspectives.....151
References....153

Supplementary 153

List of Figures

Figure 1	Overview and location of the General Secretory (Sec) and Twin-Arginine Translocase (Tat) pathways in bacteria, thylakoids and the eukaryotic ER.	3
Figure 2	Tat and Sec signal peptides.	6
Figure 3	Eukaryotic and bacterial components of the General secretory (Sec) pathway.	12
Figure 4	Schematic of Escherchia coli Tat components and complexes.	21
Figure 5	Proposed translocation mechanism for the Tat pathway.	32
Figure 6	Primary amino acid sequence and structural features of scFvM.	69
Figure 7	Ribbon, surface potential and surface hydrophobicity images of scFvM from two projections.	71
Figure 8	Mass Spec data verifying reduced state of purified scFvM used for biophysical studies.	73
Figure 9	Expression of TorA-scFvM (unmodified and Cysteine variants) with or without CyDisCo.	75
Figure 10	Structures of scFvM variants with introduced surface salt-bridges (SB) or patch of charge.	78
Figure 11	Fractionation controls for scFvM export assays.	79
Figure 12	Coomassie-stained gel of scFvM variants purified by protein A affinity column.	80
Figure 13	Export assays for salt-bridge and charged patch scFvM variants.	84
Figure 14	Export efficiencies calculated for the scFvM variants used in this study.	85
Figure 15	The Tat system tolerates significant surface hydrophobicity in scFvM.	87
Figure 16	Addition of a 26-residue disordered tail results in a complete block in export of scFvM by the Tat system.	88
Figure 17	Primary amino acid sequence and structural features of maquette BT6.	99
Figure 18	Overexpression of BT6 from the pJexpress414 plasmid causes contamination of the Periplasmic fraction.	101
Figure 19	Low-level expression of BT6 from the pEXT22 plasmid results in correct processing by the Tat system.	103
Figure 20	1D proton NMR for BT6 heme-binding variants.	106
Figure 21	Combined 1D proton NMR and export assays for the heme-binding BT6 variants.	108
Figure 22	Schematic representations of linked BT6 variants.	110
Figure 23	BT6h0 can be exported by the Tat system when linked behind BT6h2.	111

Figure 24	Targeting recombinant proteins to the Sec pathway is more metabolically draining over a 3 hr period.	118
Figure 25	hGH targeted to the periplasm via the Sec pathway is immediately available whereas Tat-targeted protein builds up over a longer time period.	119
Figure 26	Targeting hGH to the periplasm via the Tat pathway results in a larger insoluble fraction.	120
Figure 27	Ni IMAC purification of hGH from the periplasm.	122
Figure 28	Purified hGH.	123
Figure 29	Full spectrum proton NMR resonances for purified hGH processed by either the Sec or Tat pathway.	124
Figure 30	Expansion of the methyl and aromatic regions for ^{1}H NMR spectra for Sec and Tat processed hGH.	126
Figure 31	Expression of truncated TatA from the arabinose-inducible Tat0 plasmid.	132
Figure 32	Truncated TatA does not complement the filamentous phenotype of ΔtatABCDE cells	135
Figure 33	Growth of truncated TatA or TatB on LB agar plates supplemented with 2% SDS and 0.05 mM arabinose.	136
Figure 34	anti-StrepII (A&B) and anti-TatA (C&D) immunoblot for expression of truncated TatB from the Tat0 plasmid.	138
Figure 35	Truncated TatB complements the filamentous phenotype.	140
Figure 36	Schematic representation for a possible mechanism for Proofreading by the TatABC machinery.	150

List of Tables

Table 1	Primers used for amplification PCR.	48
Table 2	Restriction enzymes used in this study.	49
Table 3	Primers used for Quick Change PCR	51
Table 4	Sequencing primers used in this study	53
Table 5	Constructs generated in this study.	53
Table 6	Antibodies used in this study.	63
Table 7	Structural characteristics of scFvM variants characterised by circular dichrosim, tryptophan fluorescence and dynamic light scattering.	82

Acknowledgements

I would like to thank Professor Colin Robinson for giving me the opportunity to carry out the work presented in this thesis and his continued help, support and guidance throughout.

I would like to thank my family – Stephen, Helen and Bobbi Jones for their support throughout.

I would also like to thank the CR Group members, both past and present:
Roshani Patel and Sarah Smith for making my dissertation enjoyable, without whom I wouldn't have undertaken a PhD.
The CR Family – Kelly Walker, Daphne Mermans, Kelly Frain, Jo Roobol, Isabel Guerrero Montero, Julie Zedler, Doris Gangl, Amber Peswani, Kirsty Richards, Andrew Dean, Conner Webb, Felix Nicolaus, Ally Walters and Wayne Miller for making the lab a truly enjoyable place to work.

Special thanks to Professor David Brown for his advice, support and guidance in protein modelling using CCP4 MG.

Special thanks to Dr Gary Thompson for his advice and guidance in NMR.

Declaration

The work presented in this thesis is original work, conducted by myself 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 thank and acknowledge the following people for their help in producing data used in this thesis:
Dr James Austerberry, Dr Jim Warwicker, Dr Robin Curtis and Professor Jeremy Derrick for the scFvM biophysical work and the following useful discussions.

Dr Katy Grayson, George Sutherland, Dr Andrew Hitchcock and Professor Neil Hunter for the maquette BT6 NMR work and the following useful discussions.

Kevin Howland and Gary Thompson for the use of mass spectrometry analysis and NMR analysis respectively.

Some of the results presented in this thesis have been published in, or are in preparation to be submitted to, the following journals:

AS Jones, JI Austerberry, R Dajani, J Warwicker, R Curtis, JP Derrick, C Robinson, 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, BBA Molecular Cell Research 1863 (2016) 3116 – 3124.

KM Frain, **AS Jones**, R Schoner, KL Walker, C Robinson, The *Bascillus subtilis* TatAdCd system exhibits an extreme level of substrate selectivity, BBA Molecular Cell Research 1864 (2017) 202 – 208.

GA Sutherland, KJ Grayson, DMJ Mermans, **AS Jones**, AJ Robertson, D Auman, AA Brindley, F Sterpone, P Tuffery, P Derreumaux, PL Dutton, C Robinson, A Hitchcock, CN Hunter, Probing the quality control mechanism of the *Escherichia coli* twin-arginine translocase using folding variants of a de novo-designed heme protein (2017) (*submitted*).

Summary

The Twin Arginine Translocase (Tat) is one of two protein translocation mechanisms in *E. coli* to move proteins across the inner bacterial membrane, from the cytosol to the periplasm. A unique feature of the Tat pathway is its ability to translocate fully folded proteins, indeed, in *E. coli* the Tat pathway preferentially transports correctly folded proteins. This 'proofreading' mechanism, as it has been dubbed, is of interest to the biopharmaceutical industry, however little is known of the mechanism by which Tat proofreads a substrates conformational state.

Initial studies (chapter 3) addressed if the Tat proofreading mechanism sensed the surface charge or hydrophobicity of a substrate. To this end, surface residues of an scFv were mutated to create areas of charge and hydrophobicity without altering tertiary structure. Expression of these variants in *E. coli* revealed that Tat proofreading is tolerant of surface charge and hydrophobicity, but dependent on conformational flexibility. Further studies utilising a maquette in various folding states, confirmed Tat proofreading is sensitive to the structural rigidity of substrates (chapter 4).

Investigations then went on to assess the quality of protein entering the periplasm via the Tat pathway by comparing it to the same protein transported by the General Secretory (Sec) pathway (chapter 5). This revealed, at least for a relatively simple biotheraputic, Tat-translocated protein is of the same quality to Sec-translocated protein.

Finally, the question of what is responsible for the proofreading ability of Tat began to be addressed through C-terminal truncation studies of the Tat components that attempted to restore export of export-incompatible substrates (chapter 6).

List of Abbreviations

°C	degrees celsius
µL	microlitre
µM	micromolar
µm	micrometer
A	Alanine (Ala)
A. aeolicus	*Aquifex aeolicus*
aas	amino acids
ADP	adenosine diphosphate
Ala-x-Ala	Alanine-any-Alanine
APH	amphipathic helix
ATP	adenosine triphosphate
AxA	Alanine-any-Alanine
BiP	Binding immunoglobulin Protein
BN-PAGE	blue-native polyacrylamide gel electrophoresis
bp	base pair
C	Cytoplam fraction
C-	carboxyl-
CyDisCo	Cytoplasmic Disulphide formation in *E. coli*
D	Aspartate/Aspartic acid
DHFR	dihydrofolate reductase
DNA	deoxyribonucleic acid
DTT	dithiothreitol
E	Glutamate/Glutamic acid (Glu)
E(n)	Elution
E. coli	*Escherchia coli*
ECL	enhanced chemoluminescence
ECM	Extracellular matrix
EDTA	ethylenediaminetetraacetic acid
EM	electron microscopy
ER	Endoplasmic Reticulum, eukaryotic, mammalian
Erv1p	yeast mitochondrial thiol oxidase
F	Phenylalanine (Phe)
FeS	Iron-sulphur
Ffh	fifty-four homologue
FT	Flow through
G	Glycine (Gly)
g	grams
GFP	Green fluorescent protein
GMP	guanosine monophosphate
GTP	guanosine triphosphate
H	Histidine (His)
hGH	human Growth Hormone
hPDI	human Protein Disulphide Isomerase
hr(s)	hour (hours)

HRP	horseradish peroxidase
HSP70	Heat Shock Protein 70kDa
I	Insoluble fraction
I	Isoleucine (Ile)
IgG	Immunoglobulin G
IPTG	isopropylthiogalactoside
kDa	kilodaltons
KR	Lysine (Lys), Arginine (Arg)
L	Leucine (Leu)
L	litre
LB	Luria Bertani broth
LBA	Luria Bertani Agar
Lep1	leader peptidase
M	Membrane fraction
M	molar
mA	milliampere
MBP	Maltose Binding Protein
mg	milligram
MGD	molybdopterin guanine dinucleotide
min	minute(s)
mL	millilitre
mM	millimolar
Mtt	Membrane trgeting and translocation
N	Asparagine (Asn)
NBD	nucleotide binding domain
ng	nanogram
OD	optical density
P	Proline (Pro)
P	Periplasm fraction
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PBS-T	phosphate buffered saline and Tween-20
PhoD	Alkaline phosphatase D
PMF/pmf	proton motive force
Q	Glutamine (Gln)
QC	Quality Control
QCS	Quality Control Suppression
REMP	Redox Enzyme Maturation Protein
RK	Arginine (Arg), Lysine (Lys)
RNA	Ribonucleic acid
RNC	ribosome-nascent chain
rpm	revolutions per minute
RR	Twin-Arginine motif
S	Serine (Ser)
SDS	sodium dodecylsulphate
scFv	single-chain varible-fragment from an antibody
scFvM	Single-chain variable fragment, Manchester
Sec	General Secretory pathway

SR	SRP receptor
SRP	Signal Recognition Particle
T	Threonine (Thr)
Tat	Twin-Arginine Translocase
TC	Total Cell
TM	transmembrane
TMAO	trimethylamine-*N*-oxide reductase
TorA	trimethylamine-N-oxide reductase
TorA-	torA signal peptide fused to folowing protein
Tris	Tris (hydroxymethyl) aminomethane
Tween20	polyoxyethylenesorbitan monolaurate
U	Total cell, Uninduced
UV	ultraviolet
V	Valine (Val)
V	Volts
v/v	volume per volume
w/v	weight per volume
W	Wash
W	Tryptophan
WT	wild-type strain
XNTPs	nucleoside triphosphates (here guanosine or adenosine)
Y	Tyrosine (Tyr)
α	alpha
β	beta
ΔpH	potential of Hydrogen
ΔΨ	charge gradient
rhGH	recombinant human Growth Hormone
IMV	Inverted Membrane Vesicels

1 Introduction

Proteins, one of the essential constituents of life. They are large molecules or 'macromolecules' that carry out all manner of functions within an organism – acting as signalling molecules, replicating DNA, catalysing reactions, facilitating cellular movement, and the transport of ions/molecules from one location to another, to name a few. However as beneficial as proteins are to an organism, if they are not at their correct site of function they can be incredibly damaging and failure is often a cause of disease. This poses a problem for most organisms as cells are composed of - and are often compartmentalised by - tightly sealed phospholipid membranes that very few molecules can cross without aid. The majority of proteins are synthesised in the cytoplasm and thus by necessity these barriers must be crossed (Wickner *et al* 1991).

A variety of transport mechanisms have evolved to carry out the task of delivering substrates across lipid bilayers without compromising the structural integrity or function of the membrane. A variety of very different pathways have originated to facilitate the translocation of proteins, albeit in differing scenarios. Despite this diversity, translocation pathways have several key features in common – a gated membrane-spanning pore (composed itself of proteins), an energy requirement to facilitate substrate translocation (usually in the form of hydrolysis of nucleoside triphosphates (XNTPs) and/or utilisation of the proton motive force (pmf, ΔpH and $\Delta\psi$)) and in the case of protein translocation, delivery to the correct translocase via a specific (usually N-terminal) signal sequence.

The focus here is on translocation of proteins across lipid bilayers in bacteria, for which there are two major mechanisms. The first is shared throughout all domains of

life (Natale *et al* 2008) - Eukarya, Archaea and Bacteria. This primary and predominant answer to the protein translocation conundrum is the General Secretory Pathway (Sec). In eukarya, the Sec pathway functions in the endoplasmic reticulum (ER) and serves to secrete proteins into the lumen of the ER where they are further delivered to their site of function (e.g. extracellular matrix (ECM)) by subsequent secretory steps (Golgi etc). Sec performs a similar function in bacteria and archaea, except that it's located is the cell membrane. Plant plastids also utilise the Sec pathway to deliver proteins to the thylakoid lumen. Plastids (e.g. chloroplasts of plants or algae) and bacteria encode a second translocation pathway that also transports proteins to the thylakoid lumen and bacterial periplasm (Simone *et al* 2013). This second mechanism is the Twin Arginine Translocation (Tat) pathway. These translocation mechanisms are described in more detail in later sections and Figure 1 briefly outlines the site(s) of these translocating mechanisms.

There are several properties shared by these translocation mechanisms which evidently form the basic necessities for a translocation system to correctly function. The first is a targeting mechanism to ensure substrates are directed to the appropriate transport pathway. This is in the form of an amino acid sequence, known as a signal peptide, usually located in the N-terminus of translating polypeptides. Second, is an environment to cross the membrane. This is in the form of a channel composed from membrane embedded proteins and associated soluble partners. Finally, some form of energy requirement to facilitate translocation (Rapoport *et al* 1996, Yuan *et al* 2010).

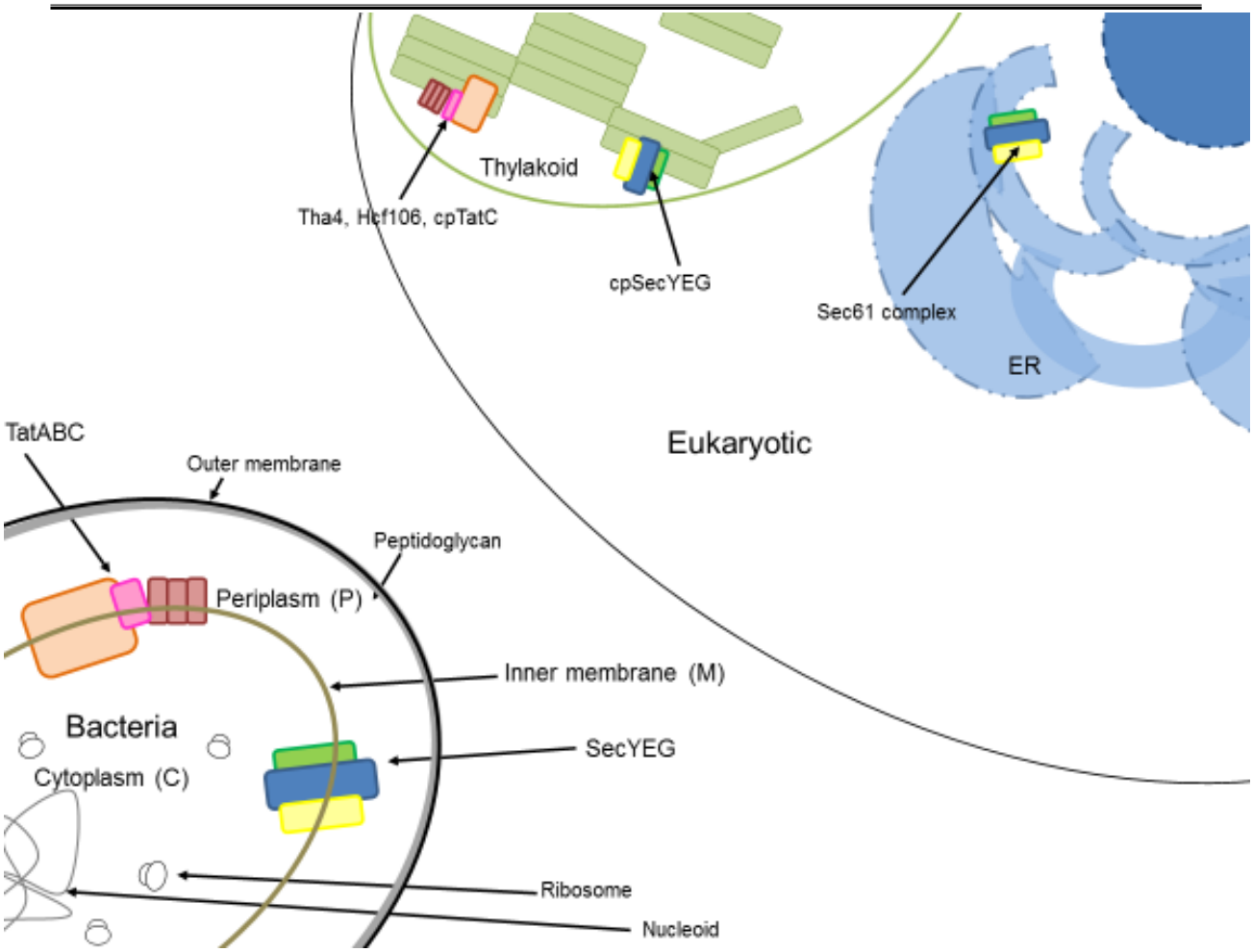


Figure 1: **Overview and location of the General Secretory (Sec) and Twin-Arginine Translocase (Tat) pathways in bacteria, thylakoids and the eukaryotic ER**. Although separate translocation mechanisms evolved, their components perform the same basic functions. Blue/Green/Yellow = General Secretory pathway. Orange/Pink/Red = Twin-Arginine Translocase.

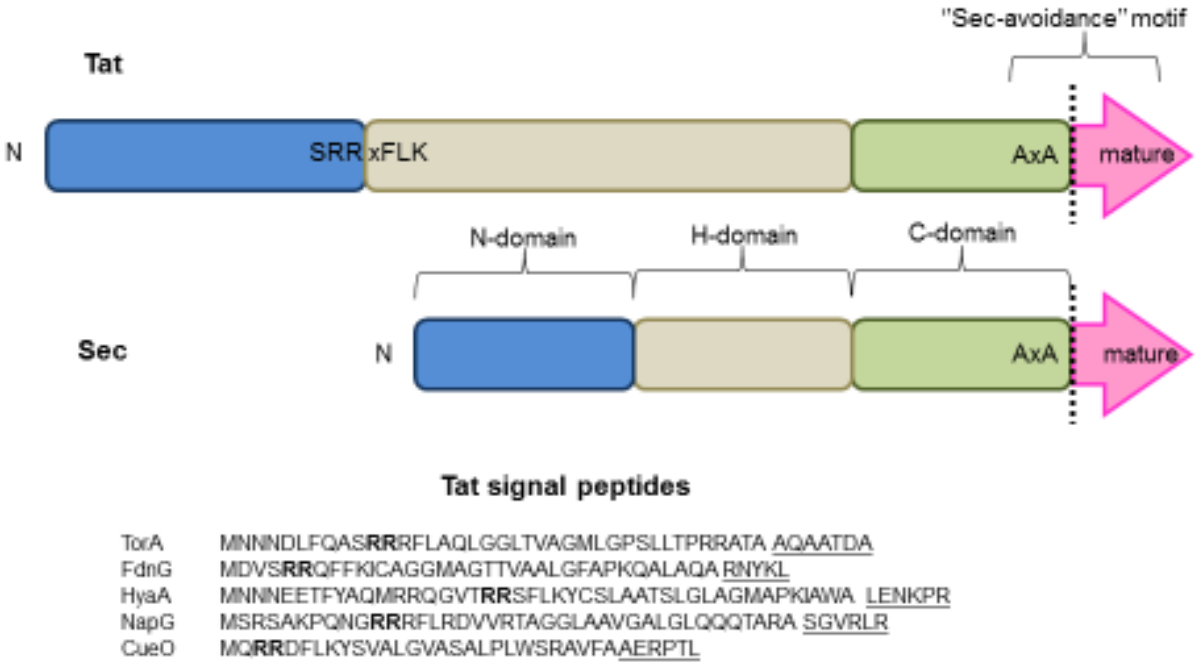
1.1 Targeting: the Signal Peptide

Proteins destined for any of the membrane-crossing translocation pathways are usually translated as precursors that include a signal peptide to specify the translocase to be utilised. For the majority of translocase substrates, the signal peptide consists of the first 20 – 50 residues, however for membrane proteins the signal peptide is usually the first TM domain (Blobel and Dobberstein 1975, Wickner *et al* 1991, Osborne *et al* 2005, Natale *et al* 2008). Signal peptides are specific for the translocase they target to and, although they vary in individual amino acid composition from substrate to substrate, they have overall features. Sec and Tat signal peptides are tripartite consisting of a hydrophilic, basic N-domain, followed by a hydrophobic core region, known as the H-domain, and finally a polar C-domain (Berks *et al* 1996). Signal peptides are cleaved after membrane translocation and this 'cleavage site' resides in the C-domain consisting of short-chain residues at the -3 and -1 positions, usually Ala-x-Ala (AxA).

The structural criteria for a signal peptide appear to be conserved across organisms as thylakoid, bacterial and eukaryotic Sec and Tat signal peptides are similarly structured. Interestingly, Tat signal peptides function across organisms as thylakoid Tat signal peptides can target substrates to bacterial Tat machinery and *vice versa* (Mori and Cline 1998, Spence *et al* 2003, Aldridge *et al* 2008).

Tat signal peptides differ from those that target substrates to Sec in several key areas. First, they are longer, at least 30 residues in length (compared to 17 – 24 for Sec substrates), with additional residues predominantly located in the basic N-domain and the hydrophobic H-domain (Cristóbal *et al* 1999, Kipping *et al* 2003). Despite the

increased length of the H-domain, Tat signal peptides are overall less hydrophobic than Sec, or even SRP, signal peptides (Cristóbal *et al* 1999). Second, Tat signal peptides contain the eponymous twin-arginine motif, ‘’RR’’ (Sargent *et al* 1998). These dual arginine residues reside within a larger S-R-R-x-F-L-K motif, found in nearly all Tat signal peptides spanning the junction of the N- and H-domains (Sargent *et al* 1998). Figure 2 gives a comparative overview of Sec and Tat signal peptides. While only the two arginine residues are invariant, the other residues are found with a frequency of > 50% (Berks 1996, Stanley *et al* 2000). While X can be any amino acid, it is usually polar. Numerous studies have shown the importance of the arginine pair for transport by the Tat system, both in bacteria and thylakoids. In bacterial and thylakoid Tat systems substitution of ‘RR’ with dual Lys (a charge-neutral change, RR > KK) blocks export completely, while in thylakoids RR > KR is tolerated, albeit with a decrease in export efficiency, and RR > RK abolishes export (Chaddock 1995, Halbig *et al* 1999). The situation in bacterial Tat is slightly different in that substitution of either arginine to lysine leads to a reduced level of export (Buchanan *et al* 2001). Some instances have been reported where a single arginine to glutamine substitution is tolerated (DeLisa *et al* 2002). These variations are only found experimentally and occur rarely in native Tat substrates with only one known protein reliably identified to cross the membrane via Tat harbouring a KR motif – the TtrB subunit of tetrathionate reductase in *Enterobacteriaceae* (Hinsley *et al* 2001). Despite the abundance of work pertaining to the importance of the twin arginine residues, relatively little work has focused on the remainder of the residues in the Tat-targeting motif. However, one study indicated a residue of high hydrophobicity is important at the phenylalanine position (+2 relative to RR) (Stanley *et al* 2000).



Tat signal peptides

TorA	MNNNDLFQASRRRFLAQLGGLTVAGMLGPSLLTPRRATA AQAATDA
FdnG	MDVSRRQFFKICAGGMAGTTVAALGFAPKQALAQA RNYKL
HyaA	MNNNEETFYAQMRRQGVTRRSFLKYCSLAATSLGLAGMAPKIAWA LENKPR
NapG	MSRSAKPQNGRRRFLRDVVRTAGGLAAVGALGLQQQTARA SGVRLR
CueO	MQRRDFLKYSVALGVASALPLWSRAVFAAERPTL

Figure 2: **Tat and Sec signal peptides. Top**: Comparison of Sec and Tat signal peptides, showing the basic N-domain (blue) followed by the hydrophobic H-domain (beige) and finally the polar C-domain (green) prior to the AlaXAla cleavage site. The Tat consensus motif, SRRxFLK, bridges the N- and H-domains.

Bottom: A collection of Tat signal peptides with the eponymous 'RR' motif in bold and the first five amino acids of mature protein underlined.

Apart from the twin-arginine motif, Sec and Tat signal peptides are highly similar and thus there must be more to the signal peptide that prevents targeting by the incorrect translocase. Several studies have suggested the C-domain contributes to avoidance of the incorrect pathway. In Tat signal peptides the region immediately prior to the AxA site is usually quite basic compared to the almost neutral charge of Sec signal peptides in this region – around +0.5 and +0.03 (Berks, Palmer and Sargent 2003) respectively. This region of charge has been dubbed a putative 'Sec avoidance' motif. Indeed, adding positively charged residues in this region blocks Sec export. One study went further and suggested the 'Sec avoidance' motif encompasses the AxA site and includes the first five amino acids of the mature protein (Bogsch *et al* 1997, Tullman-Ercek *et al* 2007).

In eukaryotes two proteins are predominantly associated with the luminal side of the translocation apparatus – signal peptidase and oligosaccharyl transferase. These highly abundant proteins are often found in association with the Sec apparatus and serve to remove the signal peptide upon translocation of substrate (von Heijne 1998) and catalyse the addition of the oligosaccharyl moiety from dolichol to translocating polypeptides respectively.

Signal peptidase generally cleaves at the AxA site and, while it performs a relatively simple function, is quite complex, at least in mammals and yeast. In mammals and yeast signal peptidase comprises 5 different membrane proteins – SP18, SP21, SP22, SP23, SP25 and SP12. Each of these spans the membrane at least once while SP22/23 is a glycoprotein (Paetzel *et al* 2002). The bacterial and thylakoid signal peptidase meanwhile is known as leader peptidase, or Lep1, and consists of only a single subunit that spans the membrane twice. Leader peptidase cleaves signal peptides from both Sec and Tat substrates once on the *trans* side of the membrane.

1.2 General Secretory Pathway, Sec

As early as the 1970s it was well understood that cells exported proteins from their site of synthesis – the cytosol – to the outer cell and the environment (Milstein *et al* 1972). However, it wasn't until the early '80s that genetic experiments first identified the exact components that facilitate export across cell membranes to the periplasm in *Escherichia coli* and later in yeast. Two different but similar approaches identified the same three genes as being essential for viability. The first created mutations in signal peptides that lead to incorrect localisation of cognate proteins to screen for mutations

that restored correct localisation. The resulting mutations were named *prl* genes for protein localisation (Bedoulle *et al* 1980, Emr *et al* 1981, Derman *et al* 1993). The second study screened for mutations that conferred conditional pleiotropic secretion defects, thus termed *sec* genes. The three genes identified in each study were SecY, SecE and SecA. Their individual roles in protein secretion were subsequently established by purification and reconstitution *in vitro*. These reconstitution assays helped to identify another component, SecG (Veenendaal and Driessen 2004). A third elegant assay utilised a chimeric protein made from the fusion of the N-terminal of maltose-binding protein (MBP), which is usually exported to the periplasm to β-galactosidase (LacZ) which normally functions in the cytoplasm, to yield a different phenotype. This fusion targets LacZ to the inner membrane thus inhibiting β-galactosidase activity (which usually occurs in the cytoplasm) and eventually stopping export machinery in wild-type cells. Thus proteins with mutations in secretion machinery have increased β-galactosidase activity, due to accumulation of MBP-LacZ chimera in the cytoplasm, and a build-up of other secreted proteins (Ito *et al* 1981, Schatz and Beckwith 1990, Danese 1998). Several important proteins of the secretion machinery were identified in this manner.

1.2.1 Components of the Sec machinery

Translocation of proteins can occur at one of two stages – during translation or after translation, termed co-translational translocation and post-translational translocation respectively. Transport across the phospholipid membrane at either of these stages utilises the same core components of the Sec machinery, composed from membrane proteins with associated soluble partners.

The largest of these is Sec61α in mammals, which shares 56% amino acid homology with Sec61p in the yeast *Saccharomyces cerevisiae* and the bacterial SecY (Golch *et*

al 1992). Eukaryotic Sec61 resides in the ER membrane whereas bacterial SecY is found in the cell membrane (Deshaies and Schekman 1987). This largest Sec component is 443 amino acids at 48 kDa (Cerretti *et al* 1983) and is composed of 10 transmembrane (TM) helices in all cases, where both N- and C-termini are in the cytosol (Akiyama and Ito 1987). The Sec pathway only translocates proteins in an *unfolded* state and as such the pore only needs to accommodate substrates of a fixed diameter (that of a polypeptide chain) (van den Berg *et al* 2004). Various reports put the pore size in the region of 10Å – 12Å as it can accommodate glass bearings of 10Å – 12Å and α-helices with a diameter of 10Å (Mingarro *et al* 2000, Kowarik *et al* 2002) however, virtual simulations and the successful translocation of biotin-incorporated proteins allow the possibility that the channel is flexible and may be able to widen to 16 – 20Å (Kurzchalia *et al* 1988, Tani *et al* 1990, Tian and Andricioaei 2006).

The primary component of the Sec machinery is SecY composed of 10 TM helices, these can be divided into two domains – encompassing TM1-5 and TM6-10. These domains are pseudosymmetrical where TM6-10 are an upside-down version of TM1-5 and are known as the C- and N-domains respectively (Akiyama and Ito 1987, van den Berg 2004, reviewed in Osborne *et al* 2005). Together these domains form an aqueous pore through which substrate polypeptides are translocated.

The other core component of the Sec machinery needed for viability is the 127 amino acid, 14 kDa SecE (Schatz *et al* 1989). One of the methods used to affirm substrates are translocated via Sec is in SecE depletions strains where Sec secretion is non-functional and *in vitro* reconstituted Sec translocases in proteoliposomes are functional with only SecYE present (Schatz *et al* 1991). The mammalian and yeast homologues

of SecE are Sec61γ and Sss1p respectively. Sec61γ and Sss1p form 1 TM helix with N-in-C-out topology, whereas in *Escherichia coli* SecE folds into 3 TM helices with only the final helix sharing sequence similarity with the eukaryotic versions. Indeed, the 2 helices at the N-terminal of *E. coli* SecE are not essential for function (Murphy and Beckwith 1994).

The third core component of bacterial Sec machinery is SecG which is smaller at 110 amino acids and 12 kDa (Nishiyama *et al* 1993). Proteins that perform a similar function can be found in mammals and yeast, Sec61β and Sbh1p respectively, however SecG shares no obvious similarity to these proteins. Although only SecY and SecE are essential for viability in the bacterial system, this third 'core' component has a stimulatory effect on translocation (Nishiyama *et al* 1993, Nishiyama *et al* 1994). SecG has N-in-C-in topology and forms 2 TM domains (Nishiyama *et al* 1996).

Together SecY, SecE and SecG form a stable membrane embedded heterotrimeric complex named SecYEG, shown in Figure 3.

The X-ray structure of *Methanococcus jannaschii* SecYEβ has been solved to 3.2 Å resolution, and shows one copy of each subunit present and is analogous to *E. coli* SecYEG (van den Berg *et al* 2004). SecE and Secβ (SecG) are located on the periphery of the complex, with SecY in the centre forming a ''clamshell'' shape laterally across the membrane (van den Berg *et al* 2004). The centre of the clamshell forms the channel through which polypeptides destined for the periplasm pass. This channel is thought to be open at the cytoplasmic face and thus needs a 'plug' to stop the unwanted leakage of ions, molecules or non-targeted polypeptides to the periplasm. A small helical

segment, termed TM2a in the second transmembrane helix of SecY is thought to perform this role. This plug rotates back into a cavity on the periplasmic face during translocation, allowing access to the periplasm (van den Berg *et al* 2004, Flower 2007).

Although overall clamshell in shape, the channel formed by SecY is not strictly cylindrical and possibly forms an hourglass shape, with two inverted 'funnels' meeting at a constriction point. These funnels are lined with hydrophilic residues to 'guide' polypeptides through the membrane to the constriction point which is composed of 6 hydrophobic residues (van den Berg *et al* 2004, Cannon *et al* 2005). In *E. coli* this pore ring is composed of Isoleucine residues and helps to form a 'seal' around the translocating polypeptide such that leakage of the membrane does not occur (van den Berg *et al* 2004). Polypeptides are then released into the periplasm via the second funnel with movement of the TM2a 'plug' (Tam *et al* 2005). The loop connecting TM5 and TM6 forms the juncture of domains in SecY and this is thought to be the 'hinge' for the clamshell. SecE serves as a 'brace' around this hinge through a network of contact points. The arrangement of SecE, SecG and each of the SecY domains leaves the front of the clamshell able to open and this is likely how TM segments of membrane proteins laterally exit the SecY channel to enter the bilayer (van den Berg *et al* 2004).

So far only the membrane components of Sec machinery have been described, the soluble components come in to play during translocation itself. As previously mentioned, membrane crossing via the Sec route takes place at one of two points – co-translationally or post-translationally.

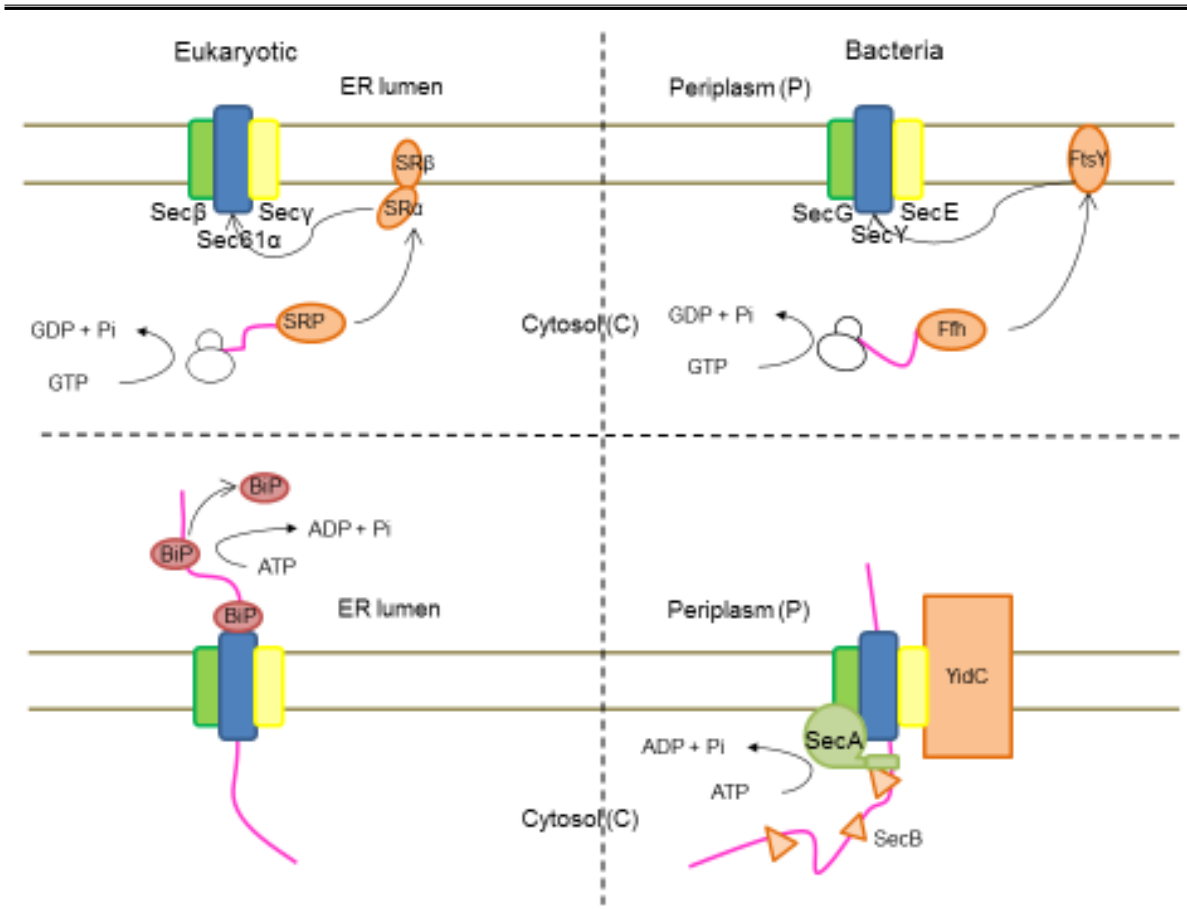


Figure 3: **Eukaryotic and bacterial components of the General secretory (Sec) pathway**. Also shown are co- and post-translational translocation (top two panels and bottom two panels respectively). In Co-translational translocation the bacterial Ffh (in mammals SRP) binds to the signal peptide of a newly translating polypeptide and guides the nascent-peptide-ribosome complex to FtsY (SRαβ) at the membrane, before dissociating. This nascent-peptide-ribosome-SR complex is subsequently shuttled to the Sec translocation channel, SecYEG (Sec61αγβ) whereupon SR dissociates. Translocation across the membrane through the Sec channel then progresses utilising the 'pushing' power of GTP hydrolysis by the ribosome as the polypeptide elongates. Homologous subunits are coloured the same.

The bottom panels show post-translational translocation, processes which differ significantly in eukaryotes and bacteria. The process is described in detail in the main text (section 1.2.3).

1.2.2 Co-translational translocation

All cells and organisms generally secrete proteins by co-translational translocation. However, the process is best described for eukaryotes. As a translated polypeptide emerges from the ribosome, information encoded in the signal peptide (section 1.1) triggers association with another soluble nucleoprotein complex. In eukaryotes this nucleoprotein is known as the Signal Recognition Particle (SRP). The eukaryotic SRP is a complex of a 300 nucleotide 7S RNA component intertwined with 6 polypeptides denoted 9, 14, 19, 54, 68 and 73, which corresponds to their molecular weights in kDa (Gundelfinger *et al* 1983). SRP binds to the signal peptide of the translating polypeptide via a methionine-rich domain (M-domain) in the 54 kDa polypeptide. Binding of SRP to the signal peptide pauses translation ('elongation arrest') and the close proximity of SRP to the ribosome increases affinity for GTP at the G-domain of the 54 kDa polypeptide (Freymann *et al* 1997, Zheng and Gierasch 1997). The ribosome-nascent chain: SRP complex now interacts with the SRP receptor (SR) on the membrane surface of the ER. SR is composed of two subunits – a 30 kDa membrane-anchored SRβ and a 70 kDa membrane-associated SRα (Montoya *et al* 1997). When GTP is bound to the G-domain, SRP interacts with SRα (Legate *et al* 2000). In eukaryotes interaction with the Sec61 channel is the only factor that leads to signal peptide release from SRP. Only after signal peptide release do SRP and SRα hydrolyse GTP causing dissolution of the RNC: SRP: SR complex, ready for new rounds of targeting (Miller *et al* 1993, Miller *et al* 1994, Powers and Walter 1995).

The co-translational process is similar in bacteria, however the signal recognition particle is named Ffh (fifty-four homologue) and is much smaller at 48 kDa, composed of a shorter 114 nucleotide 4.5S RNA component and an SRP54-like polypeptide (Romisch *et al* 1989, Bernstein *et al* 1989, Ribes *et al* 1990). It is important to note

that the co-translational route is only used in bacteria for a portion of integral membrane proteins (Macfarlane and Muller 1995). Ffh binds the signal peptide as it emerges from the ribosome forming the RNC: Ffh complex whereupon it then associates with the membrane-anchored FtsY, a homologue to SRα. There as yet appears to be no SRβ homologue in bacteria (Ladefogen and Christiansen 1997). FtsY is located at the inner cell membrane of bacteria as there is no ER (Luirink *et al* 1994, de Leeuw *et al* 1997). Here it is the charge on the head groups of membrane phospholipids that promotes release of the signal peptide by Ffh (Luirink *et al* 1994). The scene from this point is very similar to that in eukaryotes – FtsY guides the RNC complex to the core components of the Sec machinery, SecYEG.

SR or FtsY directs the RNC complex to the Sec channel. According to the most recent single-particle EM, resolved to 17 Å, the eukaryotic ribosome-channel complex is composed of four Sec61α/γ/β assemblies per translating ribosome (Manting *et al* 2000, Menetret *et al* 2000). The channel components are arranged as two side-by-side dimers which are themselves two back-to-back monomers. This arrangement is corroborated by the packing of 2D crystals for the *E. coli* SecYEG (Collinson *et al* 2001, Breyton *et al* 2002). Despite four Sec61 channels being present, only one is active and this sits within 12 Å – 15 Å of the ribosomal 'exit' site, ready to receive the translated polypeptide (Mitra *et al* 2005). The remainder of Sec61α/γ/β complexes present are thought to aid recruitment of other soluble factors on both the cytosolic and luminal side of the membrane. For this quartet arrangement of Sec complexes, it is the β (SecG) subunit that forms the significant interface between complexes and this could be the reason for the stimulatory effect seen when SecG is added to *in vitro* reconstitution assays (Nishiyama *et al* 1993). Biochemical data suggests it is the RNA component of

the ribosome that provides predominant contacts with each of the four Sec61α subunits and serves to 'anchor' the ribosome at the membrane. In this arrangement a newly translated polypeptide is passed directly from the ribosome into the membrane-spanning channel and thus it is the hydrolysis of GTP by the ribosome that drives translocation, 'pushing' polypeptide through the channel.

1.2.3 Post-translational translocation

If a polypeptide is not translocated as it emerges from the ribosome, it crosses the requisite membrane after it has been fully synthesised – post-translationally. Membrane crossing in this manner is the predominant route for bacterial polypeptides, although eukaryotic cells have a post-translational mechanism also. Unlike proteins that are translocated while being synthesised by the ribosome and thus utilise the energy derived from GTP hydrolysis to cross the membrane, proteins that cross post-translationally via the Sec route get the necessary energy from ATP and PMF. In eukaryotes, such as yeast, several channel partners are required to facilitate post-translational translocation. These are a tetramer composed of Sec62p/63p/71p and 72p that come together with the core Sec machinery to form a 7-component Sec complex. The situation is similar in mammals except there are no Sec71p/72p homologues. Once at the channel, polypeptides slide back and forth across the membrane due to Brownian motion. This alone would not achieve transport and therefore directionality is achieved through binding of a HSP70-like protein in the ER lumen termed BiP. BiP has an ATP-binding domain and in the ATP-bound state, has an open peptide-binding pocket and readily binds incoming polypeptide from the Sec channel. BiP interacts with Sec63p at the J-domain, an interaction which increases the affinity of the peptide binding pocket for polypeptide and also stimulates ATP hydrolysis, causing the pocket to close around the polypeptide (Misselwitz *et al* 1999). Thus, BiP is ''locked'' onto the

incoming polypeptide and prevents backsliding. After sufficient forward movement of the polypeptide another ATP:BiP molecule binds to the polypeptide and the loss of contact of the ADP:BiP with the J-domain results in ADP-ATP exchange and opening of the peptide binding pocket. In this manner the polypeptide is gradually ‘’ratcheted’’ into the ER lumen (Matlack *et al* 1999, Haigh and Johnson 2002). It is important to note that BiP binding to polypeptide only occurs in the vicinity of the J-domain, therefore BiP does not interfere with downstream folding stages within the ER lumen.

The post-translational scenario differs in bacteria in three key ways: first, bacteria employ a soluble partner in the *cytosol*, named SecA, that interacts with incoming polypeptide and utilises ATP and the PMF to move substrates through the SecYEG channel (Lill *et al* 1989). SecA is a large, 102 kDa protein of 901 amino acids and is composed of several domains, two of which are nucleotide binding domains (NBD), one with a high affinity for ATP (NBD1) the other with a lower affinity for ATP (NBD2) (Papanikolau *et al* 2007). Second, unlike the ‘’ratchet’’ mechanism of BiP, SecA acts as a ‘piston’, pushing polypeptides across the membrane. Finally, SecA is assisted by another soluble cytosolic protein, SecB (Driessen 2001).

A polypeptide destined for the post-translational route emerges from the ribosome, the slightly higher hydrophobicity of its signal peptide decreases affinity for Ffh and increases affinity for SecB. Multiple copies of SecB bind pre-protein as it is translated to maintain it in a Sec-translocatable conformation *i.e.* unfolded (Kumamoto 1989).

Under resting conditions, SecA exists in an equilibrium of two states – the less abundant state is a soluble dimer in the cytosol and the larger portion as a monomer at the membrane, associated with the acidic phospholipid head groups and the SecYEG channel (Lill *et al* 1990). When not bound to SecB or pre-protein, SecA only transiently interacts with SecYEG, however interaction of SecA and SecYEG

increases the affinity of SecA for SecB: pre-protein complex (Hartl *et al* 1990). Binding of the pre-protein signal peptide to SecA increases SecA-SecB interaction and leads SecB to disengage from pre-protein. This greater interaction between SecA and SecB leads to exchange of ADP for ATP at NBD1, which in turn causes the release of SecB from the SecA: pre-protein complex and insertion of the 30 kDa domain of SecA into the SecY channel. Accompanied by the 30 kDa domain is 20 – 30 amino acids of the pre-protein polypeptide. Hydrolysis of ATP at NBD1 promotes ATP hydrolysis at NBD2 which facilitates de-insertion of the 30 kDa domain from the SecY channel (Natale *et al* 2005). Pre-protein is now in the channel and could easily 'backslide' into the cytoplasm, however the PMF, although not essential for Sec-mediated protein translocation, is thought to stimulate translocation at this stage by decreasing backsliding (Schiebel *et al* 1991). In this manner, multiple rounds of SecB interaction, ADP: ATP exchange and pre-protein binding enable SecA to push pre-protein through the SecYEG channel across the membrane (Economou and Wickner 1994).

In *E. coli* a second heterotrimeric complex interacts with the SecYEG channel during protein translocation and, while not essential *in vitro* (Matsuyama *et al* 1992, Duong and Wickner 1997), deletion mutants lack viability and protein export *in vivo* (Pogliano and Beckwith 1994). This second heterotrimeric complex is found as an operon on the *E. coli* genome and is composed of SecD, SecF and yajC. All three are integral membrane proteins of 67 kDa, 35 kDa and 12 kDa respectively. SecD and SecF are composed of 6-TM domains and a large periplasmic domain (the periplasmic domain accounting for 45 kDa in SecD and 11 kDa in SecF) whereas yajC has a single TM domain and a large cytosolic domain. Despite the importance of this heterotrimeric complex *in vivo*, the exact function of SecDFyajC is largely unknown.

1.2.4 Membrane insertion and YidC

Co-translational translocation in bacteria is a route primarily reserved for integral membrane proteins (de Gier and Luirink 2003). During translocation, when a TM domain of pre-protein (which by default is largely hydrophobic) is spanning the SecY channel, the 'clamshell' that is the channel opens, allowing the TM domain to slide laterally into the membrane. A subset of *E. coli* integral membrane proteins require the aid of YidC. YidC is 62 kDa and is a homologue to the proteins Oxa1p and Alb3 that perform similar functions in the mitochondrial inner membrane and thylakoid membrane respectively (Saaf *et al* 1998, van Bloois *et al* 2005). Like SecD and SecF, YidC has 6 TM domains and a large periplasmic domain and is thought to aid TM segments of translocating polypeptides slide out of the SecY channel into the membrane. Although only involved to varying degrees with a range of IMPs, YidC is essential for the insertion of the F_0 domain of the *E. coli* F_1F_0 ATPase and cytochrome *o* oxidase of the electron transport machinery.

1.3 The Twin Arginine Translocation pathway, Tat

Some proteins are more complex, and thus require the insertion of cofactors (MGD, FeS etc) or oligomerisation before achieving functionality. In eukaryotes, secreted proteins are transferred as an unfolded polypeptide chain to the endoplasmic reticulum (ER) via the previously described Sec pathway (section 1.2). While in the ER these polypeptides fold into secondary and tertiary structures and are modified as necessary (e.g. glycosylation) to acquire their native state.

Bacteria lack the compartmentalisation of eukaryotes and, while they do not encode glycosylated proteins, there are still a number of proteins that require cofactors and/or

oligomerisation. A sub-set of these reside in the periplasm, particularly those involved in, or related to, the respiratory pathway and electron transport chain.

In the early 1990's it became evident that there was a second secretory pathway functioning in bacteria and thylakoids, working in parallel to the Sec pathway. This was first identified transporting proteins into the thylakoid lumen of plant chloroplasts. Soon after, homologues were identified in bacteria that facilitated protein transport across the cell membrane to reach the periplasm. Interests piqued in this newly discovered pathway when it was shown to enable membrane crossing for several heterologous proteins the Sec system could not handle. The first of these was dihydrofolate reductase (DHFR) bound to methotrexate (Hynds *et al* 1998), a complex known to be incredibly tightly folded. Secondly, GFP was able to be exported to the periplasm of bacteria in fluorescent state when fused to a Tat-specific signal peptide (Santini *et al* 2001, Thomas *et al* 2001), a feat not possible when GFP folds in the periplasm after translocation via the Sec pathway. Thirdly, instances of membrane crossing for protein complexes were reported (Rodrigue *et al* 1999, Wu *et al* 2000), indicating the Tat system could transport large, folded substrates. Further studies confirmed this pathway did indeed facilitate folded protein transport across the lipid bilayer. Initially named the ΔpH-pathway in thylakoids due to the sole requirement of a proton gradient for translocation (Mould *et al* 1991 and Yahr *et al* 2001) and Mtt (membrane targeting and translocation (Weiner *et al* 1998)) in bacteria, it has since been re-named and will be referred to from here on as the Twin Arginine Translocase (Tat) pathway.

1.3.1 Components of the Tat pathway

The first identified component of the Tat machinery was Hcf106, found in the thylakoids of maize plant chloroplasts (meaning High chlorophyll fluorescence 106) (Settles *et al* 1997). This was closely followed by the discovery of the remaining components of the Tat machinery – Tha4 and cpTatC (Mori and Cline 2002). Homologues to these three proteins were quickly identified in bacteria.

Minimally the Tat machinery is composed of only two classes of proteins – TatA-like and TatC-like, however this varies greatly from organism to organism and more commonly the Tat system is composed of two TatA-like proteins that perform distinct functions, and a single TatC-like protein (Goosens *et al* 2013).

Such a trimeric Tat system is found in *Escherichia coli* with the TatA-like components named TatA and TatB while the TatC-like protein is simply named TatC (homologous to chloroplast Tha4, Hcf106 and cpTatC respectively). TatAB and TatC are encoded as a *tatABCD* operon located at min 86 (Figure 4) with another gene, *tatE*, located elsewhere in the genome. All are constitutively expressed, although to varying levels, with TatA found at 50- and 25-fold abundance relative to the core Tat components TatC and TatB (Jack *et al* 2001). Expression levels are reflected in component make-up of the final translocase (section 1.3.3). Although *ΔtatABCDE* strains are still viable they exhibit reduced vitality as the bacterial Tat system is important for energy metabolism, formation of the cell envelope, biofilm formation, heavy metal resistance, nitrogen-fixing symbiosis and bacterial pathogenesis (Berks *et al* 1996, Oschner *et al* 2002, Palmer *et al* 2005, McDonough *et al* 2008, and reviewed in Buck *et al* 2008).

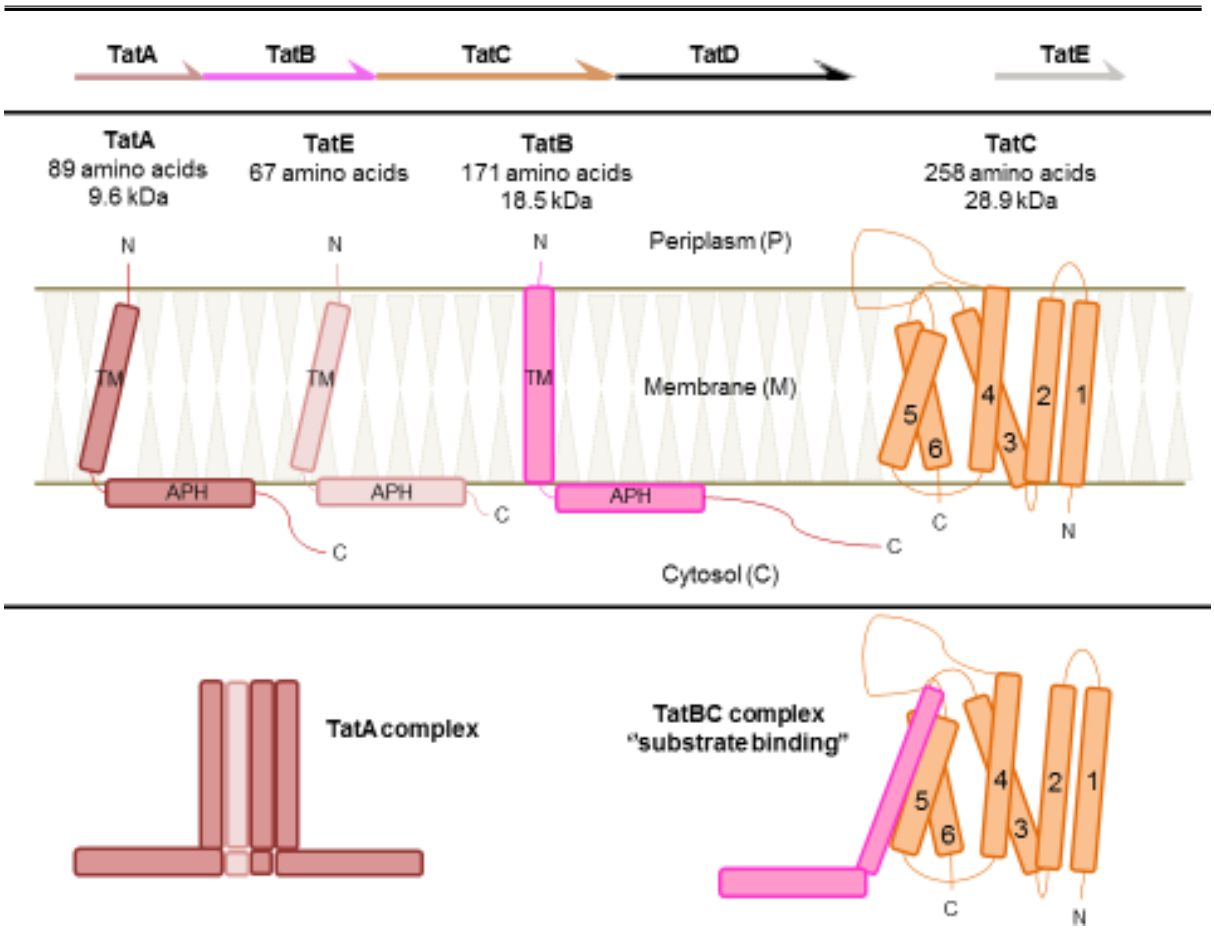


Figure 4: **Schematic of *Escherchia coli* Tat components and complexes.** TatAB and TatC are encoded in the *tatABCD* operon located at min 86 on the *E. coli* genome, while *tatE* is encoded at min 14 (top). Individual component size and orientation in the membrane are shown in the middle panel. Generally accepted complexes are shown in the bottom panel.

1.3.1.1 TatA

TatA is an 89 amino acid, 9.6 kDa protein with predominantly α-helical secondary structure (Weiner *et al* 1998). It consists of a small, unfolded N-terminal domain exposed to the periplasm (Walther *et al* 2010, Koch *et al* 2012), followed by a transmembrane (TM) helix and an amphipathic helix (APH) ending in a C-terminal domain of ~40 residues that form an unstructured tail (Hu *et al* 2010, Rodriguez *et al* 2013). Solid-state NMR indicates the TM helix spans the membrane at a 17° tilt (Müller *et al* 2007, Walther *et al* 2010), while CD and oriented CD spectroscopy shows the APH lies as a tangent to the membrane (Lang *et al* 2007). The TM-helix

and APH are connected by Glycine 21 (in *E. coli*) which is central to the ‘’hinge’’ region composed of the FGx motif (Barrett *et al* 2003) that also maintains the relative angle of the helices through packing interactions, known as the ‘’hinge-brace’’. The glycine at this position is one of the only residues conserved throughout TatA-like proteins and is essential for function. The amount of movement around this hinge region is under discussion, however it potentially plays a role in the translocation mechanism (section 1.3.4). The APH is hydrophobic facing the membrane and displays a charged face towards the cytoplasm. A recent solution NMR study of TatA has shown the APH to be anchored in the membrane by Phenylalanine 39, with substitution of this residue to alanine inactivating the translocase (Hicks *et al* 2003, Rodriguez *et al* 2013). TatA is thought to be the subunit responsible for forming the pore in the activated translocase. Cysteine scanning mutagenesis studies on TatA reveal that individual amino acids in the APH are critical to function, with single substitutions in this region commonly abolishing a functional translocase, whereas the only residue identified to be important in the TM domain is Gln8 (Greene *et al* 2007). A similar result was obtained in an elegant loss-of-function assay (Hicks *et al* 2005). This is in marked contrast to TatB in which the only essential specific amino acids appear to be those in the short N-terminal domain.

1.3.1.2 TatB

The second TatA-like protein in *E. coli* is TatB, a 171 amino acid, 18.5 kDa protein that despite sharing only 20% sequence identity with TatA, is predicted to fold to a highly similar structure (Hicks *et al* 2003). TatB possesses a short N-terminal region exposed to the periplasm, followed by a TM helix and the conserved glycine 21 residue in the ‘hinge’ region, this time in a xGP motif (Barrett *et al* 2003). Interestingly plant homologues Tha4 and Hcf106 (TatA and TatB respectively) each contain a FGP motif

(Barrett *et al* 2003). After the ‘’hinge-brace’’ region TatB begins to differ from TatA. The amphipathic helix is likely longer than that in TatA, extending by 12 amino acids (corresponding to ~ 3 extra turns of the helix) and the unstructured C-terminal domain is considerably longer (Zhang *et al* 2014). TatB also lacks the phenylalanine residue in the APH that is conserved in TatA (Phe39). Unlike TatA, TatB is not involved in the translocating pore, but plays an important role in the substrate-docking complex and thus despite having likely arisen from a gene duplication event (Yen *et al* 2002), TatA and TatB have diverged to perform different, but essential, functions (Sargent *et al* 1999). Interestingly, despite the relatively low level of sequence similarity between TatA and TatB it has been shown that only a few, minor mutations are needed in TatA such that it can complement *ΔtatB* strains (Blaudeek *et al* 2005). Complementary to this a chimeric TatA/B protein composed of the N-terminal TM domain of TatA fused with the APH and C-domain of TatB was able to support low levels of Tat activity in either *ΔtatAE* or *ΔtatB* strains.

1.3.1.3 TatC

The final core component of the *E. coli* Tat system is TatC. A considerably larger protein, TatC is 28.9 kDa formed of 258 amino acids that fold into six TM helices with both N- and C-termini in the cytoplasm (Behrendt *et al* 2004, Punginelli *et al* 2007). As well as being the largest of the Tat components, TatC has the most highly conserved amino acid sequence of all subunits across both bacteria and plants, with the majority of conserved residues residing in the cytoplasmic loops (Kneuper *et al* 2012, Ma and Cline 2013). The conserved nature of TatC is likely due to the central role it plays in recruitment of other Tat subunits and substrate binding and recognition. The recent crystallisation of *Aquifex aeolicus* TatC (which shares 40% sequence identity with *E, coli* TatC) in three environments to varying resolution (3.5 Å, 4.0 Å

and 6.8 Å), each identified the same structure for TatC indicating it has limited conformational states (Rollauer *et al* 2012). The six transmembrane domains of TatC form a concave 'bowl' structure facing the periplasm due to the shorter nature of TM5 and TM6 (Ramasamy *et al* 2013). These shorter helices contribute to the overall concave shape of TatC while also forming a cavity capped by a periplasmic loop and distort the membrane in their vicinity. This distortion of the membrane likely causes membrane thinning around the periphery of TatC and could be involved with the membrane-crossing event of Tat-targeted substrates. This cavity also harbours a glutamate residue (Glu165 in *A. aeolicus*) which, rather interestingly is conserved as a glutamate or glutamine in TatC components of all organisms (Buchanan *et al* 2002). Molecular dynamics simulations indicate this polar residue is hydrated and at least occasionally exposed to the cytoplasm thus could potentially play a role in signal peptide binding to TatC (Ramasamy *et al* 2013).

1.3.1.4 TatE

In *E. coli* the core subunits of the Tat machinery are TatABC, however in this organism there is also another TatA-like protein, TatE, located at min 14 on the genome. TatE is highly similar to TatA at 67 amino acids of which 57% are identical to that of TatA (Sargent *et al* 1998) and likely arose from a gene duplication event (Yen *et al* 2002). As with the TatABC components, TatE is expressed constitutively although to much lower levels, 50- to 100-fold less than TatA (Jack *et al* 2001). When overexpressed form an exogenous plasmid, TatE can functionally replace TatA (Sargent *et al* 1999), restoring export in *ΔtatA* strains, however its function under physiological conditions is as yet unknown although recent evidence suggests it is a regular, albeit scarce, constituent of native *E. coli* Tat complexes (Elmer *et al* 2015).

1.3.1.5 TatD

The final protein expressed from the *tatABCD* operon is TatD. Unlike the other Tat components, TatABCE, which are integral membrane proteins, TatD is a soluble cytoplasmic protein of 28.9 kDa. It has been shown to have no impact on Tat functionality and, due to the 33% sequence similarity to the *Chlamydophila pneumoniae* protein YabD and activity in the presence of Mg^{2+} is likely a magnesium-dependent DNase. It has thus been proposed that the only reason *tatD* resides in the Tat operon of *E. coli* is that it has similar transcriptional requirements.

1.3.2 Tat systems in other organisms

Nearly all bacteria, archaea and chloroplasts encode Tat systems, it is merely the number and variety of substituent components that differ. Utilisation differs from organism to organism with Gram-positive bacteria (e. g. *Staphylococcus aureus* and *Bacillus subtilis*) having few identified Tat substrates (Dilks *et al* 2003, Tjalsma *et al* 2004) whereas the enteric bacterium *E. coli* encodes 20 – 30 Tat substrates (Palmer, Sargent and Berks 2010). This variety is reflected in the encoded Tat systems, for example *B. subtilis* encodes two distinct Tat systems, both of which are minimal in that they are composed of only a single TatA-like and single TatC-like protein (Jongbloed *et al* 2004). These are TatAdCd encoded by the *phoD* operon which has a single substrate, PhoD, and is expressed under phosphate-limiting conditions and TatAyCy – a constitutively expressed Tat system with under 10 substrates. These Tat systems are variably expressed in parallel and are still able to maintain function. *B. subtilis* also encodes a third TatA-like protein, TatAc, the function of which is largely unknown, although it possibly plays a small role in both TatAdCd and TatAyCy systems. Indeed many gram-positive bacteria utilise Tat systems composed of a single TatA-like and TatC-like subunit (Jongbloed 2006). Structures have been resolved for

three of the identified TatA-like proteins to date – that of *B. subtilis* TatAd (Hu *et al* 2010) and the *E. coli* TatA and TatB (Rodriguez *et al* 2013, Zhang *et al* 2014a, Zhang *et al* 2014b). Rather unsurprisingly there is little difference between them. Moreover the TatA component (e.g. TatAd) is bifunctional in that it is able to complement either *tatA* or *tatB* null *E. coli* strains (Barnett *et al* 2008).

1.3.3 Structure of the TatABC complex

Unlike the Sec system, which transports pre-proteins in an unfolded state and thus only has to accommodate substrates of a single, small, defined size, the Tat system transports folded substrates. Folded substrates are by definition globular and of various shapes – ranging from 25 – 70 Å diameter (Berks *et al* 2000) – and sizes – encompassing small, 10 kDa high-potential iron-sulphur proteins to the large, 150 kDa formate dehydrogenase in *E. coli* (Brüser *et al* 2003, Jack *et al* 2004). Therefore, given the relatively small size of the individual Tat components a translocase must coalesce from multiple copies of each of the individual subunits.

Two such complexes have been identified to date – the first of these is composed essentially of TatB and TaC with trace amounts of TatA in *E. coli* and is known as the substrate receptor complex (Bolhuis *et al* 2001). These TatBC complexes are so named as there is significant evidence to show that this is where Tat-specific signal peptides bind (Lausberg *et al* 2012). Sequence conservation, site-specific cross-linking and genetic studies agree that the charged N-region of the Tat signal peptide (section 1.1) first binds to a surface patch on TatC formed by the N-terminal of TM1 and the cytoplasmic loop between TM2 and TM3 (Gerard *et al* 2006, Kreutzenbeck *et al* 2007, Rollauer *et al* 2012, Zoufaly *et al* 2012, Ma *et al* 2013) where the conserved ‘RR’ in the signal peptide possibly interacts with conserved ‘EE’ residues in the TatC loop (Berks, Lea and Stansfeld 2014). The remainder of the signal peptide then extends into

the interior of TatC, reaching the cavity formed by TM5 and TM6, where the C-domain of the signal peptide possibly reaches the periplasm at the periplasmic cap, exposing the cleavage site to signal peptidase (Berks, Lea and Stansfeld 2014). Similarly, cross-linking studies and modelling between TatC and TatB place the TM domain of TatB within, or near the cavity formed by TM5 and TM6 of TatC (Kneuper *et al* 2012, Rollauer *et al* 2012, Blummel *et al* 2014). Cross-links between the H-domain of signal peptide and TatB TM domain have also been observed (Aldridge *et al* 2014) and indicate TatB also plays a role in signal peptide binding (Maurer *et al* 2010).

It is well established that the substrate binding complex is composed of equimolar amounts of TatB and TatC (Bolhuis *et al* 2001), which in *E. coli* purify as complexes of 370 kDa, as determined by blue-native polyacrylamide gel electrophoresis (BN-PAGE) (Bolhuis *et al* 2001). Thus corresponding to ~ 7 copies of each subunit (Tarry *et al* 2009, Celedon *et al* 2012) and electron microscopy of TatBC reveals a hemispherical complex of 11 – 17 nm in diameter (Tarry *et al* 2009), large enough to account for an equivalent number of each subunit but not large enough to accommodate native Tat substrates and thus be the pore. The multiple copies of TatBC in these complexes would imply there are sites for multiple substrates to bind (Celedon *et al* 2012). It is unclear whether substrate binding to these sites are cooperative or not (Alder and Theg 2003a, Celedon *et al* 2012), however it is generally accepted that while the signal peptide enters deep within individual TatC units (Gerard *et al* 2007), the bulk of the substrate remains at the periphery of TatBC (Ma and Cline 2010, Tarry *et al* 2009). While TatA is not obligatory for the formation of TatBC complexed or substrate docking (Orriss *et al* 2007), there are trace amounts of TatA observed

associated with TatBC within the membrane and these are thought to be nucleation points for larger TatA oligomers to form to facilitate the membrane-crossing event (Aldridge *et al* 2014).

The second complex is highly heterogeneous in size and is composed predominantly of TatA with trace amounts of TatB. Interestingly these TatA oligomers are found in both the membrane and soluble in the cytosol (Keersmaeker *et al* 2005, Westermann *et al* 2006, Frielingsdorf *et al* 2008). Although the function of the membrane-embedded complexes is likely involved in the membrane-crossing event, the purpose of soluble TatA-like proteins remains unclear. These TatA oligomers appear as characteristic 'ladders' on BN-PAGE ranging in size from 100 kDa to 500 kDa (Oates *et al* 2005). TatA oligomers of this size only form once substrate has bound to TatBC in the presence of a PMF. While under non-translocating conditions TatA is thought to form tetrameric protomers (Leake *et al* 2008, Dabney-Smith and Cline 2009). Various studies have attempted to view these structures and identify their protein-protein interactions, as these would be important to elucidating the translocation mechanism, as TatA oligomers are thought to form the translocation 'pore'. However, it is difficult to isolate the active translocon due to their transient nature – TatA oligomers disperse upon successful translocation (Mori *et al* 2002). However, a recent study used site-directed, photo-activated cross-links to link a substrate to both TatA and TatB subunits, implying these both play a role in the translocation pore. Cross-links to TatB in this scenario could also be explained by the proximity of TatA and TatB binding sites on TatC. Current structural data on TatA-like oligomers are derived from TatAd purified from its native environment (the cell membrane) and solubilised in various detergents. Single-particle electron microscopy of these detergent-

solubilised TatAd complexes reveals ring-like structures of various sizes, matching the data of BN-PAGE. The larger of these ring-structures have internal cavities that are large enough to accommodate folded substrate proteins. One study controlled TatA oligomer size by altering the protein to detergent ratio in an effort to identify protein-protein contacts. Solution NMR of the smaller of these oligomers indicated TatA monomers interact through their TM domains while the APH 'splay' out from a central TM domain 'core'. Modelling showed this was a feasible scenario for up to 25 TatA subunits, at which point steric hindrance of residues in the APH limited further TatA monomer addition (Rodriguez *et al* 2013). These data match that garnered from spin-labelling (White *et al* 2010) and disulphide-bonded dimer experiments (Zhang *et al* 2014). These structures would be unstable however and require stabilising interaction with TatC. Cross-linking studies have placed TatA-TatC interactions between the TatA TM domain and the TM5 cavity of TatC (Berks 2015).

1.3.4 Mechanism of translocation

Translocation of proteins across the inner membrane of bacteria and thylakoid membrane initiates upon binding of substrate signal peptide to the TatBC or Hcf106-cpTatC respectively, substrate receptor complex (Alami *et al* 2003). Once docked at the membrane, TatA (or Tha4 in thylakoids) protomers are recruited in a ΔpH-dependent manner (Cline *et al* 2001, Mori *et al* 2002) to form the functional TatABC translocase.

The transient nature of the active translocase has proven difficult to 'capture' in the process of facilitating membrane crossing of a substrate. Given the heterogeneity of Tat substrates in size and shape it is difficult to conceive of a membrane-spanning channel, such as represented by the SecY protein. Such a channel would have to be quite sizeable for the Tat system to adequately accommodate the largest Tat substrates

and would prove difficult to inhibit the unwanted passage of ions and molecules across the membrane. Regardless, the initial stage of any translocation model for the Tat apparatus is guidance of substrate to the TatBC receptor complex and binding of the Tat-specific RR signal peptide at this site, a process likely driven by TatC (Frobel *et al* 2012). Following substrate docking the exact mechanism of membrane crossing becomes unclear; however, due to the experimental data available two possible mechanisms of protein translocation have been put forward.

The first is the 'Translocation pore' model, created and originally favoured due to the flexibility implied by the conserved glycine 'hinge' region in the TatA subunit (Gouffi *et al* 2004) and single particle electron microscopy data which identified TatA pore-like structures of various sizes (Gohlke *et al* 2005, Oates *et al* 2005). In this scenario, once substrate is docked at TatBC, TatA protomers are recruited to suite the size of the substrate protein with the TM helix forming the 'walls' of the pore and the APH forming a 'gate' at the cytoplasmic face of the membrane (Gohlke *et al* 2005, Walther *et al* 2013). Recruitment of protomers is supported by the homogeneity of TatE and TatAd complexes identified in *E. coli* and *B. subtilis* respectively, which, unlike TatA oligomers that range in size, are usually restricted to ring structures of a single size at ~160 kDa and a diameter of 6 nm – 9 nm, far too small to accommodate the larger Tat substrates (Baglieri *et al* 2012, Beck *et al* 2013).

Substrate would then push through this 'gate' with the APH hinging into the channel such that the hydrophobic face aligned with complementary charges on the TM helix. The charged face would then line the channel, providing a suitably hydrophilic environment for substrate to pass through. Once substrate had passed, APH would

again lie parallel to the membrane at the cytosolic face, closing the channel and limiting passage of ions. Finally, the TatA pore dissociates. Recent evidence has cast doubt on this model however as the APH is not accessible from the periplasm (Koch *et al* 2012, Aldridge *et al* 2012) and the 'hinge' region possibly lacks enough flexibility to allow this hairpin arrangement (Walther *et al* 2010). Solution NMR of TatA oligomers also indicates that the APH fan outward from a central TM helix 'core' (Rodriguez *et al* 2013).

The second possibility and interesting alternative to the translocation pore model, is that the Tat transport mechanism works by 'Membrane destabilisation' (Brüser and Sanders 2003). This model focuses on the unconventional idea that the Tat components inherently destabilise the membrane in their vicinity. Membrane destabilisation in this manner arises from two areas – a possible disruption of phospholipid head groups by the proximity and angle of the APH (Brüser and Sanders 2003, Greene *et al* 2007) and the mismatch between the length of the TatA TM helix and the width of the bilayer (the TM helix is slightly shorter than the bilayer width) (Rollauer *et al* 2012, Rodriguez *et al* 2013). This latter disruption could be enhanced by the disruption caused by TM5 and TM6 of TatC. Molecular dynamics simulations have shown that disruption of the membrane due to these effects could thin the bilayer by 80% with the phospholipids in the vicinity highly disordered (Rodriguez *et al* 2013). Substrate movement through this weakened and thinned bilayer would be far easier, especially if mechanical force is applied, whilst also maintaining the ionically-sealed nature of the membrane (Berks 2015). During membrane crossing fewer ions are likely to leak or the disordered nature of phospholipids in the vicinity could help to form a 'gasket' around the translocating protein, similar to the constriction point in

the SecY channel. Several studies have implicated the TatBC substrate-binding complex, specifically TatC in providing the mechanical energy for such a mechanism, possibly by utilising the PMF to 'pull' on the bound signal peptide (Lindenstraub *et al* 2009, Taubert *et al* 2014). These mechanisms are summarised in Figure 5.

Despite the abundance of data available, due to the lack of- and inherent difficulty in-isolating an active Tat translocon, understanding the exact mechanism of Tat transport is still a formidable challenge.

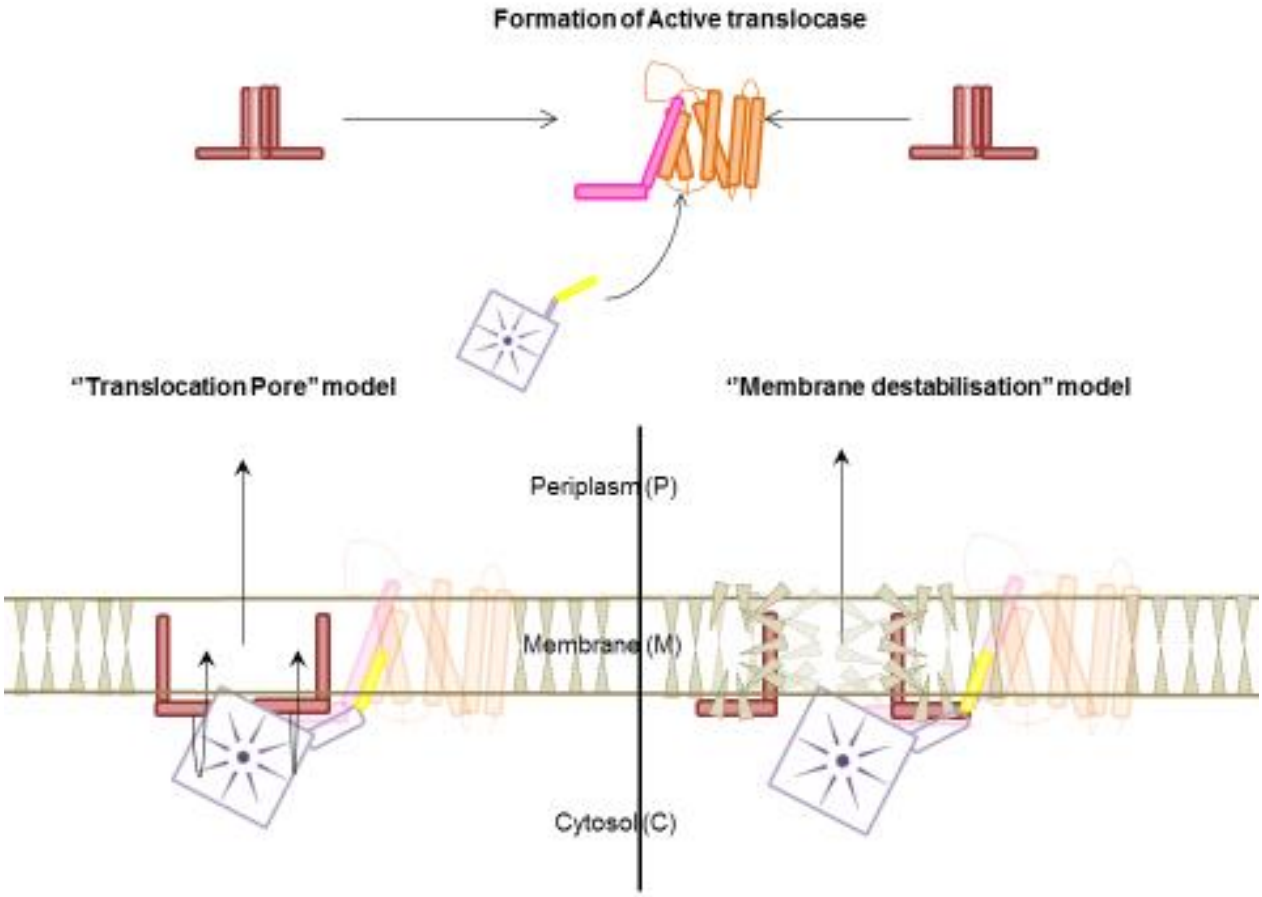


Figure 5: **Proposed translocation mechanism for the Tat pathway.** While the exact mechanism of translocation is unknown, the initial stage in any Tat translocation model is targeting and binding of substrate to the TatBC substrate receptor complex by the N-terminal signal peptide. At this point, translocation becomes less defined, however two scenarios have been envisioned. 1) Translocation pore model: TatA protomers create a channel to fit the size of the substrate with the TatA TM helix lining the channel while the APH form a ''gate''. 2) Membrane destabilisation model: The short nature of TatA TM helix causes deformation and thinning of the membrane in the local vicinity of TatA by up to 80%, allowing substrate to 'slip' across. In either model, once TatA protomers have coalesced, substrate proceeds to traverse the membrane before the signal peptide is removed by leader peptidase and substrate released into the periplasm.

1.3.5 Energy requirement of the Tat system

The current understanding for translocation by the Tat system indicates it is powered solely by the PMF (ΔpH). This is an unusual feature for protein translocation machinery, which are usually powered by hydrolysis of nucleoside triphosphates (ATP or GTP).

Early research into the energetics of the Tat pathway were carried out in the plant thylakoid system. This was due to the ease of preparing chloroplasts for both *in vivo* and *in vitro* studies combined with the ability to control the presence of a ΔpH by the presence or absence of light (Mould and Robinson 1991, Mori and Cline 2002). These found that import of the 23K protein of the oxygen evolving complex into thylakoids was only accomplished in the presence of light and a ΔpH and still occurred in the absence of nucleotides (Mould and Robinson 1991, Mori and Cline 2002). Due to a decrease in rate of import in D_2O it was thus proposed that translocation of proteins across the membrane was coupled to the transfer of protons in the opposite direction, in an antiporter mechanism (Musser and Theg 2000). The average amount of counterflow of protons necessary for a protein to cross the membrane by the chloroplast pathway was calculated to be in the region of 7.9×10^4 protons per molecule. This corresponds to the energy contained in 10 000 ATP molecules, 3% of the energy produced by chloroplasts, a considerable cost to the cell (Alder and Theg 2003b).

A similar dependence solely on the ΔpH was observed for the TatABC system of inverted membrane vesicles (IMV) prepared from *E. coli* (Alami *et al* 2002).

1.3.6 Proofreading and Quality control

Proteins identified as being routed via the Tat machinery to cross the inner membrane generally fit into three categories: 1) proteins binding complex cofactors that require enzymatic insertion, a process more suited to the protected environment of the cytoplasm and avoids exporting multiple chaperone proteins as well as the cofactor(s), which may require an extra transport mechanism of their own. 2) Proteins allowing sites potentially sensitive to external metal ions (such as the active site), to be sequestered or bind a suitable ligand first. 3) Proteins forming into hetero-oligomeric complexes that would otherwise not assemble in the periplasm (Patel, Smith and Robinson 2014). All of the three categories undertake processes that are complex and could easily go awry, resulting in malfunctioning protein being produced – a disastrous result for the cell. The majority of Tat substrates identified to date are cofactor-binding proteins, where the most abundant cofactors are Fe-S, molybdopterin and Ni-FeS centres, with a few examples of oligomeric complexes. However, a few substrates fit none of these categories and thus likely fold too rapidly to be reliably exported via Sec (e.g. AmiA, AmiB and AmiC).

There is widespread evidence indicating the Tat system has a proofreading and quality control mechanism to ensure such malfunctioning substrates do not cross the bilayer and are instead degraded (Halbig *et al* 1999, DeLisa *et al* 2003, Matos *et al* 2008).

It is apparent that native Tat substrates undergo stringent quality control prior to even reaching the Tat apparatus, particularly oxidoreductases for example trimethylamine-*N*-oxide (TMAO) reductase and dimethylsulfoxide (DMSO) reductase. These Tat substrates are encoded in an operon with subunit partners and also additional proteins that have no role in the final complex. These latter proteins have chaperone-like

features and are involved in the biogenesis of cognate cofactors and maturation of the final Tat substrate, thus they have been dubbed redox enzyme maturation proteins, REMPs.

REMPs perform many functions in oxidoreductase maturation including: protease protection of newly synthesised proteins, RR signal peptide interaction, maintaining the oxidoreductase in a cofactor-competent state, membrane interaction and targeting to the Tat apparatus. Tight binding of REMPs to their cognate protein is usually in the C-domain of the Tat signal peptide. This would serve to block the AxA motif to retain targeting by the signal peptide whilst retarding the journey to the translocase, allowing the pre-protein time to suitably fold and acquire necessary cofactors before docking at TatBC. For example, in the *torCAD* operon *torA* encodes TMAO reductase and *torC* encodes a haem-binding quinol oxidase, meanwhile *torD* is the REMP associated with TMAO. TorD has been shown to bind to apo-pre-TMAO in the cytoplasm at two sites – the C-domain of the signal peptide and with lower affinity to a sequence of the mature protein that is only in close enough proximity to the signal peptide in the folded state. In this manner TorD systematically aids protein folding and cofactor insertion while blocking premature cleavage of the signal peptide.

DMSO reductase was one of the first proteins utilizing the Tat apparatus to be studied. DMSO reductase is a heterotrimeric redox enzyme composed of the 85.8 kDa DmsA (which represents the catalytic subunit and binds a molybdopterin cofactor), the 23.1 kDa, Fe-S DmsB subunit and the 30.8 kDa DmsC subunit that acts as a membrane anchor. DmsA is synthesized with the Tat signal peptide and forms a complex with DmsB prior to export and is an example of the ‘’hitchhiker’’ mode of Tat export.

DmsD is the REMP associated with DMSO and plays an essential role in the biogenesis and maturation of DmsA (Ray et al 2003). It has been proposed that DmsD acts as the central ‘’hub’’, guiding DmsA through a cascade of accessory proteins for biogenesis. For example, DmsD has repeatedly been shown to have interactions with GroEL (the general molecular chaperone implicated in folding/refolding assistance) and members of the molybdenum cofactor biosynthesis pathway, especially MoaA and MoeB (Li *et al* 2010, Chan 2015 *et al*). Similarly, TorD, the REMP associated with maturation and biogenesis of TorA, has been shown to ‘hold’ domains I-III of apo-TMAO in a folded conformation while presenting the unfolded MGD (molybdopterin guanine dinucleotide) domain, domain IV, to the MGD synthesis enzymes (Dow 2013 *et al*). Acting as a hub not only for MGD insertion into TMAO but also for co-ordination of the various enzymes involved in the synthesis of molybdopterin and the further step of GMP incorporation to yield MGD. Upon binding of MGD, pre-TorA is released to the Tat apparatus for translocation, potentially in a GTP-dependent manner (Guymer *et al* 2010).

Another example is the Nap nitrate reductase in *E. coli* which is expressed from the large operon *napFDAGHBC* and is an important contributor to the proton gradient under anaerobic conditions. Here NapABC form the final periplasmic complex and all bind cofactors, where NapA is the nitrate reductase and binds a molybdopterin cofactor, while NapC is a quinol oxidase that transfers electrons through NapB to NapA (Potter and Cole 1999). NapC and NapB bind tetrahaem and dihaem respectively. The only other essential component is the REMP, NapD, which helps to coordinate cofactor synthesis and insertion.

Despite having high affinity for cognate signal peptides, REMPs are highly promiscuous (Chan *et al* 2009), aiding targeting and maturation of similar Tat substrates, although with reduced efficiency. For example, fusing the HybO signal peptide to mature *torA* and expression of the subsequent gene in *ΔtorD* strains still correctly localises TMAO to the periplasm, although with ~30% TMAO reductase activity. A similar result is obtained when the TorA signal peptide is fused to DmsA and expressed in *ΔdmsD* strains (Chan *et al* 2009).

Once at the Tat apparatus, a second process takes place, one of proofreading. Various studies report the ability of the Tat apparatus to reject substrates with a sufficiently destabilised structure (Halbig *et al* 1999, Maurer *et al* 2009 and for review see Robinson *et al* 2011), in some cases appearing to target them for degradation (Matos *et al* 2008, Goosens *et al* 2014). It is not quite clear at what stage this proofreading takes place and various studies have sought to determine the exact attribute the Tat machinery 'senses' in order to identify sufficiently folded precursors. Both native and heterologous substrates of the Tat pathway have been used to test the level of accuracy of Tat proofreading. For example, the necessity of cofactor insertion was addressed with the native Tat substrates NrfC and NapG, both proteins requiring FeS centre insertion in the cytoplasm prior to export. Here cysteine ligand sites were substituted to block cofactor binding, resulting in abolishment of Tat-mediated export (Matos *et al* 2008). Other studies have focused on the level of conformational stability necessary to achieve transport via the Tat pathway. In a study by DeLisa *et al* (2003) the necessity of disulphide formation in substrates was addressed through the use of the Sec substrate PhoA, an scFv (antidigoxin) and an F_{AB} fragment. These were fused with FdnG and TorA signal peptides respectively before expression in wild-type and *ΔtrxB*

Δgor ΔaphC ΔdsbA strains of *E. coli* (designated Cox, as they have an oxidising cytoplasm) and assessed for export competence via the Tat pathway. In this study, only expression in Cox strains resulted in protein export to the periplasm via the Tat pathway indicating formation of inter- and intra-disulphide bonds is necessary for a Tat substrate to be 'seen' as folded by the Tat proofreading mechanism. Slightly counter to this, recent studies have shown that both native and heterologous proteins targeted to Tat are viable for translocation in apo- and reduced states. A study by Alanen *et al* (2015) identified several heterologous proteins, including human growth hormone (hGH), an scFv and an interferon $\alpha_2\beta$ (IFN) were exported to the periplasm via the Tat pathway in the absence of disulphide formation. While neither of these studies identified the extent to which disulphide formation contributes to the structural stability of the mature protein, it is known that hGH obtains a near-native, active state in the absence of disulphides (Bewley *et al* 1969, Youngman *et al* 1995). Additionally, a study of the *E. coli* cuprous oxidase CueO, which also binds Cu^{II} at several copper centres and is a native Tat substrate, revealed translocation occurs with the substrate in the apo-form (i.e. without bound copper). The structure of CueO was shown by far-UV circular dichroism (CD) to be structurally identical in the apo- and mature forms, however (Stolle *et al* 2016).

Several studies have shown varying lengths of the hydrophilic FG-repeats taken from the yeast nuclear pore protein Nsp1p, that intrinsically lack structure, are able to cross the bilayer via the Tat pathway when fused to a Tat-specific signal peptide or folded proteins in various conformations (Richter *et al* 2007, Lindenstraub *et al* 2009, Taubert *et al* 2014). In the earliest of these, translocation efficiency decreased with increasing length or until a short hydrophobic stretch was inserted within the FG-repeat chain (Richter *et al* 2007). While in more recent studies extensive FG-repeats totalling 110

– 205 amino acids lead to late-stage translocation arrest (Lindenstraub *et al* 2009, Taubert *et al* 2014). These studies show that clearly some form of structural flexibility is accepted by the Tat proofreading mechanism and that proofreading possibly only takes place immediately before translocation occurs, not throughout the translocation process.

Complementary studies carried out in reconstituted thylakoid systems suggest the Tat machinery of thylakoids employ a proofreading process that is less stringent or has different criteria to that of the bacterial Tat system. Here Tat-directed substrates with inserted FG-repeats or (Gly_4/Ser) linkers of up to 75 amino acids successfully transported into the lumen, regardless of folding state, at least until a certain size-threshold was reached (Cline *et al* 2007). Early studies on the thylakoidal Tat system indicated a lack of a proofreading mechanism (Hynds *et al* 1998), or that it occurred on the *trans* side of the membrane after translocation completion (Roffey and Theg 1996).

Other approaches have sought to probe the proofreading process by looking at the Tat components themselves. The study published by Rocco *et al* (2012) used 3-helix-bundles that had various conformational flexibility fused to TEM-1 β-lactamase to screen for mutations within the TatABC subunits that resulted in a less-stringent proofreading mechanism, labelled quality control suppression mutants (QCS). By screening for export of a fusion that resembled the molten globule state and was thus export-incompatible in wild-type TatABC they identified residues that were key to allowing export of Tat substrates with increased conformational flexibility while maintaining the ability to reject completely mis-folded proteins. Interestingly the majority of these QCS mutations were confined to the unstructured or loop regions of

TatA, TatB and TatC. The fact that such QCS mutations can be identified proves that proofreading is at some level carried out by the Tat translocon.

A study by Goosens *et al* (2014) appeared to separate Tat-mediated quality control and proofreading by utilising mutagenesis studies. Using the Rieske Fe-S protein, QcrA, from the cytochrome bc_1 complex of Gram-positive bacteria, they substituted Cys residues involved in disulphide bridge formation or His residues needed for ligation of cofactor. Whereas cofactor-deficient QcrA was observed stalled at the membrane associated with the Tat machinery, QcrA lacking the disulphide was degraded in the cytoplasm. These results are interesting as they imply, for Gram-positive organisms at least, a ‘’hierarchy’’ between disulphide formation and, possibly REMP-assissted, cofactor insertion. The study did fail to identify the structural state of either of the mutant species however, thus one could be drastically mis-folded over the other introducing bias to the observed results.

1.4 Aims of this project

While several advances have recently been made into the possible mechanism of translocation employed by the twin-arginine translocase, information aided by high-resolution structures of the TatABC components solubilised in detergents, we are still little closer to understanding how exactly the proofreading mechanism functions and what is 'sensed' in a substrate.

This study aims to delineate some of the nuances of the Tat proofreading mechanism through alterations to surface characteristics and subtle changes to substrate folded state.

The following chapters cover:

- scFvM targeting to TatABC
- Maquette BT6 targeting to TatABC
- hGH comparison via Sec and TatABC
- Truncation of TatA and TatB components

2 Materials and Methods

Suppliers of chemicals

All chemicals used in this study were supplied by Fisher Scientific UK (Thermo Fisher Scientific Inc, USA), Sigma (Sigma-Aldrich, USA) or Formedium (UK) unless otherwise stated.

2.1 DNA Techniques

2.1.1 Preparation of plasmid DNA

All plasmids were isolated through mini-preparation using QIAprep spin miniprep kit (Qiagen, Hilden, Germany) as per the manufacturer's instructions. Plasmids were recovered in 30 µL elution buffer (10 mM Tris-Cl pH 8.5) and concentration determined using NanoDrop 2000c (Thermo Fisher Scientific Inc, USA) before use immediately or stored at -20°C until needed.

2.1.2 Amplification Polymerase Chain Reaction (PCR)

PCR solutions were made using 1 µL template DNA (80-100 ng), 1 µL dNTP mix (200 µM, Promega, USA), 10 µL GC buffer, 0.3 µL of each primer (0.5 µM), 1.5 µL DMSO (3%), 0.5 µL Phusion high fidelity DNA polymerase (2U/µL, New England Biolabs, UK) and made up to a final volume of 50 µL with milliQ H_2O (GenPure UV/UF water purification system). PCRs were generally carried out in Biometra T3 Thermocycler (Biometra Anachem, UK) and then analysed through agarose gel electrophoresis.

An example amplification reaction is given below:

Initial denaturation	98°C	30 sec	
Denaturation	98°C	10 sec	x35 cycles
Annealing	primer T_m minus 3-5°C	30 sec	
Extension	72°C	30 sec	
Final extension	72°C	5 min	
Pause	4°C		

2.1.3 Amplification (PCR) Primers

Table 1 shows all amplification primers used in this study, supplied by Eurofins MGW. The nucleotides encoding affinity tags are shown in bold while restriction sites are underlined.

2.1.4 Agarose gel electrophoresis

Agarose gels were prepared by adding 1% (w/v) agarose (Bio-Rad Laboratories Ltd, USA) to 50 mL 1x TAE buffer (0.04 M Tris-acetate, 0.001 M EDTA, pH 8.2). After mixing, the solution was microwaved at MAX power for a total of 45 seconds with occasional swirling until agarose had dissolved. Once set the gel was submerged in 1x TAE buffer. DNA samples were mixed with either 4 µL or 10 µL (restriction digest or PCR amplification respectively) 6x Gel Loading Buffer (Thermo Fisher Scientific Inc, USA) and 2 µL or 4 µL SYBER Green Nucleic Acid Gel stain (20x in DMSO, Invitrogen, Thermo Fisher Scientific Inc, USA) and left to sit for 5 min prior to gel loading. Electrophoresis was then carried out at 150V for 40 min. DNA was visualised and photographs taken using Bio-Rad Gel doc (Bio-Rad Laboratories Ltd, USA).

Table 1. Primers used for amplification PCR.

Primer	**Sequence**
scFvM study	
JASCFV1F	GGTGGACCCCATATGCAGGAACAGCTGGTTGAATCTGGT
scFvMARO	**ATG ATG ATG ATG ATG** GCG TTT GAT TTC CAC TTT TGT ACC
scFvMBRO	**ATG ATG ATG ATG ATG** ACG TTT GAT CTC AAC TTC GTG CCC
scFvMERO	**ATG ATG ATG ATG ATG** ACG TTT GAT CTC AAC CTT CGT GCC
scFvMGRO	**ATG ATG ATG ATG ATG** ACG TTT GAT TTC AAC TTT GGT GCC
scFvMHRO	**ATG ATG ATG ATG ATG** TTT TTT GAT TTC CAC TTT GGT GCC
scFvMIRO	**ATG ATG ATG ATG ATG** ACG ACG AAT TTC AAC ACG TGT ACC'
scFvMJRO	**ATG ATG ATG ATG ATG** TTT TTT GAT TTC CAC TTT GGT GCC CTG
His_stop_bamhi_R	GAATTCGGATCCGGATCCTTATT**AGTGATGATGATGATGATG**
Maquette BT6 study	
TorABT6KpnIF	TAA TAA GGT ACC CAT ATG AAC AAT AAC GAT CTC TTT
TorABT6XbaIR	GGC GGC TCT AGA CTC GAG TTA ATG GTG GTG ATG ATG
BT6KpnIF	TAA TAA GGT ACC ATG GCG CAA GCG GCG GGC GGC
TorABT6XhoIR	TTA TTA CTC GAG ATG GTG GTG ATG ATG GTG CAA TTG
BT6HisXhoIF	TAA TAA CTC GAG GCGCAAGCGGCGGGCGGCGACGGC

BT6HisXbaIR	TTA TTA TCT AGA TTA **GGC ATA ATC CGG AAC ATC ATA CGG ATA** CAG CTG TTT CAG ATC CTC AAA CTG

2.1.5 Purification of DNA from agarose gels

Appropriate bands were excised from agarose gels under UV transilluminescence using a sharp scalpel blade. Agarose was then removed using QIAprep gel extraction kit (Qiagen, Hilden, Germany) as per manufacturer's instructions. DNA was recovered in 30µL elution buffer before use or stored at -20°C until needed.

2.1.6 Restriction Digests of DNA

Typically, 2 µg DNA was mixed with 1 µL appropriate buffer and 20units 1st restriction enzyme and incubated at 37°C for 30 min (AccuBlock digital dry bath, Labnet, USA) before adding 20units 2nd restriction enzyme and incubation at 37°C for a further 1 hr. Table 2 shows restriction enzymes used in this study. DNA products were then isolated through agarose gel electrophoresis and purification as detailed previously.

Table 2. Restriction enzymes used in this study.

Enzyme	Sequence (5' → 3')	1st or 2nd
NdeI	CATATG	1st
BamHI	GGATCC	2nd
KpnI	GGTACC	1st
XbaI	TCTAGA	2nd
XhoI	CTCGAG	2nd

2.1.7 Ligation of DNA fragments into plasmid vector backbone

DNA fragments were cloned into open plasmid vectors using T4 DNA ligase (Roche, Sussex, UK). Insert and vector were mixed in a 3:1 ratio (typically 9 µL to 3 µL) with 1 µL ligase buffer and 1 µL T4 ligase then incubated overnight in Biometra T3 Thermocycler. The next day 10 µL ligation reaction was used to transform competent *E. coli* DH5α, XL1 Blue or NEB Turbo cells as described in section 2.2.4. The ligation program is given below:

4°C	30 min	∞ cycles
16°C	30 min	

2.1.8 Ligation of DNA fragments

DNA fragments to be joined were mixed in a 1:1 ratio and incubated overnight in Biometra T3 Thermocycler on ligation program as detailed in section 2.1.7. The next day ligation reaction was subject to agarose gel electrophoresis and appropriate fragments purified (section 2.1.4 and 2.1.5). DNA fragments were then amplified through PCR (section 2.1.2) and again subject to agarose gel electrophoresis to purify, before restriction digest with the appropriate enzymes and ligation into an open plasmid vector as described in section 2.1.6 and 2.1.7.

2.1.9 Site-specific DNA mutagenesis

Point mutations were introduced using *Pfu Turbo* alongside QuickChange site-directed mutagenesis kit (Agilent Technologies, Cheshire, UK) as per manufacturer's instructions. QuickChange mutagenesis was carried out in Biometra T3 Thermocycler. Afterwards 5 µL of DNA was used to transform competent *E. coli* DH5α or XL1 Blue cells as described in section 2.2.4.

2.1.10 Site-specific mutagenesis Primers

Table 3 shows all amplification primers used in this study, supplied by Eurofins MGW.

Table 3. Primers used for Quick Change PCR.

Primer	Sequence
scFvM study	
scFvMmutMA88LF	C TCT CTG CGT CTG GAA GAT ACC GCA GTT TAC TA
scFvMmutMA88LR	CGG TAT CTT CCA GAC GCA GAG AGT TCA TCT GGA
scFvMmutMS122LF	G GTT ACC GTG TTA TCT GGC GGT GGT GGC AGC GG
scFvMmutMS122LR	CAC CGC CAG ATA ACA CGG TAA CCA GAG TGC CTT
1SBF	CAAAGTTGAAGTGTCTTCTGGCGGTGGTGGC
1SBR	GTGCCTTGGCCCCAGAGGTCAGAG
scFvMwtC14F (C97S)	GCA GTT TAC TAT AGC GCG CGT GAT CTG
scFvMwtC14R (C97S)	CAG ATC ACG CGC GCT ATA GTA AAC TGC
scFvMwtC79F (C162S)	GCC ACT CTG TCC AGT ACG CTC TCT TCT
scFvMwtC79R (C162S)	AGA AGA GAG CGT ACT GGA CAG AGT GGC
TatABC study	

TatA-20F2	ACT ATC GCC GAT AAG CAG GCG TAA ACG AAT CAG GAA CAG GC
TatA-20R2	GCC TGT TCC TGA TTC GTT TAC GCC TGC TTA TCG GCG ATA GT
TatA-30F2	CC AGT CAG GAT GCT GAT TTT TAG GCG AAA ACT ATC GCC
TatA-30R2	GGC GAT AGT TTT CGC CTA AAA ATC AGC ATC CTG ACT GG
TatA-40F	AGC GAT GAT GAA CCA AAG TAG GAT AAA ACC AGT CAG GAT GC
TatA-40R	GCA TCC TGA CTG GTT TTA TCC TAC TTT GGT TCA TCA TCG CTC
TatA-45F2	GGC TTT AAA AAA GCA ATG AGC TAG GAT GAA CCA AAG CAG GAT AAA ACC
TatA-45R2	GGT TTT ATC CTG CTT TGG TTC ATC CTA GCT CAT TGC TTT TTT AAA GCC
TatB-20F	CCACGCCAGAGCGGTGGTAAAATAAGCTGCGGACGCTG
TatB-20R	GCGGTTTTCGGTTCAGCGTCCGCAGCTTATTTTACCACCG G
TatB-30F	CGCAGGCCAGTTCGCCGGAACAGAAGTGAGAAACCACGC CAGAGCCG
TatB-30R	ACCACCGGCTCTGGCGTGGTTTCTCACTTCTGTTCCGGCG
TatB-50F	CCGGTGGTGAAAGATAATGAAGCTTAGCATGAGGGCGTA ACG
TatB-50R	GCAGCGGCAGGCGTTACGCCCTCATGCTAAGCTTCATTAT C

TatB-75F	GCGGAGTCGATGAAGCGTTCCTACTAAGCAAACGATCCT GAAAAG
TatB-75R	CGCCTTTTCAGGATCGTTTGCTTAGTAGGAACGCTTCATC G
TatB-130F	GGCTGGATTCGCGCGTTGCGTTCATAGGCGACAACGGTG
TatB-130R	CAGTTCGTTCTGCACCGTTGTCGCCTATGAACGCAACGC

2.1.11 Sequencing of plasmid DNA

Sequencing was carried out using Beckman Coulter or GATC sequencing service.

Table 4 all sequencing primers used in this study.

Primer	Sequence	Use
pTacF	GAGCGGATAACAATTTCACACAGG	Sequencing pTac vectors
CM_pET23_SEQF	AGCTGTTGACAATTAATCGGC	Sequencing pET23 vectors

2.1.12 Constructs generated in this study

Table 5. Constructs generated and used in this study.

Plasmid	Function	Reference
scFvM study		
pYU49	pET23 based vector with pTac expression of protein of interest targeted to Tat via TorA signal peptide. Polycistronic expression of codon-optimised Erv1p and mature human PDI expression (CyDisCo).	Matos *et al* 2014
pHIA554	pET23 based vector with pTac expression of protein targeted to Tat via TorA signal peptide.	Matos *et al* 2014
pHAK13	pYU49 TorA-scFvM wild-type	Jones *et al* 2016
pAJ15	pHIA554 TorA-scFvM wild-type	Jones *et al* 2016
pAJ5	pYU49 TorA- 3SB (L11Q, Q13K, A88E, L112K, T114E, S116K)	Jones *et al* 2016
pAJ6	pYU49 TorA-5Lys (L11K, Q13K, A88K, L112K, S116K)	Jones *et al* 2016
pAJ7	pYU49 TorA-5Glu (L11E, Q13E, A88E, L112E, S116E)	Jones *et al* 2016

pAJ8	pYU49 TorA-1SB (L112K, T114E)	Jones *et al* 2016
pAJ9	pYU49 TorA-5R>K (R19K, R87K, R150K, R203K, R240K)	Jones *et al* 2016
pAJ10	pYU49 TorA-7K>R (K65R, K76R, K162R, K183R, K186R, K235R, K239R)	Jones *et al* 2016
pAJ22	pYU49 TorA-26tail (Additional - SNAIIIIITNKDPNSSSVDKLAAALE ' 3' to C-terminus)	Jones *et al* 2016
pAJ24	pYU49 TorA-2SB (A88E, L112K, T114E, S116K)	Jones *et al* 2016
pAJ31	pYU49 TorA-4NLeu (Q13L, R19L, K65L, K76L)	Jones *et al* 2016
pAJ35	pYU49 TorA-5Arg (L11R, Q13R, A88R, L112R, S116R)	Jones *et al* 2016
pAJ36	pYU49 TorA-5NLeu (Q13L, R19L, K65L, K76L, A88L)	Jones *et al* 2016
pAJ39	pYU49 TorA-6NLeu (Q13L, R19L, K65L, K76L, A88L, S116L)	Jones *et al* 2016
pAJ18	pYU49 TorA-scFvM (C97S)	Jones *et al* 2016
pAJ19	pYU49 TorA-scFvM (C162S)	Jones *et al* 2016
pAJ34	pYU49 TorA-scFvM (C97S, C162S)	Jones *et al* 2016
Maquette BT6 study		
pCM219	pEXT22 Single-copy (1 - 1.5 copies/cell) vector with pTac promoter.	Dykxhoorn *et al* 1996
pKG100	pJexpress414 BT6h2	This study
pKG101	pJexpress414 RR-BT6h2	This study
pKG102	pJexpress414 KR-BT6h2	This study
pKG103	pJexpress414 KK-BT6h2	This study
pKG104	pJexpress414 BT6h0 (H53A, H88A)	This study
pKG105	pJexpress414 RR-BT6h0 (H53A, H88A)	This study
pAJ21	pEXT22 RR-BT6h2	This study
pAJ25	pEXT22 RR-BT6h0 (H53A, H88A)	This study
pAJ26	pEXT22 KK-BT6h2	This study
pAJ27	pEXT22 KR-BT6h2	This study
pAJ28	pEXT22 BT6h2	This study

pAJ29	pEXT22 BT6h0 (H53A, H88A)	This study
pAJ40	pEXT22 RR-BT6h2-BT6h0 (H53A, H88A)	This study
pAJ41	pEXT22 KR-BT6h2-BT6h0 (H53A, H88A)	This study
pAJ44	pEXT22 RR-BT6h2-BT6h2	This study
pAJ45	pEXT22 KR-BT6h2-BT6h2	This study
pAJ46	pEXT22 RR-BT6h0 (H53A, H88A) - BT6h2	This study
pGS01	pEXT22 RR-BT6h1 (H53A)	This study
pGS02	pEXT22 BT6h1 (H53A)	This study
Tat truncation study		
pWM-Tat1	pLysSBAD (KasI digest to remove lysozyme) *Escherichia coli* TatABC	This study
pAJ48	pLysSBAD (KasI digest to remove lysozyme) *Escherichia coli* TatABC, TatA truncated by 20 amino acids	This study
pAJ50	pLysSBAD (KasI digest to remove lysozyme) *Escherichia coli* TatABC, TatA truncated by 30 amino acids	This study
pAJ54	pLysSBAD (KasI digest to remove lysozyme) *Escherichia coli* TatABC, TatA truncated by 40 amino acids	This study
pAJ52	pLysSBAD (KasI digest to remove lysozyme) *Escherichia coli* TatABC, TatA truncated by 45 amino acids	This study
pAJ62	pLysSBAD (KasI digest to remove lysozyme) *Escherichia coli* TatABC, TatB truncated by 20 amino acids	This study
pAJ59	pLysSBAD (KasI digest to remove lysozyme) *Escherichia coli* TatABC, TatB truncated by 30 amino acids	This study
pAJ60	pLysSBAD (KasI digest to remove lysozyme) *Escherichia coli* TatABC, TatB truncated by 50 amino acids	This study
pAJ61	pLysSBAD (KasI digest to remove lysozyme) *Escherichia coli* TatABC, TatB truncated by 75 amino acids	This study
pAJ63	pLysSBAD (KasI digest to remove lysozyme) *Escherichia coli* TatABC, TatB truncated by 130 amino acids	This study
human Growth Hormone Quality study		
pAJ13	pHIA554 OmpA-human Growth Hormone	This study
pHAK14	pHIA554 TorA-human Growth Hormone	Alanen *et al* 2015

2.2 Growth and Maintenance of *E. coli* cultures

2.2.1 Glycerol stocks

All constructs used and generated in this study were stored as glycerol stocks in either DH5α, XL1 Blue or NEB Turbo *E. coli* strains. Glycerol stocks were prepared by combining 1 part stationary phase culture with 2 parts 50% glycerol, then immediately frozen on dry ice and placed at -80°C for storage.

2.2.2 Preparation of competent cells

A 5 mL LB pre-culture, with appropriate antibiotics (at 1:1000 concentration), was inoculated with *E. coli* strain and incubated overnight at 30°C, 200 rpm in a shaking incubator (Multitron Pro shaking incubator, Infors, Switzerland). The next day 10 mL fresh LB was inoculated from the pre-culture and incubated at 30°C 200 rpm until OD_{600} = 0.3 – 0.4 at which point cells were pelleted by centrifugation (Jouan GR4_22) at 3000 rpm, 4°C, 10 min. Supernatant was discarded and cell pellet resuspended in 10 mL ice-cold 100 mM $MgCl_2$ and incubated on ice for 5 min. Cells were re-pelleted before resuspending in 1 mL ice-cold 100 mM $CaCl_2$ and stored at 4°C for 1 – 24 hrs before use.

2.2.3 Media

All liquid cultures were grown aerobically in Luria Bertani (LB) medium (10 g/L sodium chloride, 10 g/L tryptone, 5 g/L yeast extract), while Luria Bertani Agar (LBA) (10 g/L sodium chloride, 10 g/L tryptone, 5 g/L yeast extract, 10 g/L bacto-agar (Becton Dickinson and company, USA)) was used for aerobic growth on plates.

Antibiotics were added just prior to inoculation of cultures and were used at the following concentrations: Ampicillin (100 µg/mL), Kanomycin (50 µg/mL) and Chloramphenicol (35 µg/mL). All antibiotics were dissolved in milliQ H_2O except

Chloramphenicol which was dissolved in Ethanol (Fisher Scientific UK, Loughborough, UK).

2.2.4 Transformation of competent *E. coli* cells

In pre-cooled 1.5 mL eppendorfs 2 µL DNA (100 ng/µL) was gently mixed with 30 µL competent cells and incubated on ice for 30 minutes. The cells were then placed in a water bath (Grant Instruments, Cambridge, UK) at 42°C for exactly 30 seconds before placing back on ice for 5 min. Next 0.5 mL LB was added and cells incubated at 37°C for 60 min. After incubation, cells were pelleted at 6500 rpm for 5 min. After which 500 µL supernatant was discarded and cell pellet gently resuspended in remaining liquid (≤ 100 µL). Total cell suspension was plated on LBA plates containing appropriate antibiotic(s) which were incubated in Plate incubator (MLR-162-PE, Panasonic, Japan) inverted, at 37°C overnight.

2.3 Protein production

2.3.1 Culture of *E. coli* and Plasmid induction

For fractionation of *E. coli* cells (section 2.3.2 and 2.3.3), typically 5 mL LB with appropriate antibiotics was inoculated and grown aerobically at 30°C, 200 rpm overnight. The following day cultures were diluted to OD_{600} = 0.05 in 50 mL fresh LB, and antibiotic, in 250 mL erlynmeyer flask before incubation at 30°C until OD_{600} = 0.4 – 0.6. At this point plasmid(s) were induced. Several vectors were used throughout this study: pYU49 (pET23 based) and pEXT22 were typically induced with 0.5 mM IPTG while pBAD24 and pWM-Tat0 were induced with 20 µM freshly made Arabinose.

2.3.2 Preparation of spheroplasts

Cell cultures were separated into Spheroplast (Sp) and Periplasmic (P) fractions using the EDTA/lysozyme/cold osmotic-shock method (Randall and Hardy 1986). After the appropriate induction period, typically 3 hrs, cells equal to OD_{600} = 10 were harvested by centrifugation at 3000 rpm, 10 min, 4°C. Supernatant was discarded and the cell pellet resuspended in 0.5 mL chilled Buffer 1 (100 mM Tris-acetate pH8.2, 500 mM sucrose, 5 mM EDTA) before addition of 0.5 mL chilled milliQ H_2O and 40 µL lysozyme (from egg-white, 1 mg/mL stock (Sigma Aldrich, UK)) and incubation on ice for 5 min. Subsequently 20 µL $MgSO_4$ (1 M stock) was added to stabilise inner membrane before pelleting by centrifugation at 14 000 rpm (Eppendorf Centrifuge, 5417R, Hamburg), 4°C, 2 min. 0.5 mL of supernatant was then collected as the Periplasmic fraction (P) and stored at -20°C. Remaining supernatant was discarded and the pellet was washed with 1 mL chilled Buffer 2 (50 mM Tris-acetate pH 8.2, 250 mM sucrose, 10 mM $MgSO_4$) and pelleted by centrifugation at 14 000 rpm, 4°C, 5 min. Supernatant was discarded and cell pellet resuspended in 1 mL Lysis buffer (20 mM sodium phosphate pH 7.3) and lysed by sonication (3 x 10 sec, 8 µM (Soniprep 150 plus, Sanyo Gallenkamp, Loughborough, UK)) and 2 x freeze-thaw cycles before addition of 10 µL DNase (10 mg/mL) and storage at -20°C.

2.3.3 Cytoplasm and Membrane separation

Spheroplasts were prepared as in section 2.3.2, however after washing with Buffer 2 the cell pellet was resuspended in Buffer 3 (50 mM Tris-acetate pH 8.2, 2.5 mM EDTA) and membranes lysed by sonication (6 x 10 sec, 8 µM). Membranes were then separated from the Cytoplasm by ultracentrifugation at 70 000 rpm (TL-100 Ultracentrifuge, TLA 100.3 rotor, Beckman Coulter Inc, USA) 30 min, 4°C. Following this 0.5 mL supernatant was taken as the Cytoplasmic fraction (C) and

remaining supernatant discarded before resuspending the pellet in 0.5 mL Buffer 3 as the Membrane fraction (M).

2.3.4 Separation of Insoluble fraction

To separate insoluble fraction (I) from membrane fraction (M) samples were spun at 14 000 rpm for 15 minutes, 4°C after the sonication step. The supernatant was drawn off and fractionation continued from the ultracentrifugation step (section 2.3.3).

2.3.5 Time course assay

To generate time courses of protein production and export to the periplasm, samples equal to OD5 were taken from 50 mL cultures at appropriate time-points after induction. Fractionation was carried out as detailed in section 2.3.2 and 2.3.3.

2.4 Protein purification

2.4.1 Immobilised Metal Affinity Chromatography (IMAC), Nickel

For purification of 6x Histidine-tagged (C-term) proteins by Nickel IMAC, 10 mL LB with appropriate antiobiotic(s) was inoculated and grown aerobically overnight at 30°C, 200 rpm. The next day cultures were diluted to OD_{600} = 0.05 into 400 mL fresh LB and antibiotic(s) in 2 L erlynmeyer flask and incubated at 30°C, 200 rpm. Plasmid(s) were induced with 0.5 mM IPTG when culture OD_{600} = 0.4 – 0.6 and then incubated at 30°C, 200 rpm for 4 hours. After the induction period, cells were harvested by centrifuge 3000 rpm, 30 min, 4°C (Beckman Avanti J-25, JA 10 rotor). A sample of the supernatant was taken (Media) and the rest discarded. Cell pellet was resuspended in 10 mL chilled Buffer 1 without EDTA (so as not to interfere with IMAC down-stream, 100 mM Tris-acetate pH8.2, 500 mM sucrose) and 10 mL milliQ

H_2O before adding 800 µL lysozyme and incubating on ice for 10 min. To stabilise membranes 800 µL $MgSO_4$ was added and ½ crushed tablet Complete EDTA-free protease inhibitor before centrifugation 14 000 rpm, 20 min, 4°C (Beckman Avanti J-25, JA 25.5 rotor). Supernatant was taken as the Periplasmic fraction and stored at 4°C. The remaining pellet was resuspended in 10 mL 50 mM sodium phosphate pH 7.2, 150 mM NaCl and sonicated (6x 10 sec) before adding 400 µL DNase and frozen at -80°C as the Spheroplast.

Prior to IMAC purification, Spheroplast samples were spun for 20 min 5000 rpm and the cell pellet discarded, the supernatant was further spun at 45 000 rpm for 20 min to remove unbroken cells and cell debris.

IMAC columns were run on the same day as Spheroplast/Periplasm fractionation. A bed volume of 5 mL IMAC sepharose 6 fast-flow (GE Healthcare, Buckinghamshire, UK) beads were added to a 20 mL syringe with glass wool used as a filter. Storage solution (20% ethanol) was washed off with 10 column volumes (CV) milliQ H_2O before adding 5 mL 0.2 M NiCl, followed by another 2CV milliQ H_2O wash. Columns were equilibrated with 3CV Equilibration buffer (50 mM sodium phosphate pH 7.2, 0.3 M NaCl) before loading Periplasmic sample and collecting Flow Through (FT). Unbound matter was removed with 6CV Wash buffer (50 mM sodium phosphate pH 7.2, 50 mM Imidazole, 0.3 M sodium chloride) and sample collected as Wash (W). Finally 6x Histidine-tagged protein was eluted in 9 x 2 mL fractions of Elution buffer with an imidazole gradient (50 mM sodium phosphate pH 7.2, 0.3 M NaCl with 100 mM, 150 mM, 200 mM or 250 mM Imidazole).

All samples (Media, Sp, P, FT, W, E1-9) were run on SDS-PAGE (section 2.5.1 and 2.5.2) and western blot (section 2.6.1) to find peak elution fractions, these were assessed for their purity and concentrated.

2.4.2 Protein quantification assay

Proteins were quantified by a Bradford assay (Bio-Rad *DC*™ Protein Assay, Herts, UK) as per manufacturer's instructions. Or by reading A_{280} from a quartz cuvette (Starna Scientific Ltd, Essex, UK) and using the molar extinction co-efficient to calculate concentration.

2.4.3 De-salting IMAC peak fractions

Peak IMAC elution's were pooled in dialysis tubing (SnakeSkin® dialysis tubing, Thermo scientific) and dialysed overnight into 50 mM sodium phosphate, 150 mM sodium chloride.

2.4.4 Concentrating IMAC peak fractions

After dialysis, proteins were concentrated with VIVASPIN 20 10,000 MWCO columns (Sartorius, Gloucestershire, UK) until an appropriate volume was reached, typically 0.25 – 0.5 mL.

2.5 Protein separation

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

Protein separation was carried out using a vertical gel apparatus from Bio-Rad Mini-PROTEAN® Tetra System (UK), the gels were cast and run according to manufactures instructions. Typically, 0.75 mm gels were used, however the Maquette work presented in chapter 4 (p. 96) use 1.5 mm gels. The separation gel was composed of: 15% acrylamide, 0.3% bis-acrylamide (37.5:1, BioRad, Herts, UK), 375 mM Tris-HCl pH 8.85 and 0.1 % SDS. Stacking gel was composed of: 5% acrylamide, 0.0375 % bis-acrylamide 125 mM Tirs-HCl pH 6.8 and 0.001% SDS. 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, 5% β-mercaptoethanol) and heated for 10 minutes at 50°C. Protein gel running buffer was 25 mM Tris, 192 mM Glycine and 0.1% SDS at a pH of 8.3. Gels were run until the dye-front had run off the gel, typically 40 minutes at 60 mA.

2.5.2 Detection of proteins with Coomassie

The protein profile of samples was visualised with Coomassie stain (0.05% w/v Coomassie Brilliant Blue R-250 (Bio-Rad, Herts, UK), 50% Ethanol, 1.74 M Acetic acid). Newly run SDS gels were covered in Coomassie stain and incubated at room temperature for 1 hr. Stain was subsequently removed and excess removed from the gel by incubation with D-stain (5% Ethanol, 7.5% Acetic acid).

2.6 Protein imaging

2.6.1 Western-blotting

After SDS-PAGE visualisation of C-terminal 6x His- or C-terminal HA-tagged protein was achieved by the wet-western-blot method. Protein was transferred to PVDF-membrane (GE Healthcare, Buckinghamshire, UK) running for 1 hr at 80V, submerged in wet-blot transfer buffer (192 mM Glycine, 25 mM Tris, 10% ethanol). Membrane was prepared by soaking in methanol prior to contact with the gel. After transfer, PVDF membrane was then immersed in blocking milk (2.5% (w/v) skimmed milk powder in 50 mL 1x PBS-tween20 (0.1%)) or 3% BSA solution at 4°C overnight.

2.6.2 Detection of proteins by immunoblotting

The following day blocked membrane was washed 3x 5 min with 1x PBS-tween20 (0.1%) before incubating primary antibody (in 20 mL 1x PBS-tween20 (0.1%), 3%

(w/v) BSA (Life technologies, CA, USA)) for 1 hour at room temperature on shaking table (GyroTwister, Labnet, New Jersey, USA). Membranes were then washed for 3x 5 min with 1x PBS-tween20 (0.1%) before incubating with secondary antibody (in 20 mL 1x PBS-tween20 (0.1%)) for 1 hour at room temperature on shaking table. Finally, membranes were washed for 8x 5 min with 1x PBS-tween20 (0.1%) before imaging. Antibodies and their concentrations used in this study are shown in Table 6.

Immunoreactive bands were detected using enhanced chemiluminescence detection kit (ECL, Bio-Rad, Herts, UK) according to the manufacturer's instructions. Bands were visualised using BioRad Gel-doc chemiluminescence imager and associated software. Comparison of band-intensity was also carried out on BioRad Gel-doc imager with appropriate software.

Table 6. Antibodies used in this study.

Primary antibody	**Concentration**	**Source**	**Secondary**	**Concentration**	**Source**
anti-hexahistidine (C-term)	1 in 8000	Invitrogen	anti-Mouse-HRP	1 in 5000	Promega
anti-HA (C-term)	1 in 1100	Sigma	anti-Rabbit-HRP	1 in 5000	Promega
anti-TatA	1 in 2000	Laboratory stock (Bolhuis *et al*2000)	anti-Rabbit-HRP	1 in 5000	Promega
anti-hGH	1 in 10 000	Laboratory stock (Alanen *et al* 2015)	anti-Rabbit-HRP	1 in 5000	Promega

Testing the proofreading capacity of the Twin-Arginine Transolcase through export of an scFv with altered surface charge.

The Tat system of *Escherichia coli* consists of the three membrane proteins, TatA, TatB and TatC and is the most well-understood TatAC-like system. The TatABC system functions in the cell membrane of *E. coli* to transport proteins across the plasma membrane. Tat substrates generally preclude export via Sec due to one of several reasons – rapid folding, co-factor binding or oligomerisation (Berks *et al* 2003, Rodrigue *et al* 1999, Wu *et al* 2000). Substrates are targeted to the Tat pathway via an N-terminal signal peptide which is then cleaved upon successful translocation (Stanley *et al* 2000), as discussed in section 1.1. Evidence suggests the Tat pathway has a proofreading mechanism to ensure incorrectly folded proteins are not translocated (section 1.3.6). A multitude of studies have sought to probe the mechanism by which Tat proofreads substrate wherein sufficient destabilisation of substrate structure results in 100% rejection by the Tat system (Halbig *et al* 1999, DeLisa *et al* 2003, Richter and Brüser 2005, Matos *et al* 2008). Initial studies in this area utilised the native *E. coli* protein alkaline phosphatase (PhoA) fused with a Tat-specific signal peptide (TorA), to show that without prior formation (i.e. in the cytoplasm) of disulphide bonds, these proteins were rejected for export via the Tat pathway (DeLisa *et al* 2003). Fortunately the Tat proofreading system is capable of functioning on heterologous proteins that it would not normally encounter in nature. Thus further studies utilised heterologous recombinant proteins (e.g. hGH, interferons and scFvs) to probe the Tat proofreading mechanism (DeLisa *et al* 2003, Alanen *et al* 2015). These further highlighted the ability of the Tat system to distinguish between folded and mis-folded substrates. The majority of these studies however, utilised either grossly mis-folded substrates or substrates for which the degree of miss-folding is unclear, and thus it is difficult to pinpoint exactly what it is that Tat 'senses' as 'unfolded'. Here we have sought to assess the proofreading mechanism of the Tat system through the

use of a biotechnologically relevant protein fused with a Tat targeting signal at the N-terminus.

scFv's are ideal candidates to probe the proofreading capacity of the Tat system as they are heterologous, they usually contain disulphide bonds and they are the initial step to the recombinant production of large, complex molecules such as bi-specifics and antibodies (e.g. IgGs). Here, we have taken an scFv (named scFvM) developed for use in a solubility study (Austerberry *et al* 2017) and used it to probe the proofreading capacity of the Tat system on a more nuanced level. The primary sequence of scFvM is shown in Figure 6 where various features are highlighted (e.g. bridge-forming cysteine pairs, TorA signal peptide and 6x histidine-tag). The rationale behind this study is two-fold – we have a number of scFvM variants for which the T_m, peak fluorescence and secondary structure (through CD) have been determined, thus for the first time we can probe the Tat system with substrates of defined structure. Second, the effect of structural flexibility can be observed through the export competence of the reduced and oxidised forms.

We first tested the necessity of prior disulphide formation on export for scFvM by expressing it in wild-type and CyDisCo strains. The cytoplasm of *E. coli* is naturally reducing, which inhibits the formation of disulphide bridges. Recently this has been overcome through the expression of a yeast mitochondrial thiol oxidase (Erv1p) and human protein disulphide isomerase (hPDI), collectively termed CyDisCo (Cytoplasmic Disulphide formation in *E. Coli*), in the cytoplasm of *E. coli* alongside protein(s) of interest (Hatahet *et al* 2010, Nguyen *et al* 2011). Second we created three groups of mutants designed to probe the proofreading mechanism of the Tat apparatus through

modifications to the surface charge of scFvM while having minimal effect on the secondary or tertiary structure. These groups are: 1) substitution of 2, 4, or 6 non-polar residues to create 1, 2 or 3 surface salt-bridges formed of Lys-Glu pairs. 2) Substitution of 5 non-polar residues with Lys, Glu or Arg to create a concentrated 'patch' of charge and finally, substitution of surface Lysines with Arginines or Arginines with Lysines to represent a more distributed charge.

Richter *et al* (2007) used short chains of hydrophilic FG repeats, some interspersed with a hydrophobic stretch to suggest the Tat system proofread by identifying exposed hydrophobic regions, thus we created a fourth group of scFvM variants where 4, 5 or 6 hydrophilic residues on the surface were substituted with Leucine.

Figure 6: **Primary amino acid sequence and structural features of scFvM.** The scFv in this study was fused to the signal peptide from TMAO reductase (yellow) along with the first four amino acids of the mature TMAO protein (green). Numbering starts at the first methionine residue of the mature scFv with residues forming restriction sites from NdeI and BamHI shown in italics. Bridge-forming cysteine residues are shown in orange and underlined (Cys23-Cys97 and Cys162-Cys232), while the $(Gly_4/Ser)_3$ linker between immunoglobulin domains is bold and underlined. The C-terminal 6x histidine-tag is shown in brown.

Results

scFvM is folded and active in the reduced form but is exported by Tat more efficiently in the oxidised form

In this study we aimed to get a more discerning view of the Tat proofreading capacity. To this end we chose a biotechnologically-relevant model protein as a substrate for which we had a well-defined structure along with multiple variants thereof. These variants had either altered surface features while maintaining tertiary structure, or the folding and thus tertiary structure was deliberately disrupted. Some of the variants with altered surface features are sourced from a biophysical study on the scFv in solution (Edwardraja *et al* 2010, Austerberry *et al* 2017), these are summarised in Table 7. The scFv is raised against c-Met, a protein which plays an important biological role in the proliferation of cancer cells (Edwardraja *et al* 2010). Although there is no crystal structure available, scFvM is homology modelled to PDB file 2GHW and is comprised of two immunoglobulin folds connected by a short 'linker' region traditionally used in scFv's ($(Gly_4Ser)_3$) (Miller *et al* 2010), shown in Figure 7 from two projections as a ribbon diagram, with surface potential and finally with surface hydrophobicity. Each domain contains a disulphide bond sequestered within the β-sheets (circled in Figure 7).

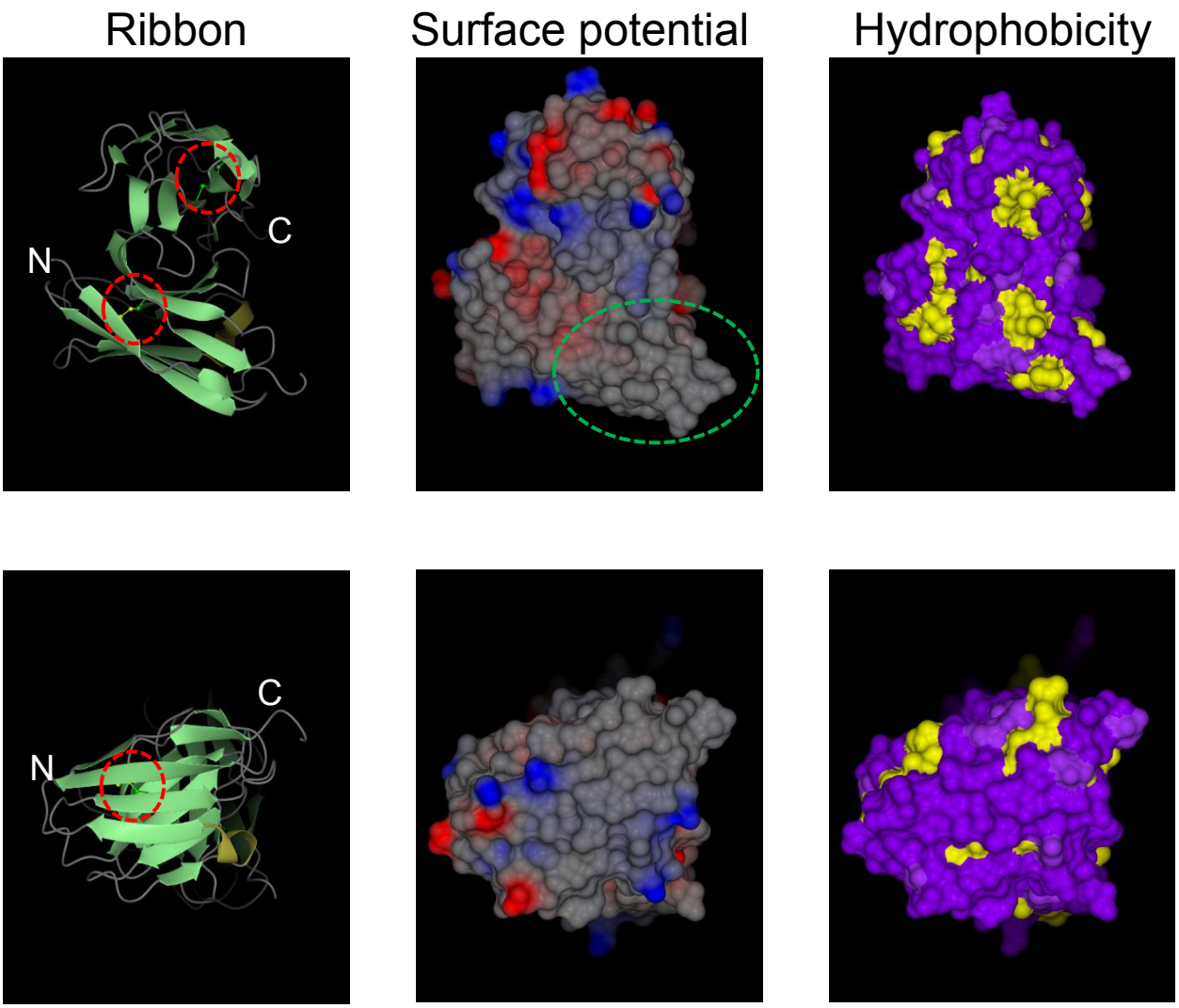


Figure 7: **Ribbon, surface potential and surface hydrophobicity images of scFvM from two projections.** Top panel: projection centred on patch that is uncharged in the wild-type. Bottom panel: projection centred on N-terminal. Ribbon diagram: β-sheets = light green, α-helix = gold, loop regions = grey. Disulphide bridges are circled in red. Surface potential: blue = positive, red = negative, grey = uncharged/neutral. The uncharged patch where salt-bridge or charge substitutions are located is circled in green. Hydrophobic surface: purple = hydrophilic, yellow = hydrophobic. Images were made using CCP4 Molecular Graphics.

Previous studies have shown that although the TatABC system has a discerning proofreading mechanism, some Tat-targeted heterologous proteins do not require prior formation of their disulphide bonds before export (Alanen *et al* 2015) and are thus less likely to be in a dynamic state i.e. they are rigid. These proteins are known to take on their native conformation, be stable and are active without formation of these disulphides (Bewley *et al* 1969, Youngman *et al* 1995). In light of this, we first tested the necessity of prior disulphide formation for export of scFvM, even though this protein is also soluble and active in the reducing cytoplasm of *E. coli* (Edwardraja *et al* 2010, Austerberry *et al* 2017). We verified the reduced state of the disulphide bonds in the *E. coli* cytoplasm and during purification through mass spectrometry analysis of scFvM immediately after purification with and without the addition of the reducing agent DTT. The resulting data show a mass that corresponds to the reduced state of 26, 806 Da in both cases, shown in Figure 8, and thus the biophysical/activity data presented here (see below) were carried out on the reduced form of scFvM.

scFvM was fused with a TorA signal peptide (similar to other studies on the Tat system e.g. Thomas *et al* 2001, shown highlighted in yellow in Figure 6) and expressed either in the presence or absence of the CyDisCo components, as a previous report by our group has shown a different scFv does not require formation of disulphide bonds prior to export (Alanen *et al* 2015). Figure 9.A and B show time-course analysis of export after induction with IPTG. An equivalent amount of cells were fractionated into cytoplasm, membrane and periplasm (C/M/P) after the time periods indicated in the wild-type strain (Figure 9.A) where the cytoplasm is reducing (*i.e.* in the absence of CyDisCo components). Protein is evident in all three fractions after 2 hr with a gradual build-up observed over the 5 hrs shown. The band in the periplasmic fraction (P)

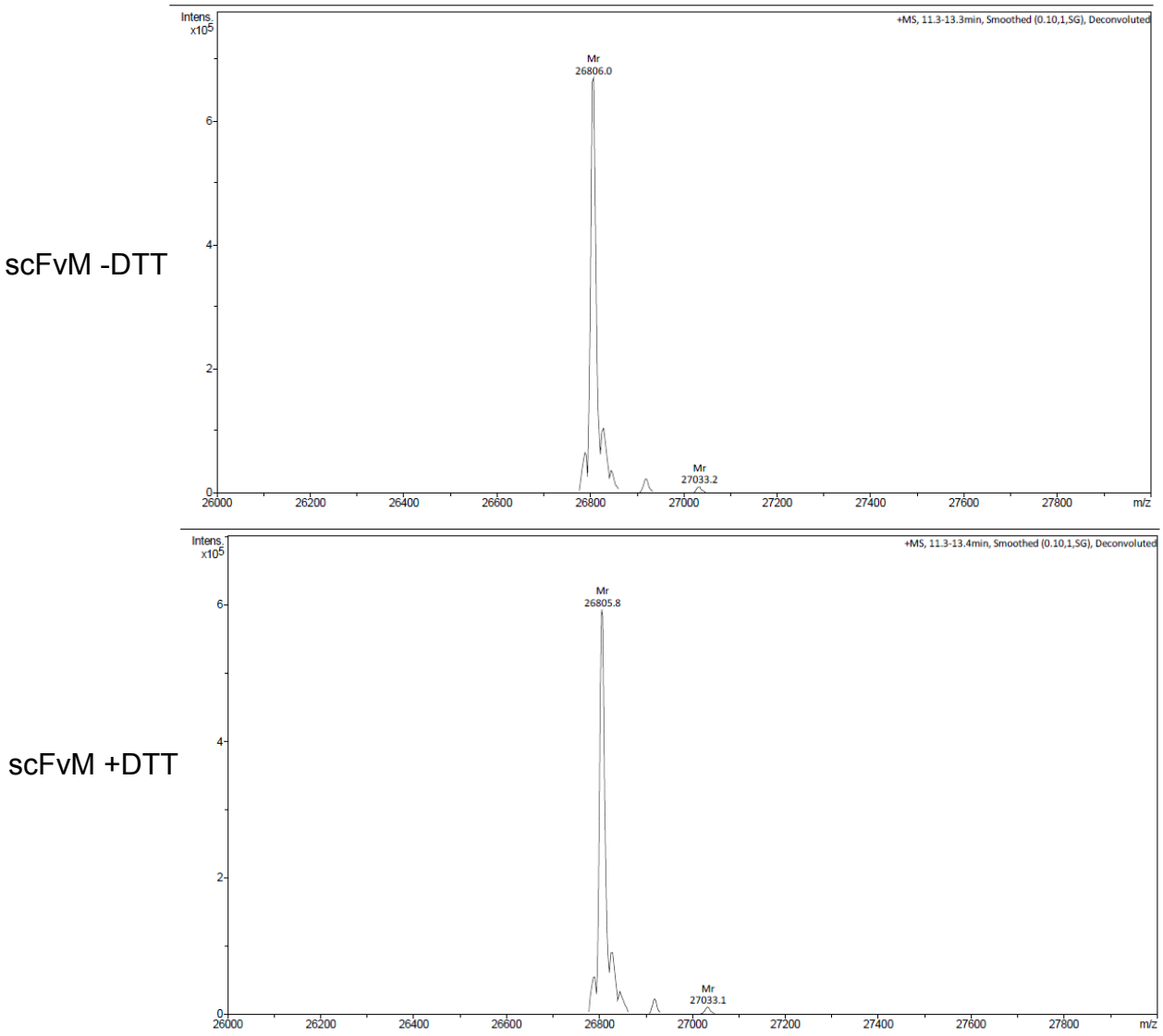


Figure 8: **Mass Spec data verifying reduced state of purified scFvM used for biophysical studies.** DTT, dithiothreitol. Mature scFvM = 26 806 Da.

indicates scFvM is exported by the Tat system in the reduced form. However the strength of this band compared to the cytoplasmic and membrane bands at this time-point indicates export of the reduced form is very inefficient. A precursor band can be seen at a higher molecular weight (32.3 kDa) which corresponds to unprocessed TorA-scFvM. The lower molecular weight bands correspond to the mature size scFvM (26.8 kDa) and the presence of these in the cytoplasm and membrane fractions is likely due to clipping by cytosolic proteases; this phenomenon is often observed when overexpressing Tat-targeted protein in strains with native levels of the Tat machinery.

Figure 9.B shows a parallel time-course analysis of scFvM strains co-expressing the CyDisCo components, here the same build-up of protein is observed over the period shown. However, even from the first time point bands corresponding to mature scFvM are much more prominent indicating a larger quantity of protein present. The build-up of mature protein in the periplasmic fraction (P) is more evident in CyDisCo expressing cells. To confirm the observed band intensities are not just a between-gel artefact, the bands from multiple experiments were subject to densitometry from the same gel. This revealed that export is circa 10x more efficient in CyDisCo expressing cells (see Figure 9.D and Figure 14, discussed more below).

scFvM contains two disulphide bridges, one in each immunoglobulin fold. These are formed between Cys23–Cys97 and Cys162–Cys232 (orange in Figure 6 and circled in the ribbon diagram of Figure 7). In order to further test the necessity of disulphide formation upon export, Cysteine residues were substituted with Serine to break each bridge individually and both bridges at the same time (denoted C97S, C162S and Dual Cys respectively). These were co-expressed alongside CyDisCo as export is more efficient and Figure 9.C shows fractionation of these cells after 3 hrs. Here it can be seen breaking either disulphide has a similar effect to expression in wild-type cells (i.e. without

CyDisCo components) and the dissolution of both disulphides enhances the reduction in export efficiency.

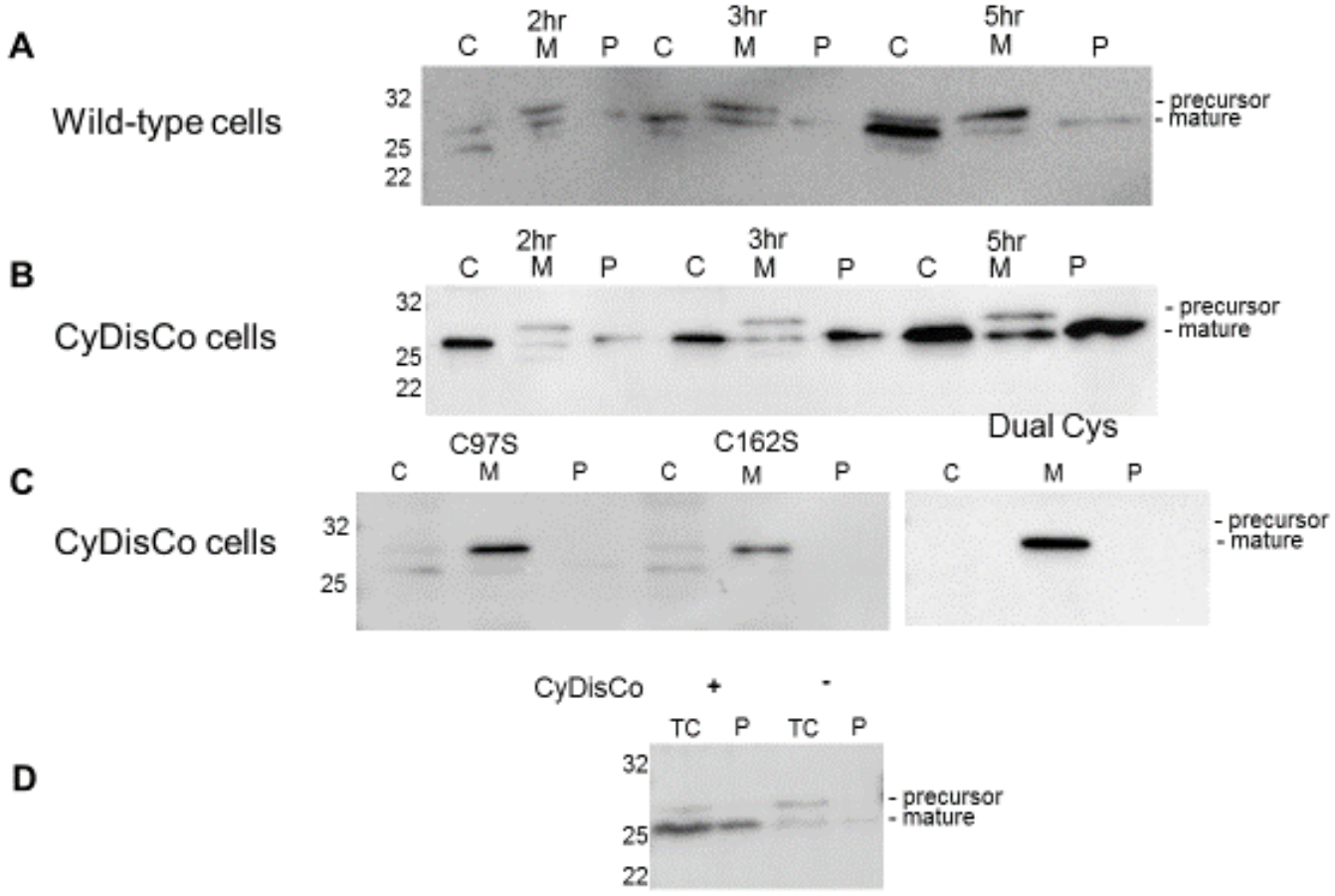


Figure 9: **Expression of TorA-scFvM (unmodified and Cysteine variants) with or without CyDisCo. A & B:** Expression of TorA-scFvM in wild type or CyDisCo cells with fractionation into Cytoplasm (C), Membrane (M) and Periplasm (P) after 2 hrs, 3 hrs or 5 hrs induction. **C:** Expression of cysteine variants that could no longer form one or both disulphide bridges in CyDisCo cells. **D:** Comparison of Total cell (TC) and Periplasm samples for expression of scFvM with (+) or without (-) CyDisCo components. In all cases samples were immunoblotted to the C-terminal 6x histidine-tag. Mobility of precursor and mature forms of scFvM are shown on the right, while molecular weight markers are shown on the left.

Finally, Figure 9.D shows the periplasm from scFvM expressed in CyDisCo or wild-type cells run adjacent to total cell (TC) extract that was not fractionated. This helps to highlight the difference in efficiency of scFvM export in oxidising and reducing cells and also confirms that export efficiency is comparable to that observed as a proportion of the combined (C+M+P) samples.

Altogether these results indicate export of scFvM by the Tat system is far more readily tolerated in CyDisCo expressing cells in which cysteine residues are able to form disulphides, an interesting result considering scFvM is suitably folded after extraction from the cytoplasm of wild-type cells (discussed below) and we propose that formation of the disulphide bonds generates a more rigid substrate structure that is more acceptable to the Tat proofreading mechanism.

The Tat proofreading mechanism is tolerant of large-scale alteration to substrate surface charge

Evidence suggests the Tat system may prefer certain surface characteristics in a substrate, with certain surface features being equated with unfolded protein and thus rejected for export (for example, a large area of hydrophobicity may indicate incomplete folding) (Richter *et al* 2007). To assess this point we altered the surface charge or hydrophobicity while ensuring the structure remained unaffected. These scFvM variants with altered surface residues were then fused with a TorA signal peptide and tested for export via the Tat system. As previously mentioned, the initial surface alterations can be grouped into three groups: 1) 2, 4 or 6 uncharged residues were substituted to form 1, 2 or 3 surface salt-bridges (SB) composed of Lys-Glu pairs and these were denoted 1SB, 2SB and 3SB respectively. 2) 5 uncharged surface residues (L11, Q13, A88, L112 and S116) were

substituted with Lys, Glu or Arg to create a 'charged patch' as high levels of surface charge have been shown to reduce aggregation (Lawrence *et al* 2007) and this could be a factor in improving export. These are denoted 5Lys, 5Glu and 5Arg respectively. 3) Finally, several surface arginines were substituted with Lysine (5R>K) or *vice versa* (7K>R), to test if the identity of surface positively charged side-chains effected export. The substitutions in groups 1 and 2 are located in a region which is uncharged in the wild-type while group 3 substitutions are distributed globally across the surface of scFvM.

Surface images displaying electrostatic potential of group 1 and 2 variants are shown in Figure 10 with positive charge in blue, negative charge in red and uncharged in grey. The top row shows salt-bridge variants with increasing number of salt-bridges from left to right (1SB, 2SB, 3SB), with each successive salt-bridge indicated by arrows. The image on the far-left details the wild-type scFvM with the uncharged region circled. In the bottom panel group 2 variants are shown with the 'charge-patch' circled, the same area circled in the wild-type. Images from left to right are 5Lys (blue), 5Glu (red) and 5Arg (blue), and it can be seen that a significant area of charge has been introduced.

The variety of mutant versions of scFvM analysed in this study have wide ranges of surface charge and this could alter the individual variants' proclivity to be correctly separated by the fractionation protocol used. For this reason control experiments were consistently carried out to ensure the levels of protein in each of the fractions is an accurate representation of the intact cell. A sample of these is shown in Figure 11. Figure 11.A shows a Coomassie-stained gel of the cytoplasm, membrane and periplasmic fractions from a 3 hr induction of scFvM expressed alongside CyDisCo. It is easily evident that the protein profiles of each fraction are distinct.

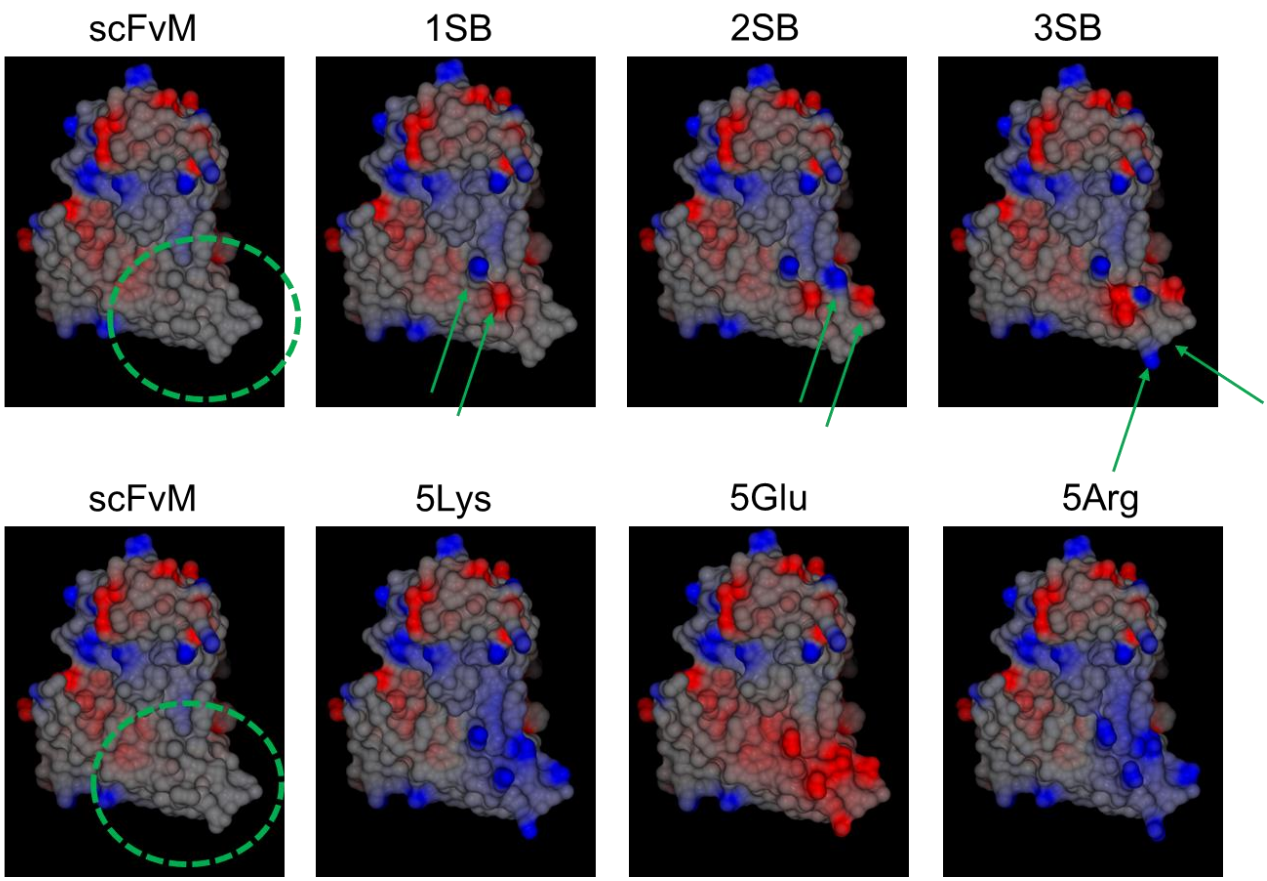


Figure 10: **Structures of scFvM variants with introduced surface salt-bridges (SB) or patch of charge.** Top row, from left: unmodified scFvM with uncharged domain circled in green. Incremental salt-bridges are shown from right to left with each new salt-bridge indicated by green arrows. Salt-bridges are formed from Lys-Glu pairs. Bottom row, from left: unmodified scFvM with uncharged domain circled in green. Patches of 5x Lys, Glu or Arg are shown from left to right. Blue = positively charged at physiological pH. Red = negatively charged at physiological pH. Images were made using CCP4 Molecular Graphics.

Figure 11.B shows a western blot for fractionation of scFvM expressed alongside CyDisCo and a *tat* null strain: a periplasmic band is apparent in the former but completely absent in the latter, demonstrating there is no leakage of cytoplasmic or membrane fractions into the periplasmic fraction. The *tat* null fractionation is also important as it also demonstrates proteolytic processing of TorA-scFvM to the mature size in the cytoplasm. Figure 11.B also shows fractionation of an scFvM mutant termed 26tail (discussed in more detail below) expressed alongside CyDisCo; again the absence of a band in the periplasm indicates no contamination of cytoplasmic proteins.

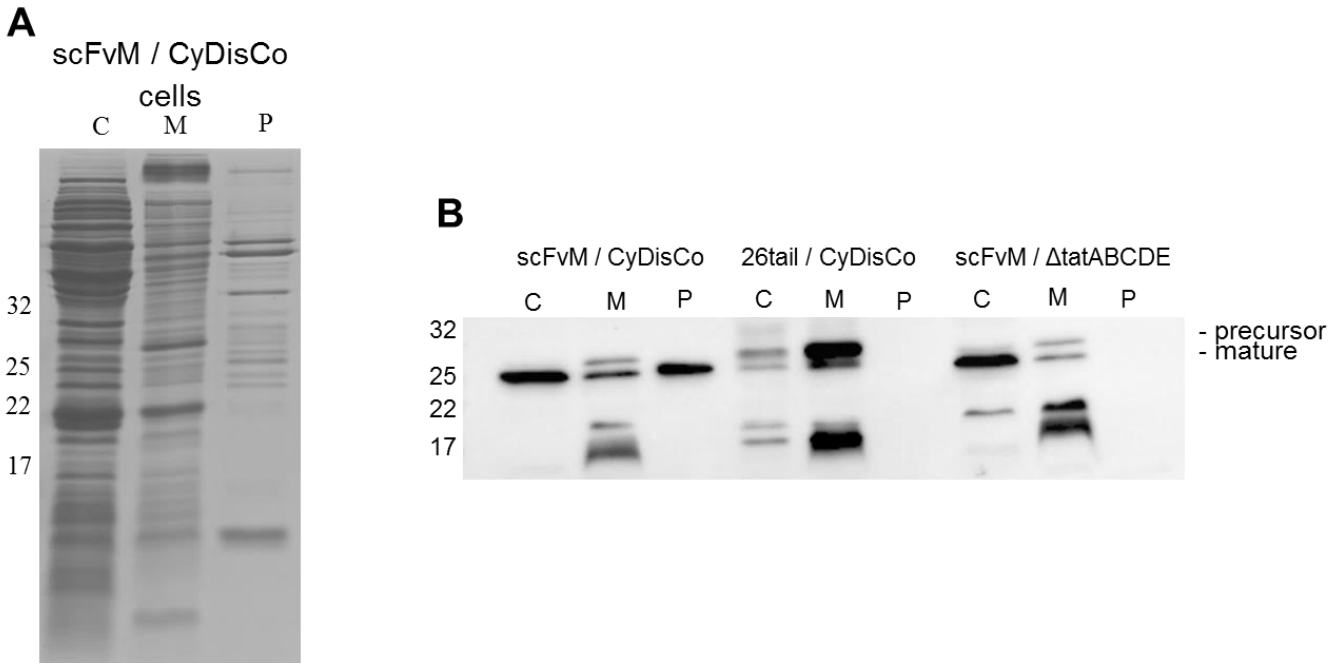


Figure 11: **Fractionation controls for scFvM export assays.** TorA-scFvM was expressed in CyDisCo cells and *ΔtatABCDE* strains. After 3 hr induction, cultures were fractionated into Cytoplasm, Membrane and Periplasm fractions. **A**: Coomassie-stained gel highlighting the distinct protein profile of C/M/P fractions. **B**: Immunoblot for fractionation of cells expressing TorA-scFvM and CyDisCo or in *tat* null strains (left and right lanes respectively). The central lanes show fraction of cells expressing a 26tail variant (Figure 16). Mobility of molecular markers (in kDa) are shown on the left, while precursor and mature forms of scFvM are shown on the right.

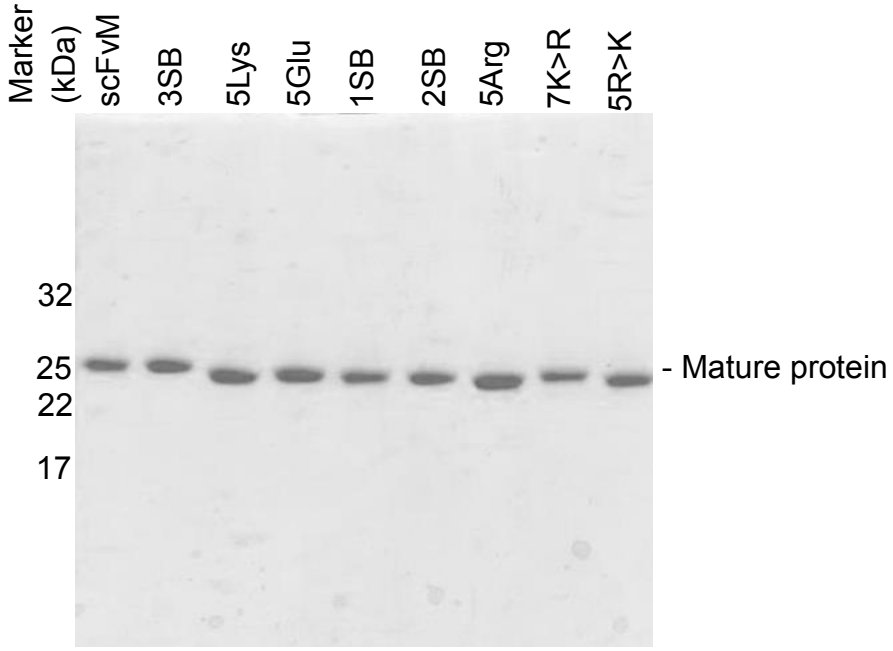


Figure 12: **Coomassie-stained gel of scFvM variants purified by protein A affinity column.** Molecular weight markers are shown on the left (in kDa) with mature protein indicated on the right. Note the lack of contaminating bands in any lane.

Before testing these mutations for the effect on export by the Tat system, we wanted to ensure none had introduced any significant structural changes, thus variants were expressed and purified from the cytoplasm of wild-type cells (*i.e.* cells not expressing CyDisCo) and biophysical characterisation of the variants undertaken. The characteristics of the wild-type and variants are shown in Table 7. Purification from the cytoplasm of wild-type cells and biophysical characterisation was carried out by James Austerberry at the University of Manchester. Biophysical studies were carried out on the reduced protein, indicated by the mass spectrometry data, whereas export assays were conducted in cells co-expressing CyDisCo components because export is more efficient. Nevertheless, the reduced protein is folded, thus it is reasonable to assume disulphide bond formation by the CyDisCo components would only serve to decrease the structural flexibility of the domains, increasing rigidity and will not alter the tertiary structure or surface topology of the variants in any significant manner.

Inspection of Table 7 reveals that each of the variants in groups 1, 2 or 3 share a high degree of structural similarity to the wild-type scFvM deviating by, at most, 0.03 from wild-type values; a negligible amount. Fluorescence peak values deviated by no more than 0.6 nm from the wild-type peak of 334.6 nm and had a hydrodynamic radius with a maximum of 0.2 nm difference to the wild-type of 2.4 nm. Fluorescence peak and hydrodynamic radius give an indication of overall compactness of a protein, providing an indication of globular density and flexibility. In order to conduct these biophysical experiments on scFvM and associated variants, protein first needed to be purified. After expression in the cytoplasm, protein was purified via Protein A affinity chromatography. Figure 12 shows a Coomassie stained SDS-PAGE for samples of the elutates from the Protein A column. All variants have a band similar to that of the wild-type. This is significant as Protein A affinity requires an intact V_H domain-fold, confirming all variants have an intact Protein A binding site and a structural character similar to the wild-type. The combination of biophysical data and ability to bind a Protein A column of the variants presented here serve to confirm the surface variations have minimal effect on overall tertiary structure of scFvM and substantiate continuation of this study.

As with the wild-type, all variants were cloned into vector pYU49 (TorA signal peptide and co-expression of CyDisCo) and plasmids were induced with 0.5 mM IPTG for 3 hrs before fractionation into cytoplasm, membrane and periplasm. Figure 13 shows the results of these fractionations after western blotting to the C-terminal 6xHis-tag. Figure 13.A shows introduction of 1, 2 or 3 Lys-Glu salt-bridges to the surface results in an export profile with little difference to the wild-type – a band at mature size (26.2 kDa) is prominent in the periplasm (P) with mature (32.3 kDa) and precursor bands present in the cytoplasm and membrane. With a cursory glance it could be argued that 2SB and 3SB

show slightly improved export over the wild-type, however after comparison of several repeat experiments

Mutant	pI	Secondary Structural Element						Fluoresence Peak [nm]	Hydrodynamic Radius [nm]
		α_R	α_D	β_R	β_D	T	U		
Wild type	**7.8**	0	0	0.34	0.20	0.12	0.34	334.6 ± 0.1	2.41 ± 0.01
5x Lys	**9**	0	0	0.35	0.20	0.13	0.33	334.6 ± 0.1	2.39 ± 0.01
5x Glu	**5.3**	0	0	0.35	0.20	0.13	0.33	334.6 ± 0.1	2.60 ± 0.02
5x Arg	**9**	0	0	0.36	0.20	0.11	0.33	334.6 ± 0.1	2.52 ± 0.04
3x salt-bridge	**7.8**	0	0	0.35	0.20	0.12	0.33	334.1 ± 0.1	2.40 ± 0.01
2x salt-bridge	**7.8**	0	0	0.34	0.20	0.13	0.34	334.6 ± 0.1	2.39 ± 0.01
1x salt bridge	**7.8**	0	0	0.36	0.20	0.11	0.34	334.9 ± 0.2	2.40 ± 0.01
Global K→R	**7.8**	0	0	0.35	0.21	0.09	0.35	334.5 ± 0.1	2.37 ± 0.01
Global R→K	**7.8**	0	0	0.35	0.19	0.11	0.35	335.15 ±0.1	2.43 ± 0.01

Table 7: **Structural characteristics of scFvM variants characterised by circular dichroism, tryptophan fluorescence and dynamic light scattering.** Secondary structural elements were obtained through fits of the spectra between 190 – 260 nm by CDNN secondary structural analysis programme. Samples are at a concentration of 1 mg/mL as determined by UV absorbance at 280 nm, in 25 mM sodium acetate buffer, pH 5. αR = regular alpha helix, αD = disordered alpha helix, βR = regular beta-sheet, βD = disordered beta sheet, T = beta turns, U = unordered. Secondary structural fractions are rounded. Work contributed by James Austerberry at University of Manchester.

it is evident that all three salt-bridge variants are exported to a degree broadly similar to the unmodified scFv. Figure 13.B shows fractionation for variants 5Lys, 5Glu and 5Arg. Bands appear to be weaker overall, however further tests show that 5Lys and 5Glu are still exported with moderate efficiency (export efficiency is discussed in more detail below). The introduction of an 'arginine patch' in the 5Arg variant however, greatly reduces export efficiency, evident here by a comparatively weak band in the periplasm (P).

Figure 13.A and B also show a control export of the unmodified scFvM run against total cells (TC) to show the combined (C+M+P) samples are representative of a total cell. These panels also show scFvM total cell sample before induction (U), where the absence

of band confirms expression of scFvM from this plasmid is tightly regulated by the presence of IPTG. Figure 13.C shows similar controls for scFvM group 1 and 2 variants with total cell samples run adjacent to periplasmic samples. Not only are the export efficiencies comparable to those in panels A and B, but Figure 13.C also serves to highlight the reduced export efficiency of the 5Arg variant.

Due to the poor export efficiency observed for the 5Arg variant, we tested whether the Lys:Arg ratio and distribution of positively charged side-chains affected export by substituting these residues in other places across the scFvM surface. To this end 5 arginine residues were substituted with lysine in 5R>K (note these are different residues to those in variant 5Arg) and 7 lysine residues were substituted with arginine in 7K>R. These substitutions account for all of the surface exposed Arg/Lys residues. Export assays for these two variants are shown in Figure 13.D, revealing that 5R>K and 7K>R are both exported with reasonable efficiency and thus it is likely the reduced export observed in 5Arg is not due to an 'aversion' to arginine on the part of the Tat machinery (discussed further below).

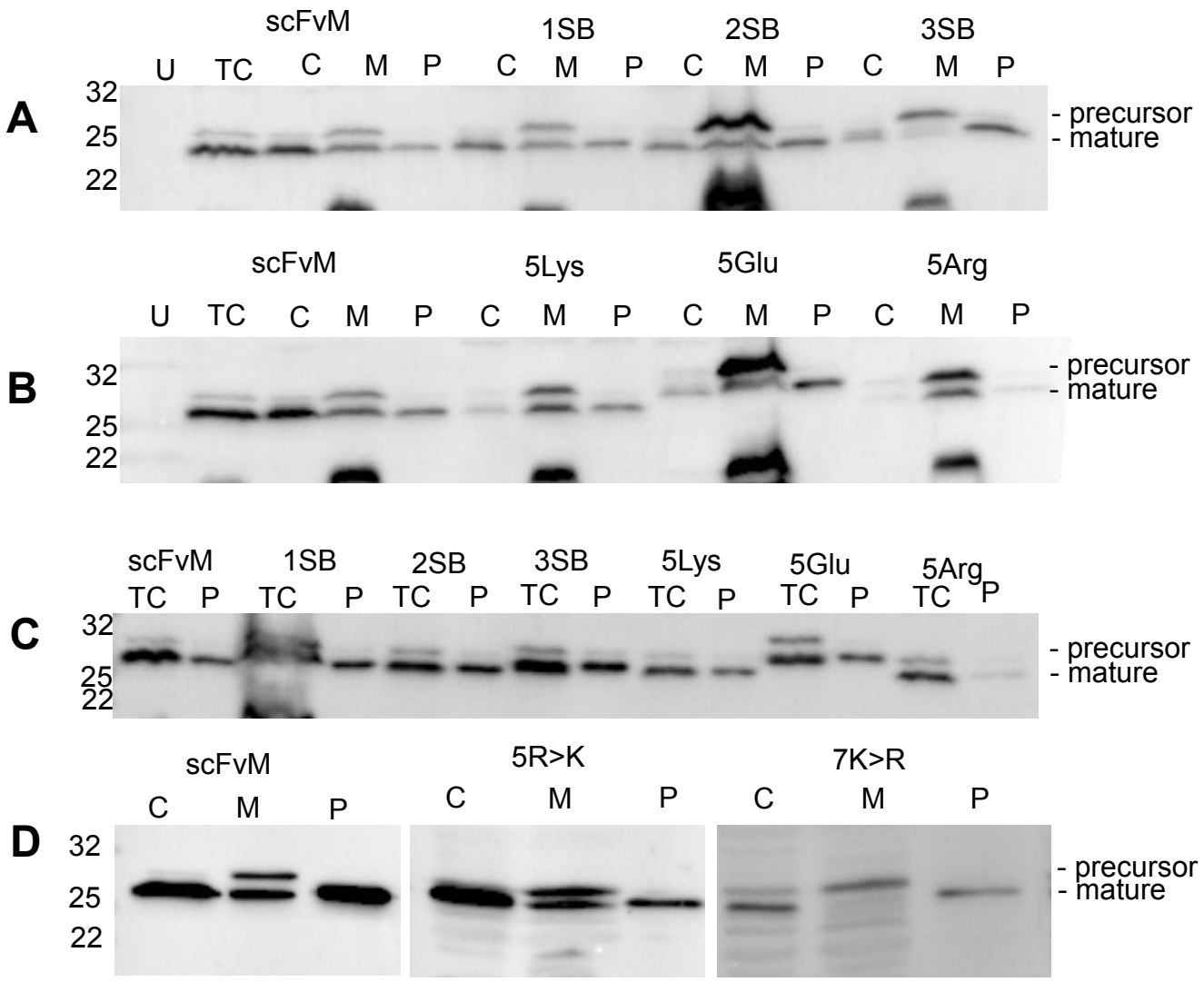


Figure 13: **Export assays for salt-bridge and charged patch scFvM variants. A**: Export assay for unmodified scFvM and salt-bridge variants where 2, 4 or 6 surface residues were substituted to form 1, 2 or 3 salt-bridges (1SB, 2SB, 3SB) formed of Lys-Glu pairs 3 hrs after induction. Samples were generated as in Figure 9 and were run against Total Cell (TC) and an equivalent number of cells prior to induction (U). **B**: Export assay for the charge patch variants where 5 residues were substituted to introduce 5x Lys, Glu or Arg residues. Shown in the same format as in A. **C**: Total cell and Periplasmic samples from unmodified and all variants were run on the same gel for comparison. **D**: Export assays for unmodified scFvM and variants in which 5 surface arginines were substituted with lysine or 7 surface lysines were substituted with arginine. Molecular weight markers are shown on the left (in kDa) while precursor and mature forms of scFvM are indicated on the right.

To more accurately compare the export efficiency of the individual scFvM variants, the bands on the blots for multiple repeat experiments were subject to densitometry and compared to the export efficiency of the wild-type. Figure 14 shows the average export efficiencies calculated after 4 separate experiments, relative to the wild-type. Here it becomes apparent that introduction of up to 3 Lys-Glu salt-bridges has minimal effect on export efficiency, while a patch of 5Lys or 5Glu residues is moderately tolerated. However, creation of a 5Arg patch greatly reduces export efficiency. Counter to this, exchanging surface lysines with arginines or *vice versa* does not drastically affect export, although exchange of lysines in the latter (7K>R) is less well tolerated.

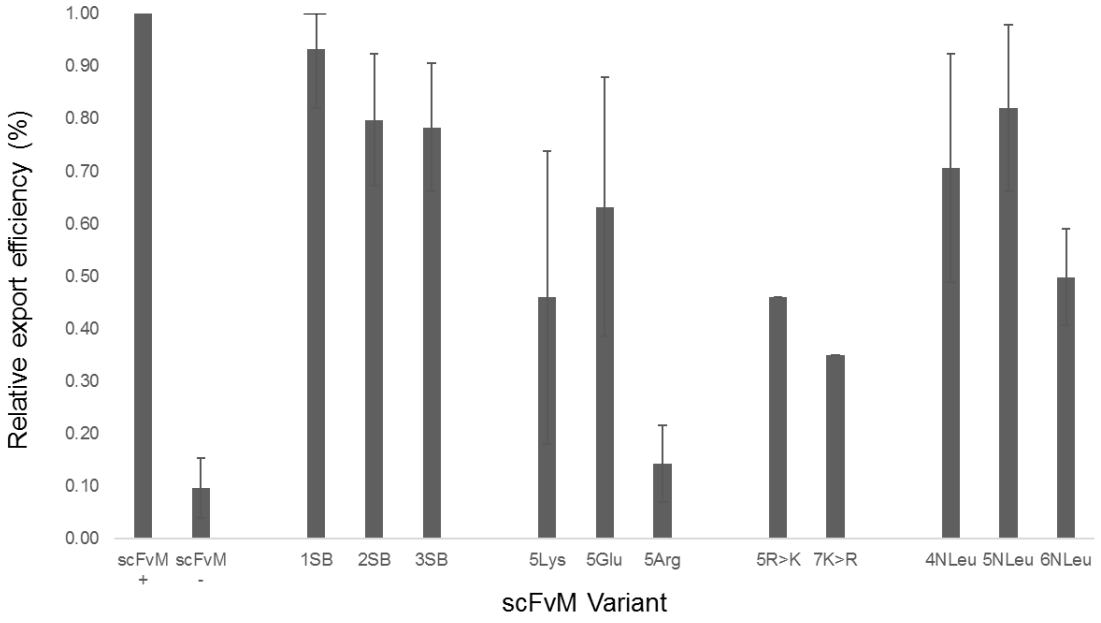


Figure 14: **Export efficiencies calculated for the scFvM variants used in this study.** After each export assay, the signal strength of each periplasmic band was calculated as a percentage of the total strength of bands for that variant (C+M+P). Shown is the average from 4 separate export assays per variant with export efficiency relative to that of the unmodified variant. scFvM +/- = TorA-scFvM unmodified expressed with or without CyDisCo components respectively.

The Tat system can tolerate significant surface hydrophobicity in substrates

Previous studies have suggested that hydrophobicity in a substrate could be a key factor in how the Tat machinery assesses the folding state (Richter *et al* 2007), with high levels of hydrophobicity equated with an unfolded protein as these are usually sequestered within the interior. In continuation of our alteration to the surface features, we created another group of variants where 4, 5 or 6 hydrophilic surface residues were substituted with leucine. These variants are denoted 4NLeu, 5NLeu and 6NLeu respectively, with substitutions beginning in the N-terminal domain. The top panel of Figure 15 shows hydrophobicity profiles, centred on the N-terminal domain of group 4 variants compared to the wild-type (left). The surface of the unmodified scFvM is predicted to be relatively hydrophilic with hydrophobicity (shown in yellow, hydrophilic residues are purple) contributed by alanines, two valines and a leucine, therefore introducing multiple leucine residues to the surface will likely significantly increase the hydrophobicity of this region. Unmodified scFvM is shown on the left with pre-existing hydrophobic residues labelled. 4NLeu, 5NLeu and 6NLeu are then shown with additional Leu residues arrowed in each case. It should be noted that structural characterisation of this group of variants was not possible as they could not be prepared in sufficient quantities.

Fractionation of cells expressing these variants after 3 hrs is shown in the bottom panel of Figure 15. Interestingly, despite the increase in surface hydrophobicity the periplasmic band (P) for 4NLeu is fairly strong, indicating export still takes place and the changes to the surface of this variants is tolerated by the Tat proofreading system. The periplasmic band for 5NLeu and 6NLeu appears to be weaker, possibly indicating less efficient export. However, as with the previous surface alterations, an average for export efficiency of these variants can be seen in the final three bars in Figure 15, this highlights that creation

of a significantly hydrophobic patch is largely tolerated by the Tat proofreading system and only has a moderate effect on export efficiency.

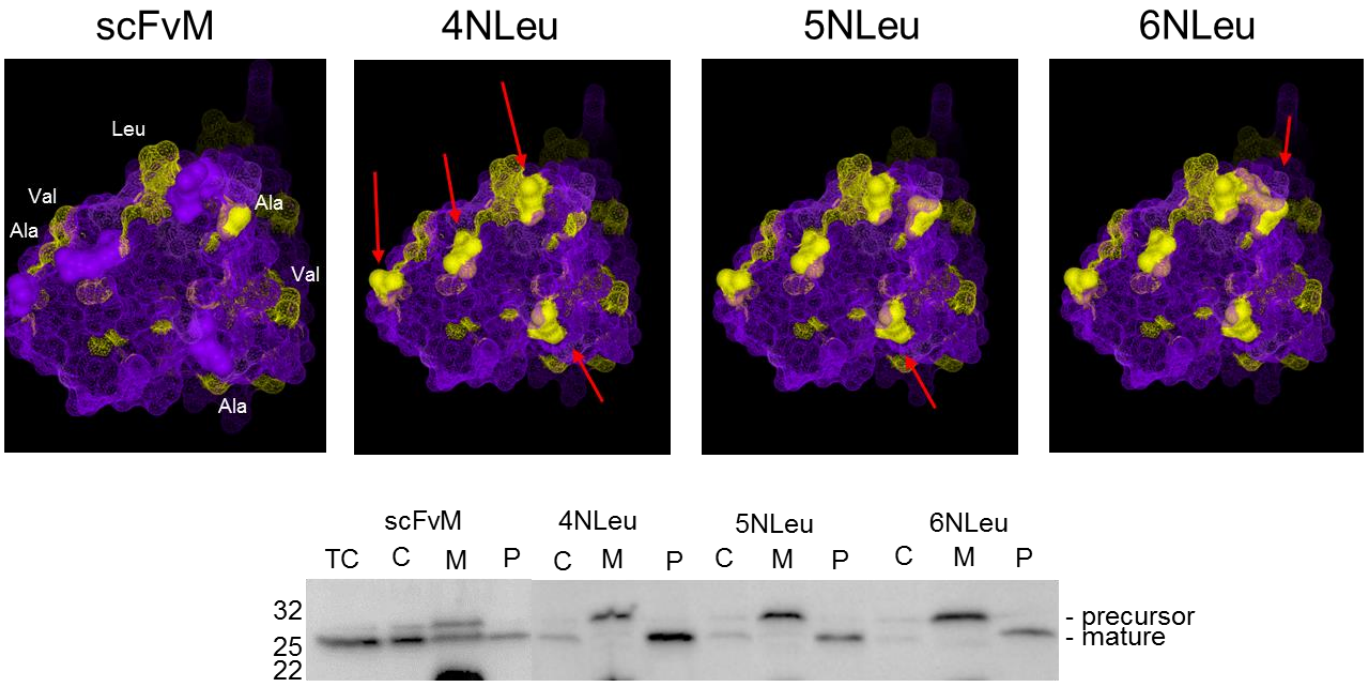


Figure 15: **The Tat system tolerates significant surface hydrophobicity in scFvM.** Top: surface hydrophobicity profiles for the N-terminal of scFvM. The unmodified variant is shown on the left with the progressive introduction of 4, 5 or 6 leucine residues to the right. Introduced leucine residues are indicated by red arrows and shown in solid yellow. Purple = non-hydrophobic, yellow = hydrophobic. Bottom: Export assay of the hydrophobic variants compared to the unmodified scFvM. Molecular weight markers are shown on the left (in kDa) with precursor and mature scFvM indicated on the right.

Export of scFvM by the Tat system is blocked by the addition of an unstructured 'tail'

This study has currently shown that the Tat proofreading mechanism is tolerant of changes to the surface charge of a heterologous substrate targeted for export via the Tat machinery. As a final part to this study we wanted to see if Tat proofreading could recognise an unfolded domain on a largely folded substrate. To this end we added 26 residues, chosen at random, to the C-terminus of scFvM, creating scFvM-26tail (SNAIIIIITNKDProNSSSVDKLAAALE). Secondary structure modelling using the Phyre2 programme (Kelley *et al* 2015) indicates the core scFvM is unaffected by the addition of this tail while the tail itself is largely disordered (supplementary data Figure 1). Export assays carried out over a 120 min time-course for scFvM-26tail is shown in Figure 16. It is clearly evident in this figure that this variant is not exported via the Tat system at any point throughout the 120 min induction period, showing Tat unambiguously identifies the 26-residue 'tail' as unfolded and rejects the protein for export.

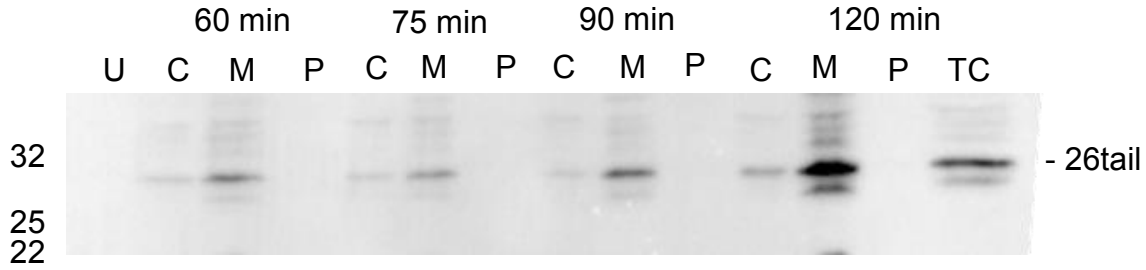


Figure 16: **Addition of a 26-residue disordered tail results in a complete block in export of scFvM by the Tat system.** Immunoblotting to fractionation samples (as in Figure 9) from cultures expressing the '26tail' scFvM variant bearing a 26-residue C-terminal extension. Equivalent number of cells were fractionated into C/M/P at 60, 75, 90 and 120 minutes post-induction, with a total cell sample taken pre-induction (U) and at the 120 min time-point (TC). Molecular weight markers are shown on the left (in kDa).

Discussion

The study conducted in this chapter sought to more accurately identify the factors that dictate whether a substrate is transported or rejected by the Tat machinery. The predominant method by which this was addressed was by identification of any alteration to surface features were accepted or rejected by the Tat system. Secondly, tolerance for substrate conformational flexibility was also touched upon. Although the proofreading or transport mechanism is not understood in any detail at this stage, we assumed the Tat system may associate surface features, such as large areas of hydrophobicity, with an unfolded and therefore transport-incompetent state. We therefore began by manipulating the surface-exposed residues of a structurally characterised scFv in an attempt to induce rejection by the Tat machinery. Alongside this, due to recent evidence indicating the Tat machinery tolerates substrates that possess greater structural flexibility (Alanen *et al* 2015, Stolle *et al* 2016), we also targeted scFvM to the Tat apparatus in the reduced and oxidised state as well as a version with a disordered C-terminal 'tail'.

A range of scFvs have been efficiently exported via the Tat system in both oxidised (DeLisa *et al* 2003) and reduced states (Alanen *et al* 2015), we have found the export of scFvM to be almost totally dependent on the ability to be in the oxidised state (i.e. in the presence of CyDisCo or with bridge-forming Cys residues present). This indicates that the protein is more readily accepted by the Tat proofreading mechanism if intra-domain disulphides are formed. This is an interesting finding as scFvM is suitably folded (see biophysical data, Table 7, and protein A binding, Figure 12) and active (Edwardraja *et al* 2010) when obtained from the cytoplasm in a reduced state. We interpret this to mean that formation of the disulphide bonds 'locks' the scFvM domains in a more rigid

conformation while without these anchors (i.e. in the reduced state), scFvM is in a more dynamic state, possibly reminiscent of the molten globule and is thus deemed insufficiently folded by the Tat system, akin to for example TMAO reductase lacking a MGD cofactor. Additionally, reduced cysteine residues (SH HS) are more hydrophobic than their cysteine (S-S) counterparts (Nagano *et al* 1999), suggesting loss of these covalent linkages in scFvM results in increased flexibility, rather than destabilisation of the hydrophobic cores of the two domains. The entropic contribution of disulphide bridges to protein stabilisation has been well studied, and used as the basis for engineering stability (Zavodszky *et al* 2001). Cysteine and serine side-chains have very similar electrochemical properties, with the latter often used to test the effect disulphide bridges contribute to a protein. Therefore, the 10% export we observe in WT cells may well represent those molecules that happen to be in their most compact conformation when docking at the TatBC substrate-binding complex. Moreover, exchanging cysteine residues with serine, inhibiting bridge formation even in CyDisCo strains, produces a similar reduction on export efficiency. These data suggest that the folded state of a substrate is assessed by the Tat system with an ability to identify substrates in a highly dynamic state as opposed to an 'unfolded' or mis-folded protein. Indeed, scFvM was specifically chosen and engineered for stable expression in the reducing environment of the *E. coli* cytoplasm (Edwardraja *et al* 2010, Austerberry *et al* 2017).

The initial scFvM variants created in this study were designed to assess the capacity of Tat proofreading mechanism to sense surface charge of a substrate. In all but two of these variants substitutions were made in the same, specific area, creating a patch with a variety of different properties. In the first group of variants, up to three charge-neutral pairs (Lys-

Glu) were introduced in the place of non-polar residues, however these variants are still exported with an efficiency comparable to the wild-type.

The second group introduced a significant increase in either positive (Lys or Arg) or negative (Glu) charge in this same area through substitution of 5 uncharged residues by Lys, Glu or Arg. In the case of the Lys and Glu substitutions only a moderate decrease in export efficiency was observed (by ~30%). Evidently the Tat system can tolerate significant increases in the surface charge of a substrate, even when localised to a focused area. Interestingly, introduction of 5 Arg residues had a detrimental effect on export (by ~80%) when localised to the same area, however replacement of 7 surface lysine residues with arginine only produced a moderate decrease in export efficiency, similar to that observed in 5Lys, 5Glu and 5R>K variants, suggesting it is not the specific side-chain for which Tat is intolerant but possibly the concentration of charge.

The final group of variants created in this study were designed to give some idea of the level of hydrophobicity tolerated by the Tat system. Exposure of significant regions of hydrophobicity are hallmarks of unfolded or aggregation-prone proteins and we speculated that introduction of such a region may severely inhibit, or block export. Moreover, it was previously suggested that exposed hydrophobic domains blocked export for small, unstructured proteins (Richter *et al* 2007). In light of this it is interesting that creation of a substantial hydrophobic patch by substitution of uncharged residues with 4, 5 or 6 leucine residues, has no significant effect on export efficiency. Modelling of these variants shows a significant patch of hydrophobicity is introduced at the N-terminal domain of scFvM. Given these data we conclude that a certain threshold of hydrophobic

character in a substrate is needed before rejection occurs, or that it is sensed in conjunction with other characteristics of Tat-targeted substrates, such as conformational flexibility.

Finally, an unfolded domain was added to a folded substrate in the form of a disordered 26-residue 'tail'. In contrast to the other variants used to test the Tat proofreading mechanism in this study, addition of this unfolded 'tail' completely blocks export. This is in-line with the data obtained for the formation of intra-domain disulphide bridges, which have a stimulatory effect on export efficiency. These data represent various levels of an unfolded substrate, with a completely disordered domain (26tail variant) and a more dynamic, transiently-unfolded domain (scFvM lacking disulphide bridges) in an otherwise globular and relatively stable protein.

For the structurally dynamic variants, scFvM 26tail and variants lacking disulphide bridges, the protein is not confined to a specific structure and is able to 'breathe'. It is beyond the scope of this study to more accurately identify the structural flexibility of scFvM and the variants generated, and further work could possibly do this in a more detailed manner, using a variety of substrates and utilising more sensitive techniques, for example NMR, to link conformational flexibility with translocation competence.

Testing the proofreading capacity of the Tat system with a wholly synthetic, heme-binding BT6 Maquette.

In the previous chapter we have shown the Tat proofreading system is tolerant of significant alterations to the surface charge and increased hydrophobicity of a Tat-targeted scFv. Conversely Tat proofreading has little tolerance for an scFv that lacks structural rigidity or contains a disordered C-terminal domain. Evidence abounds for the Tat systems' ability to transport heterologous proteins such as GFP (Thomas *et al* 2001), scFvs (Chapter 3, Matos *et al* 2014, Alanen *et al* 2015), interferons, V_H domains (Alanen *et al* 2015) and F_{AB} fragments (DeLisa *et al* 2003). Here we set out to test the export and proofreading capacity of the Tat system with a wholly artificial protein maquette, termed BT6.

Maquettes were born of the desire to create proteins fit for purpose outside those found in nature. Due to the 'evolutionary baggage' accumulated in the amino acid sequence of natural enzymes, it is difficult to assign function to any one amino acid and derive any one function from a motif and thus achieve the desired outcome when importing a natural motif from an enzyme into a man-made construct (DeGrado *et al* 1989, Edelman and Gally 2001). More recently a new approach was undertaken that aimed to minimise complexity in the amino acid sequence while increasing engineering freedom. This was a bottom-up approach that generated heptad repeats to create soluble 4 α-helical bundles. Within this framework the functionality of each amino acid is largely understood due to the small and simple nature of the maquette. Cofactor ligation sites were then inserted into this framework without mimicking natural enzyme motifs to create a protein with oxidoreductase capabilities e.g. electron transfer proteins (Shifman *et al* 1998). These first 4 α-helical maquettes were helix-loop-helix homodimers tethered by a disulphide bridge. Due to their sequence symmetry, they are only able to bind single cofactors and perform relatively simple functions, such as proton coupling, electrochemical charge coupling and ligand exchange (Grosset *et al* 2001). This was overcome by removing the

disulphide tether and fusing the N-terminus of one unit to the C-terminal of another, generating a helix-loop-helix-loop-helix-loop-helix design (Lichtenstein *et al* 2012). The maquette used in this study, BT6, is of this single-chain design, developed by Farid *et al* 2013. The sequence of BT6 is shown in Figure 17.A while a ribbon diagram and the heme *b* cofactor are shown in Figure 17.B. The BT6 four-helix bundle forms a water-excluding cavity where histidine residues form ligation sites for two heme *b* cofactors, producing an artificial protein resembling a *b*-type cytochrome. To be accepted by the Tat proofreading mechanism, BT6 needs to fold and acquire its heme cofactors in the cytoplasm prior to interaction with the Tat translocase.

A

MNNNDLFQAS RRRFLAQLGG LTVAGMLGPS LLTPRRATA AQAA
GGDGENL YFQG EIWKQHE DALQKFE EALNQFE DLKQL **GGSGSGSGG**
EIWKQHE DALQKFE EALNQFE DLKQL **GGSGSGSGG**
EIWKQHE DALQKFE EALNQFE DLKQL **GGSGSGSGG**
EIWKQHE DALQKFE EALNQFE DLKQL HHHHHH

B

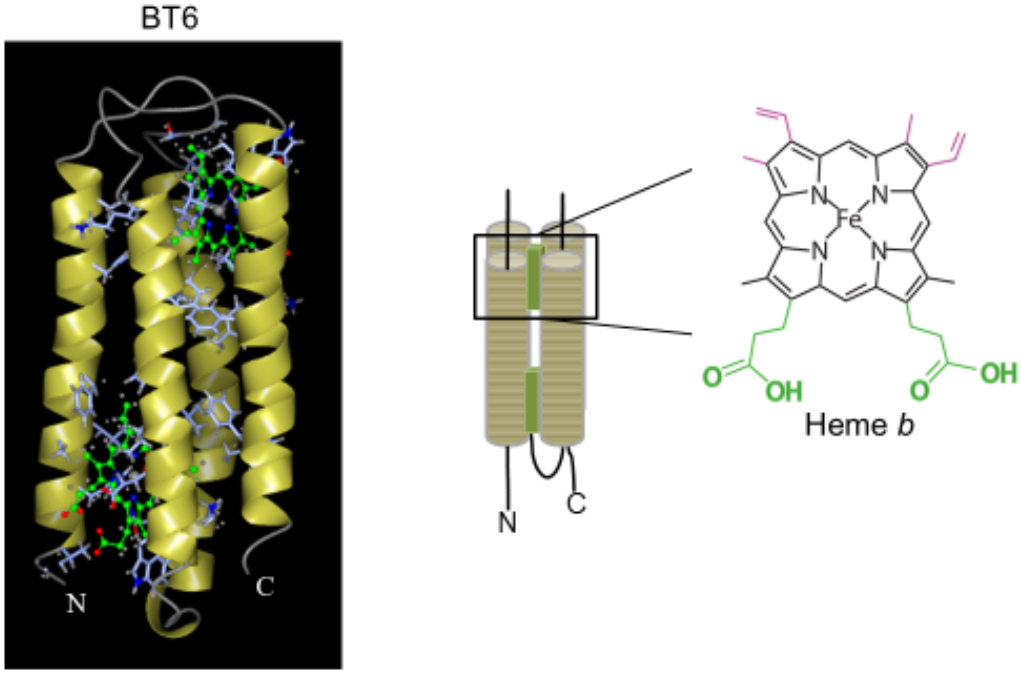


Figure 17: **Primary amino acid sequence and structural features of maquette BT6**. **A**: The signal peptide from TorA was fused to the N-terminal of BT6 (yellow) along with the first four amino acids of the mature TorA protein (green). The Gly/Ser linker between heptad repeats is in bold and underlined with heme *b*-binding histidine residues in orange and underlined. The C-terminal 6x histidine-tag is shown in brown. **B**: Left: Ribbon diagram of the holoprotein BT6 with two bound heme *b* molecules. α-helix = gold, β-sheet = red, unstructured = grey and heme *b* = light green. Right: Cartoon diagram of BT6 bound to its co-factor with an expanded section showing the heme *b* molecule. Images generated with CCP4 Molecular Graphics.

Results

High-level overexpression of BT6 results in contamination of Periplasm with cytoplasmic protein

Here we set out to test the ability of the *Escherichia coli* TatABC system to export a wholly artificial protein. This represents a challenge for the Tat system as, not only is BT6 a recombinant protein and thus not usually found in *E. coli*, it is wholly artificial protein and therefore not found in nature. The protein sequence shares no identity with natural proteins and avoids natural motifs. Given the expansion of synthetic biology and the growth of the ‘’protein toolbox’’ in recent years, it would be of interest to know the capabilities of the Tat system with regards to entirely man-made proteins.

The sequence for the signal peptide from *E. coli* TorA (trimethylamine-N-oxide reductase) along with the first 4 amino acids of the mature protein (highlighted in Figure 17.A), were fused to the N-terminus of a heme *b*-binding maquette (denoted RRBT6h2), and expressed from plasmid pJExpress414, a high-copy vector with a T7 inducible promoter. In parallel to this, mutations to the TorA signal peptide that are commonly used to indicate protein is processed solely by the Tat system (and not by the Sec system) were introduced. These are as follows: mutation of the first Arginine in the ‘RR’ motif to Lysine (denoted KRBT6) which commonly has little effect on translocation in *E. coli*, and mutation of both Arginines to Lysine (denoted KKBT6) which leads to a block in export (Halbig *et al* 1999).

Anti-His western blots to the C-terminal 6x His-tag of cultures expressing these plasmids and fractionated as described in section 2.3 are shown in Figure 18.A. Here it can be seen that induction with 0.5 mM IPTG results in protein being present in all three fractions (Cytoplasm, Membrane and Periplasm, C/M/P respectively) at the mature size with the

wild-type ‘RR’ signal peptide. Mutation to ‘KR’ results in a decreased amount of export, indicated by a less-intense band in the periplasm (‘P’ lane) at mature size, additionally there is no mature size protein present in the periplasm for the non-exportable ‘KK’ BT6h2.

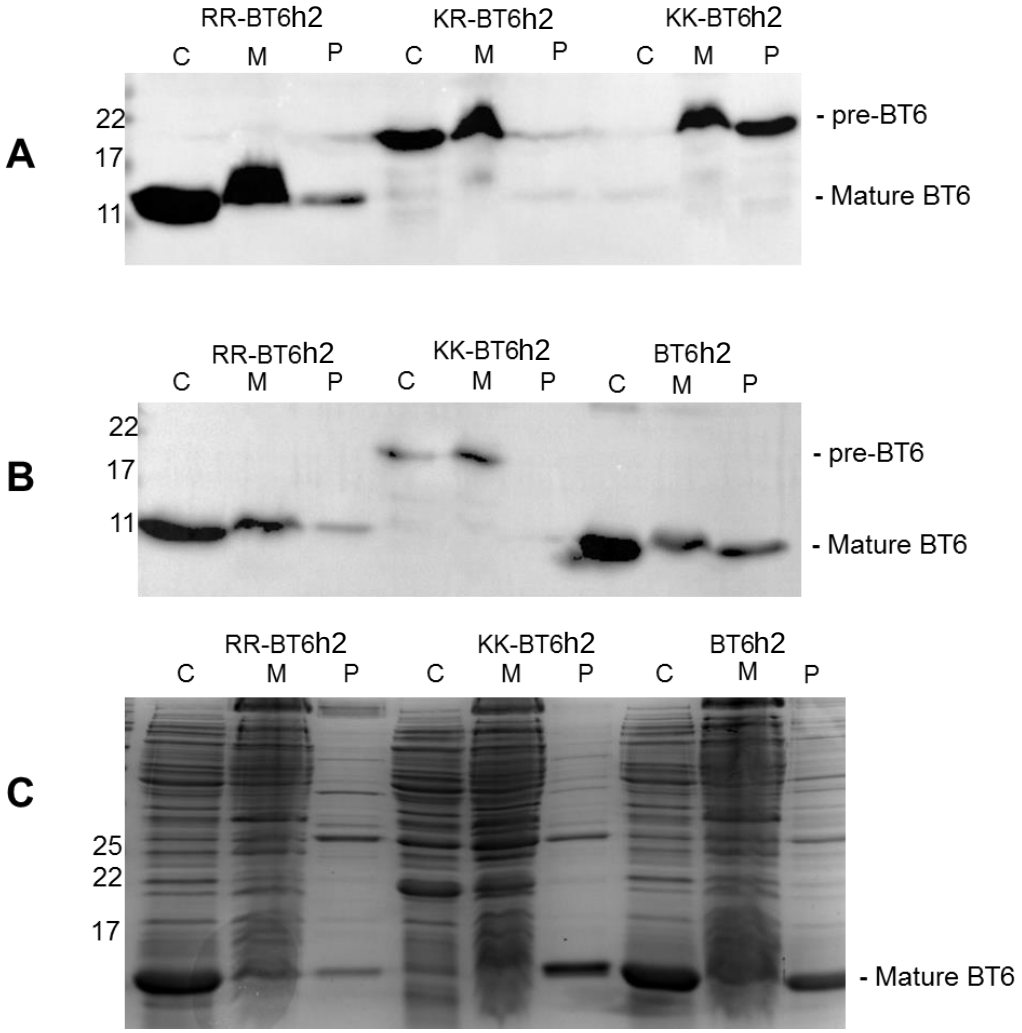


Figure 18: **Overexpression of BT6 from the pJexpress414 plasmid causes contamination of the Periplasmic fraction. A**: Immunoblot to the C-terminal 6x histidine-tag for C/M/P samples for expression of BT6h2 from the pJexpress414 plasmid. Functional TorA (RR) and a version blocked in transport (KK) are shown. **B**: Functional TorA signal peptide (RR) is shown alongside the two controls for non-functional signal peptide (KK) and BT6h2 lacking a signal peptide. Again precursor and mature bands are observed in all three (C/M/P) fractions. **C**: Coomassie-stained gel of the immunoblot shown in B. Bands corresponding to BT6h2 are easily the most abundant in each fraction. Molecular weight markers are shown on the left (in kDa) and precursor and mature BT6 indicated on the right. Precursor – 21.8 kDa, mature – 17.11 kDa.

However, there are periplasmic bands at precursor size for both KR and KK BT6h2 products which indicates expression at this level causes leakage across the inner membrane of BT6. Indeed at mature size these are not processed by leader peptidase and thus likely do not cross the membrane via Sec/Tat.

To further test whether expression level was the cause of contamination of the periplasmic fraction or if the Tat system simply could not recognise the synthetic BT6, the maquette was cloned into pJExpress414 without the signal peptide and expressed under the same conditions. In Figure 18.B a similar scenario to that in A is observed, namely both RRBT6h2 and KKBT6h2 are evident at mature size in the periplasm. Additionally, the final three lanes show BT6h2 with no signal peptide at all is found in all three fractions. The final panel (Figure 18.C) shows a Coomassie-stained gel of the blot in B. The different protein profiles of the C/M/P lanes indicate an efficient fractionation and no contamination across the fractions, indicating the mature BT6 present in the periplasmic fraction is due to some other factor, likely the very high expression level of protein from pJexpress414.

Low-level expression of BT6 results in correct processing by the Tat system

IPTG-inducible promoters are not always very tunable and thus we sought a different method to decrease the expression level of our BT6 constructs. We decided to clone the various constructs into the low copy-number plasmid pEXT22 which on average has 1 – 1.5 copies per cell (Dykxhoorn *et al* 1996) and also places the various BT6 constructs under the lower expression pTac promoter. Figure 19 shows fractionation of cells expressing RRBT6h2 and various controls from this vector. In Figure 19.A BT6h2 fused to a functional TorA signal peptide, is exported via the Tat system to the periplasm and

processed to a mature size as evidenced by a band comigrating with the 17 kDa marker (16 981 Da). Successive lanes show RRBT6h2 is processed solely by the Tat system as KKBT6h2 is completely blocked for export, as is expression of the mature-size protein only (i.e. lacking the TorA signal peptide). The anti-His blot in Figure 19.B shows efficient export can be achieved for the KRBT6h2 construct whereas KKBT6h2 is blocked for export. BT6h2 lacking a signal peptide is also not exported to the periplasm. A lack of signal in the cytoplasm and membrane fractions we take to indicate almost all protein is exported due to the low levels of protein produced in this vector. A lack of signal in the non-exported constructs is likely due to degradation of non-exported Tat substrates as has been identified previously (Lindenstrauß *et al* 2010).

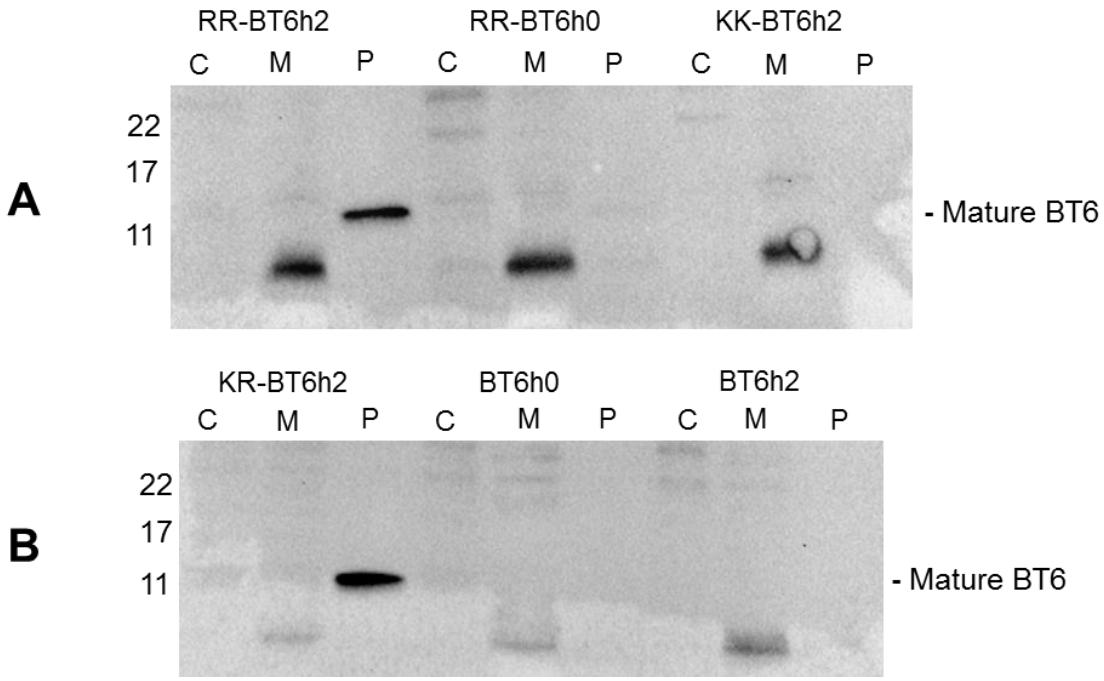


Figure 19: **Low-level expression of BT6h2 from the pEXT22 plasmid results in correct processing by the Tat system. A**: C/M/P samples for BT6h2 fused to the TorA signal peptide (left) alongside the signal peptide control, KK, that is blocked in export (right). The middle lane shows a non-heme binding variant that is also blocked in export. **B**: Fractionation of BT6h2 with no signal peptide to act as further controls for correct processing by the Tat system. Molecular weight markers are shown on the left (in kDa) and mature BT6h2 is indicated on the right.

The E. coli TatABC system rejects a synthetic protein in its apo form

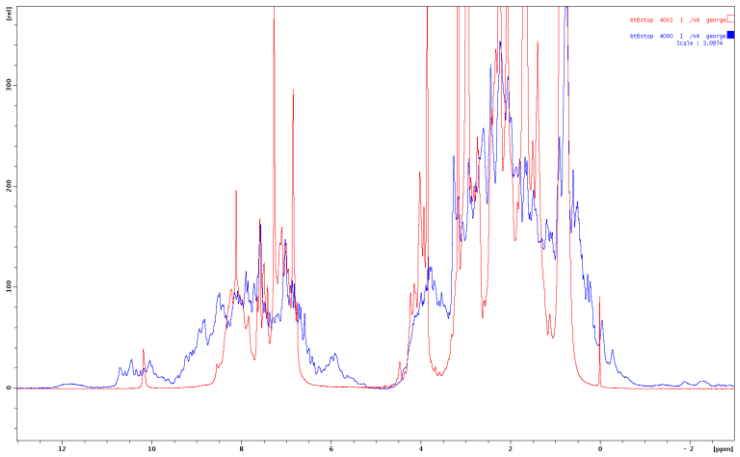
Ligation of heme into the interior of maquettes is known to increase the structural stability over the apo-form (Farid *et al* 2013), 'tightening' structure and significantly increasing T_m. We used this characteristic to probe if Tat proofreading could identify a synthetic substrate that is not strictly folded i.e. a substrate that has obtained significant structure but lacks its cofactors. In our BT6 each heme is coordinated by two histidine residues and thus we created two versions of BT6: the first has a single His to Ala substitution such that only a single heme *b* ligates, denoted BT6h1. Second we substituted two His with Ala so that no hemes ligate, generating the apo-form denoted BT6h0.

Each variant was expressed in the cytoplasm of *E. coli* BL21 (DE3) and purified via a Nickel affinity column. Figure 20 shows 1D proton NMR with purified BT6 variants (work carried out by George Sutherland at University of Sheffield), with and without bound heme. The broadening of the peaks observed when BT6h2 is bound to two heme cofactors is indicative of tertiary folding (panel A). The same phenomenon is observed with BT6h1 when the single heme is bound, although to a lesser degree (panel B). Finally, panel C shows no change in tertiary structure even in the presence of heme for the non-heme-binding BT6h0. The top and middle panels in Figure 21 show overlays of all three variants without and with their maximum heme binding (left and right respectively).

Cells expressing RRBT6h2, RRBT6h1 and RRBT6h0 were fractionated and anti-His blots to the cytoplasm/membrane/periplasm fractions are shown in the bottom panel of Figure 21. Here holo-BT6 (RRBT6h2) is exported and processed to the mature size as before, evidenced by the strong band in the periplasm. The single heme mutant (RRBT6h1), although exported and processed to mature size, has a distinctly weaker band than that of the wild-type indicating not all of the substrate is accepted as folded by the

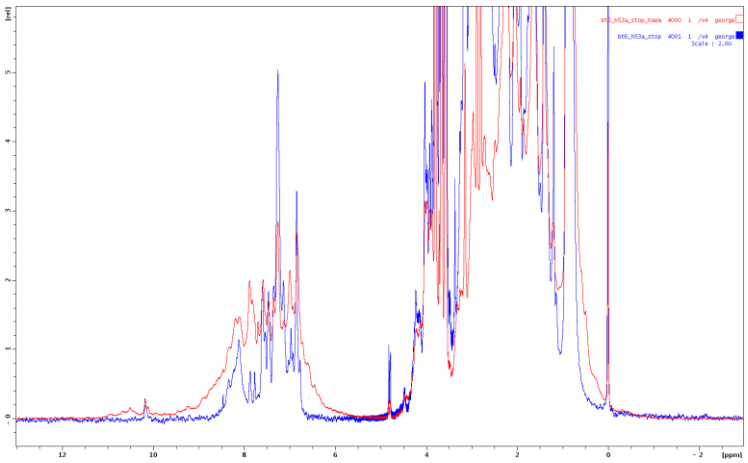
Tat machinery. Finally, the apo-form of BT6 (RRBT6h0) which doesn't bind either heme, is blocked for export. Precursor bands are evident in the cytoplasm (faintly) and membrane indicating BT6h0 is expressed and targeted correctly however is not accepted by the Tat proofreading system as no band is apparent in the periplasm.

A



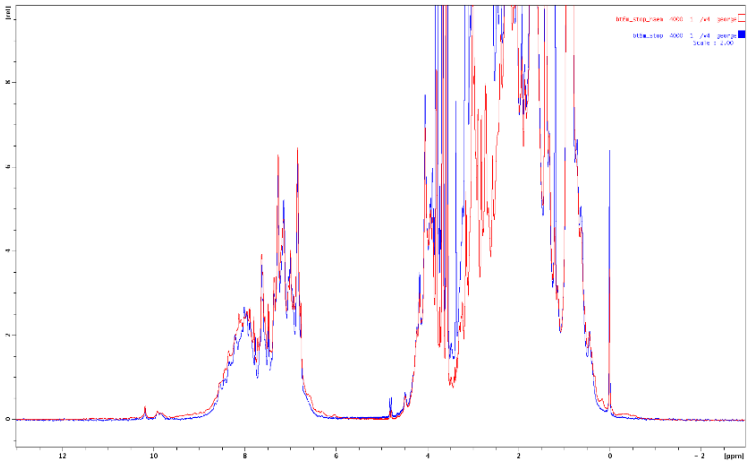
BT6h2: blue = with heme, red = without heme

B



BT6h1: red = with heme, blue = without heme

C



BT6h0 red = with heme, blue = without heme

Figure 20: **1D proton NMR for BT6 heme-binding variants. A**: the broadening of the peaks in the absence and presence of heme *b* for BT6h2 indicates a tightening of tertiary structure (red and blue respectively). **B**: Peak broadening is slightly less pronounced for BT6h1, which is only able to bind a single heme *b* molecule (blue – without heme, red = with heme). **C**: there is no change in the NMR trace in the absence or presence of heme *b* for the BT6h0 variant (blue/red respectively). Images generated by George Sutherland at the University of Sheffield.

The apo form of BT6 can be exported by the Tat system only when linked after the mature form

Our next step in this study was to see if we could ‘‘trick’’ the Tat system to accept the apo- form as folded and we thought it possible this could be accomplished if it was accompanied by a folded companion. Therefore, RRBT6h2-BT6h2, RRBT6h2-BT6h0 and RRBT6h0-BT6h2 were created. These are the wild-type BT6 linked to the apo-BT6 in various configurations via a 10 amino-acid linker and a C-terminal HA-tag, shown schematically in Figure 22.A.

Early studies with maquettes attempted to increase the number and variety of cofactors able to bind to a single molecule by increasing the number of α-helices. Initially these were homodimers created by tethering two helix-loop-helix domains via a disulphide bond. The repetitive nature of the amino acid sequences in these constructs limited the number and variety of cofactors able to bind however. This problem was overcome by replacing the disulphide tether with a linker between helix 2 and 3. This tether is a variation of the flexible repetitive $(Gly_4/Ser)_3$ linker used in scFvs to link the two immunoglobulin domains and is composed of the sequence GGSGSGSGG. Our linker was composed of six histidine residues

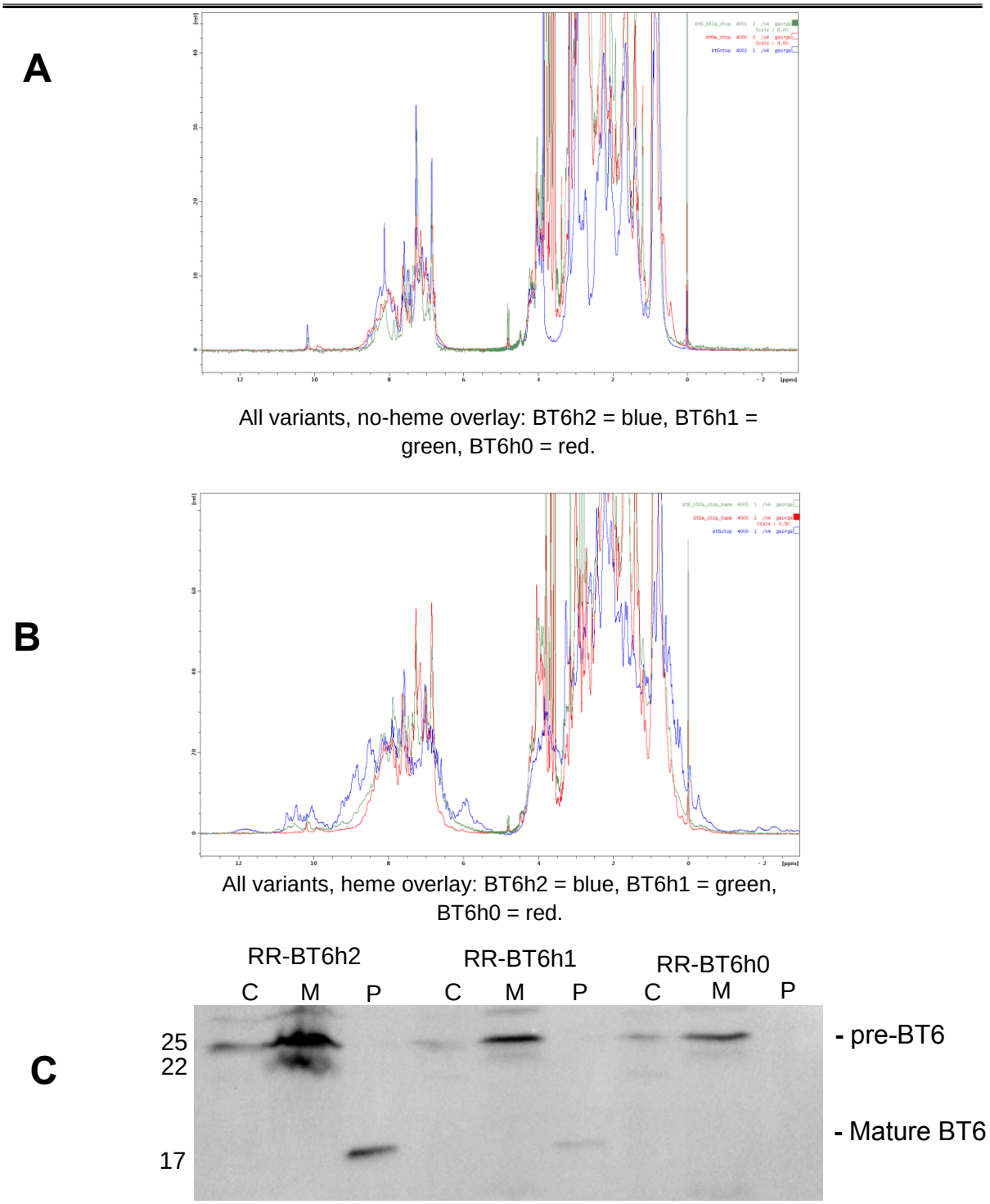


Figure 21: **Combined 1D proton NMR and export assays for the heme-binding BT6 variants. A**: Overlay for all variants in the absence of heme *b*. **B**: Overlay in the presence of hemeB. In the absence of heme *b* the traces are superimposable for each variant whereas in the presence of heme *b* peaks become more broad with increasing number of hemes able to bind, indicative of tighter tertiary structure (BT6h2 = blue, BT6h1 = green, BT6h0 = red). **C**: Export assay for each of the heme-binding variants. A decrease in export efficiency is observed for decreasing number of hemes able to bind. Molecular weight markers are shown on the left (in kDa) with precursor and mature forms indicated on the right.

and the first four amino acids of mature TorA, HHHHHHAQAA. This was deemed a suitable linker as the residues can form fairly flexible chains and the previously exported BT6 has both the N-terminal signal peptide (including AQAA) and the C-terminal six histidines (as a 6xHis-tag), neither of which hinder translocation compatibility.

The single-chain form of maquette generated by incorporation of these linkers not only had greater freedom for variety of ligation sites and enabled the sequence of each α-helix to be different, but it was also shown to still correctly fold (Farid *et al* 2013) into 4-helix bundles. Thus we can be reasonably certain our $(\text{helix-loop-helix-loop-helix-loop-helix})_2$ maquette is folded as shown in Figure 22.B.

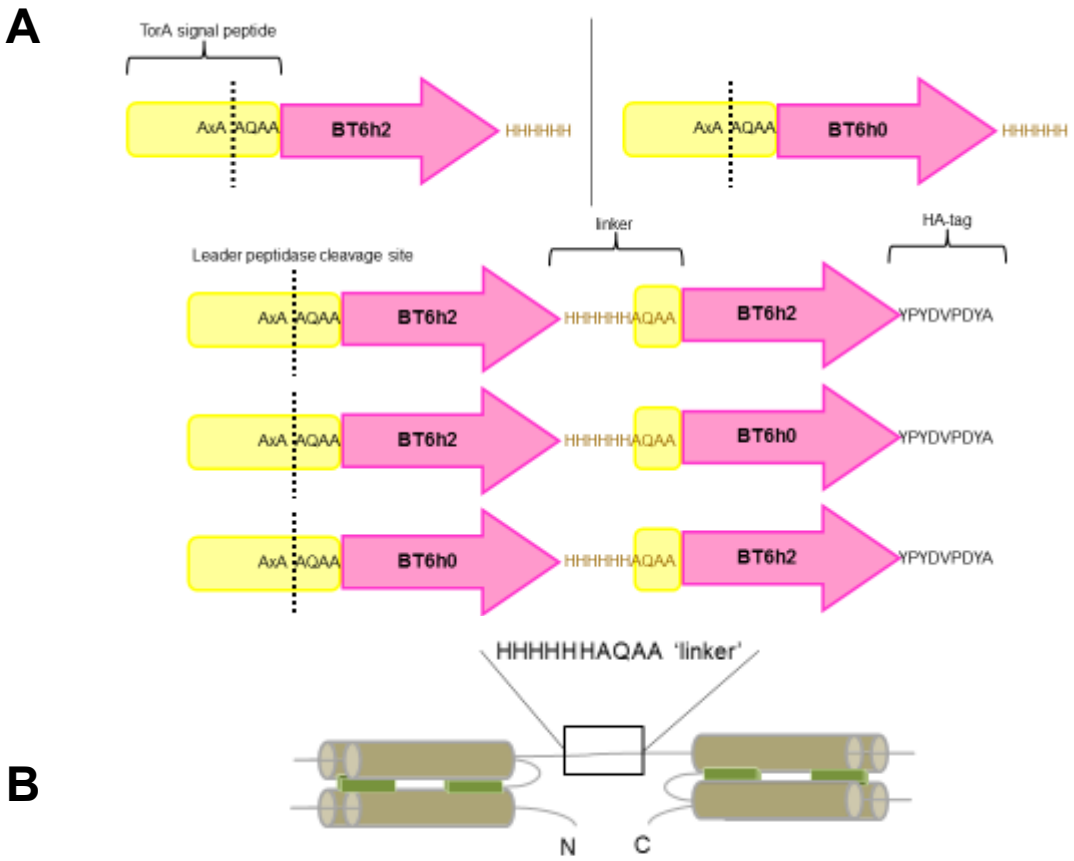


Figure 22: **Schematic representations of linked BT6 variants. A**: Dual heme-binding BT6 was linked to the apo-BT6 (zero heme-binder) in various configurations. TorA signal peptide is yellow, with AxA cleavage site and first four amino acids of mature TorA (AQAA) shown. Pink indicates the mature protein. 10 amino acid linker and HA-tag are indicated in brown and black respectively. **B**: Schematic representation of predicted fold for linked BT6 constructs shown in A.

Due to the more acceptable expression level for our purposes, constructs were expressed from vector pEXT22 under the same conditions and fractionated. The top panel of Figure 23 shows anti-HA blots of C/M/P fractions from cells expressing RR-BT6h2-BT6h2, RR-BT6h2-BT6h0 and RR-BT6h0-BT6h2. Here it can be seen that the non-mutated form RRBT6h2-BT6h2 is exported by the Tat system, albeit not very efficiently, with the periplasmic (P) protein at the processed size of 34, 448 Da, just above the 32 kDa marker. The middle column is interesting because surprisingly, there is also mature-size protein evident in the periplasm for the non-mutated BT6 linked to apo-BT6 (RRBT6h2-BT6h0).

Whether this is due to the mature form stabilising the apo-form or if the Tat system only proofreads upon initial docking of substrate and not throughout translocation, is beyond the scope of this study. Evidence for the latter scenario is shown in the final column however, where placing apo-BT6 first results in rejection by the Tat proofreading mechanism as there is no band evident in the periplasmic fraction.

The bottom panel in Figure 23 shows a control to ensure the linked BT6 variants are exported exclusively by the Tat system by expression of the constructs in *ΔtatABCDE* cells. As expected, no protein is present in the periplasm lane, indicating protein is indeed processed by the Tat system alone.

TorA KR versions of the linked BT6 variants were also created to see if a lack of the dual arginine residues has an effect on the proofreading capacity of the Tat system (data not shown). The same results as 'RR' signal peptide was observed.

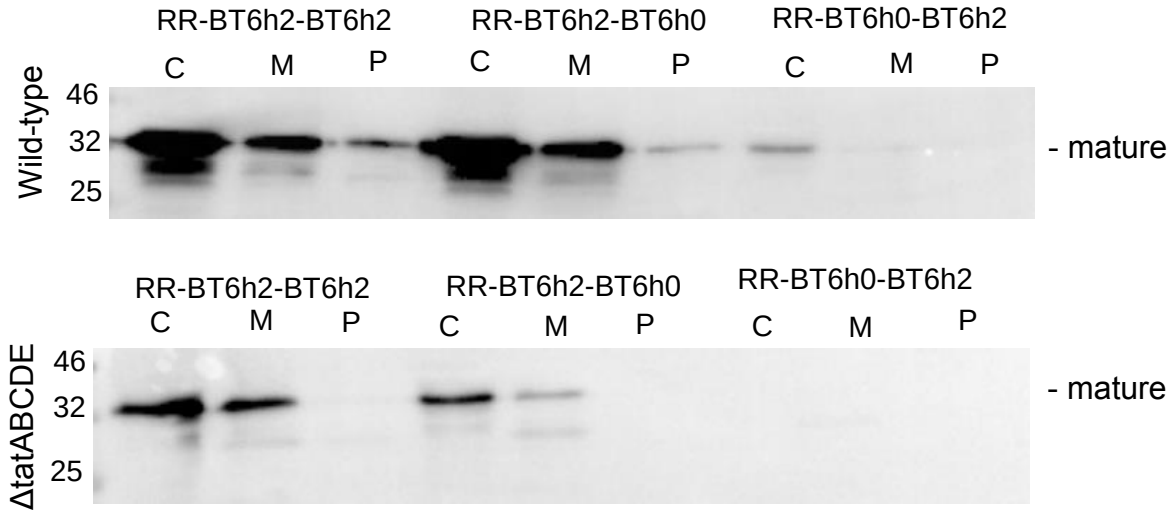


Figure 23: **BT6h0 can be exported by the Tat system when linked behind BT6h2.** A linked BT6h2 is efficiently exported via the Tat system, while BT6h2 linked to BT6h0 is exported with reduced efficiency. Placing BT6h0 first results in a block in export (left, centre and middle lanes respectively). Constructs are shown schematically in Figure 22. Cultures were fractionated into C/M/P as in previous studies in this thesis. While the bottom panel shows expression in *tat* null strain. The lack of band in the periplasm (P lane) signifies these linked BT6 variants are still processed by the Tat system. Molecular weight markers are shown on the left (in kDa) with mature protein indicated on the right.

Discussion

This study set out to analyse the capacity of the Tat system to export a wholly synthetic protein maquette, one that is completely designed *de novo*, to the periplasm in *E. coli*. In the initial stage of this study, BT6 was expressed in vast quantities from plasmid pJexpress414. Although expression at this level resulted in some processing by the Tat system, as shown by a mature-size band in the periplasm fraction, the vast majority of protein was found at precursor size, or in the periplasm at mature size even when no targeting signal was present. There are two scenarios that could cause this effect: 1) BT6 interacts with the membrane and destabilises it. As maquettes are attempts to mimic naturally membrane-associated proteins, such as those of the electron transport chain, this is a high possibility. 2) cytoplasmic contamination of the periplasmic fraction during fractionation could occur. This scenario cannot be ruled out, but is unlikely given the protein profiles of cytoplasm, membrane and periplasm fractions are different and do not share banding patterns (Figure 18.C). Additionally, a combination of these scenarios cannot be ruled out - where membrane destabilisation by membrane-associated BT6 leads to cytoplasmic leakage during fractionation. Nevertheless we attempted to counter these issues by expressing the Tat-targeted BT6 from a lower-expression plasmid. These experiments show the Tat system can indeed identify the *de novo* designed BT6 and transport it to the *E. coli* periplasm. Moreover, mutations to the 'RR' motif in the Tat signal peptide indicate BT6 is transported solely by the Tat pathway.

In the next stage of this study we took advantage of the effect on BT6 structure for binding to the heme *b* cofactor, and we thus substituted histidine ligation sites with alanine to block binding of a single or both heme *b* cofactors. NMR has previously shown the increase in tertiary structure exhibited by BT6 in a heme-free, single-heme and bis-heme states (Farid *et al* 2013 and Figure 20 A, B & C and Figure 21: Top panel), and we thought

this an excellent opportunity to test the Tat proofreading system, not only in general but also on a protein not found in nature. In the absence of heme i.e. when BT6 has the most structural flexibility, the Tat proofreading system rejects the BT6 substrate for export. Moreover, as heme *b* ligation sites are increased from zero to two the suitability for Tat-mediated export also increases. These data suggest that as BT6 loses structural flexibility upon increasing number of cofactor binding it becomes more acceptable to the Tat proofreading mechanism, similar to the scenario observed when scFvM is expressed in cells with an oxidising cytoplasm and is able to form disulphide bonds (chapter 3). In this scenario, a more flexible substrate would interact with the Tat machinery upon initial binding, which would in turn preclude transport of substrate across the membrane.

In the final section of this study we sought to address whether we could get the Tat proofreading mechanism to accept a substrate with a structurally flexible domain. To this end we joined the bis-heme binding BT6h2 to the heme-free BT6h0 in various configurations via a short, 10 amino acid, linker. The data obtained show that, perhaps surprisingly, the more structurally flexible BT6h0 is accepted for translocation via the Tat system only when linked behind a bis-heme-binding (i.e. more structurally ridgid) BT6h2. This scenario suggests that proofreading of substrate takes place only upon initial interaction of substrate with the Tat machinery, likely after signal peptide binding to TatBC but before translocation, and once accepted by the Tat proofreading mechanism, substrate is committed to export. The data presented here along with that observed by others (Cline *et al* 2007, Lindenstraub *et al* 2009, Taubert *et al* 2014) would make sense if such a scenario was the case and proofreading isn't an ongoing process throughout the translocation process. This mechanism would also explain the delay between substrate binding at the TatBC complex and translocation across the membrane, both processes that occur on a timescale of seconds. Further work would need to look at the exact structure

or folded state and heme-occupancy of the double BT6 maquettes presented in this study in order to more closely identify characteristics that Tat proofreading system identifies.

This study has not only provided interesting insights into the proofreading by the Tat machinery but has also shown the system is amenable to the translocation of protein substrates never before seen in nature to the periplasm of *E. coli*. This is convenient as it could prove useful for downstream processes (e.g. purification) of *de novo* designed proteins which is easier to achieve with a periplasmic protein.

Comparison of quality of human Growth Hormone (hGH) exported to the periplasm via the General Secretory (Sec) or Twin Arginine Translocase (Tat) pathways.

The market value of all licenced recombinant biotheraputic proteins has easily surpassed a hundred-billion dollars per year. 68% of recombinant proteins are derived from microbial platforms with purification from *Escherichia coli* the most common non-mammalian system (Walsh *et al* 2014). Expression takes place in the cytoplasm, at which point several options are available: purification from the cytoplasm, refolding from inclusion bodies, export to and purification from the periplasm via the General Secretory (Sec) pathway or, in rare cases, secretion to the culture medium via OsmY fusions. Secretion to the culture medium has significant benefits as it would enable continuous growth while siphoning off product, decreasing time and costs. However, this method is still very much in its infancy (Georgiou and Segatori 2005). Secretion to the periplasm is relatively simple, only requiring an N-terminal signal peptide and associated protein-conducting channel (section 1.1 and 1.2 and reviewed in Yuan et al 2010). Periplasmic expression enables a simplified purification procedure by selectively rupturing the bacterial outer membrane (Pierce et al 1997). This minimises the number and abundance of contaminating proteins, DNA and cellular debris (Balasundaram et al 2009). As discussed in section 1.2, the Sec pathway requires proteins to be in an unfolded state for translocation. As an alternative method of localising proteins to the periplasm, the Tat pathway can transport fully folded proteins. A unique feature of the Tat pathway though is its in-built proofreading mechanism that can detect structurally dynamic substrates and reject them for export (chapter 3 & 4), an important point for those recombinant proteins requiring disulphide bond formation (e.g. mAbs and hormones) or cofactor insertion. This proofreading capability could indicate a purer, more homologous product is present in the periplasm and thus a greater yield upon purification. In this chapter, we have sought to address this issue by targeting a model recombinant biotheraputic – human Growth Hormone (hGH) – for export to the periplasm of *Escherichia coli* via either the Sec or

Tat pathway. hGH was then purified from the periplasm by Ni IMAC before being subject to a variety of assays to assess any structural or folding differences or heterogeneity of product.

Results

The Sec system is more metabolically draining that the Tat system

In this study the aim was to highlight any structural differences between proteins produced and purified from the periplasm when targeted to either the Sec or Tat pathways. To this end we chose a well characterised model protein with biotheraputic relevance – human Growth Hormone (hGH). We created two identical plasmids that transcribed hGH and placed either a Sec or Tat signal peptide at the N-terminal. The Sec signal peptide was taken from the Outer membrane protein (OmpA-hGH), while the Tat signal peptide was taken from TMAO reductase (TorA-hGH). These are the traditional signal peptides used for targeting proteins to the Sec and Tat pathways respectively.

Figure 24 shows growth curves for *E. coli* BL21(DE3) transformed with either the Sec or Tat plasmid (blue and orange respectively) from inoculation, through induction with 0.5 mM IPTG (3 hrs, red arrow) and on to fractionation 3 hrs post-induction (6 hrs growth total, green arrow). Here it can be seen that although cultures harbouring Sec or Tat plasmids have the same overall growth profile, cultures expressing TorA-hGH reach significantly higher OD_{600} over this period (OmpA: 0.678, TorA: 1.965 average of six curves). Prior to induction, growth is similar for cultures with either plasmid. However, post-induction growth for cultures expressing OmpA-hGH decreases dramatically, appearing to almost enter stationary phase. Conversely, cultures expressing TorA-hGH continue in exponential phase until 2½ hrs post-induction. This becomes clearer when

growth curves are combined in Figure 25.A and compared to an empty strain (maximum OD_{600} = 2.295).

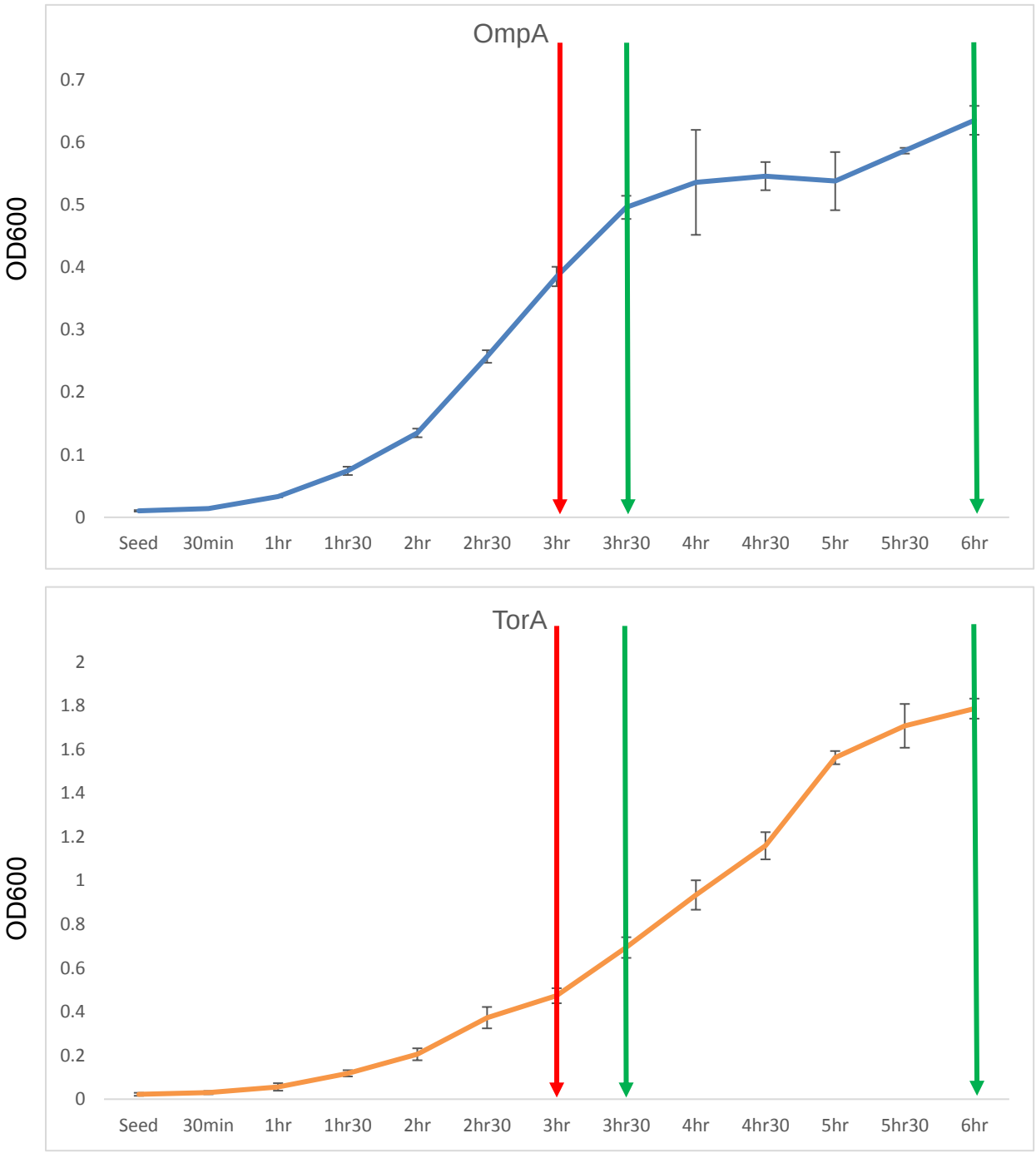


Figure 24: **Cells grow more slowly over a 3 hr period when rhGH is targeted to the Sec pathway over the Tat pathway.** Growth curves for cultures harbouring plasmids encoding human growth Hormone (hGH) targeted to the periplasm via either the Sec (OmpA) or Tat (TorA) pathway. Induction with 0.5 mM IPTG was at 3 hrs (red arrow) and fractionation was carried out at 30 min post-induction and 3 hr post-induction (green arrows). Each curve was generated from the average of 6 separate cultures. Maximum OD_{600} were: OmpA 0.678, TorA 1.965 and Empty 2.295.

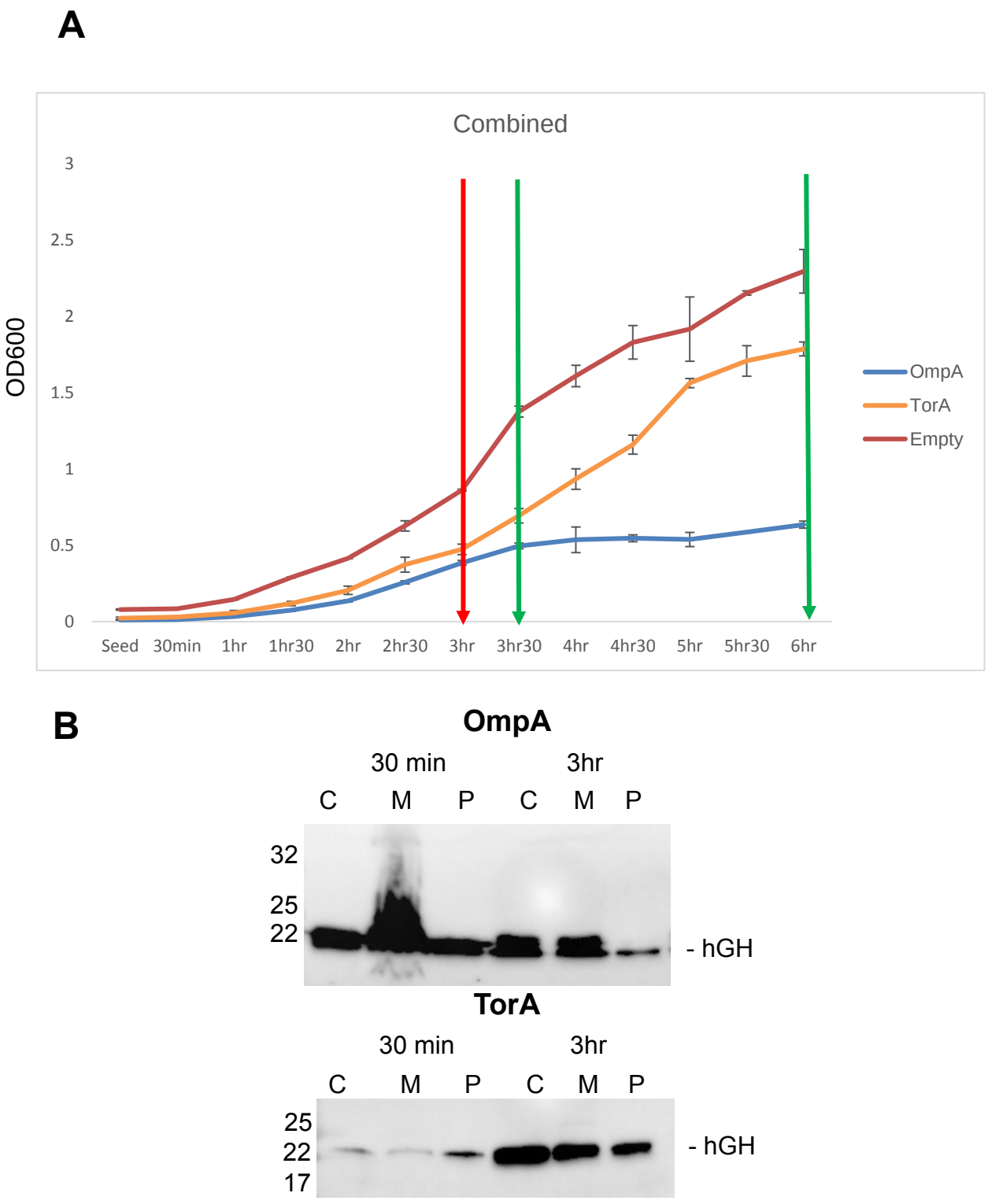


Figure 25: **hGH targeted to the periplasm via the Sec pathway is immediately available whereas Tat-targeted protein builds up over a longer time period**. **A:** The combined growth curves more clearly show the inhibitory effect on cell growth from targeting hGH to the periplasm via the Sec pathway. **B**: Samples 30 minutes post-induction are shown in the left three columns (C/M/P) and after 3 hrs in the right. Fainter bands are observed for 30 minutes post-induction for Tat-targeted protein. Molecular weight markers are shown on the left (in kDa) and mature hGH is indicated on the right.

Fractionations of culture into cytoplasm, membrane and periplasm at time points 30 min post-induction and 3 hr post-induction are shown in Figure 25.B. Here we see that after 30 minutes induction, OmpA-hGH has already built up significantly across all three fractions (left three columns) and remains consistent up to 3 hrs (right three columns). Whereas TorA-hGH is barely evident 30 minutes after induction and a consistent increase in protein happens over the 3 hrs of induction. After 3hrs induction, cells were fractionated into cytoplasm, membrane, periplasm and insoluble (I) fractions in an attempt to identify any bottlenecks or saturation of the pathway. Figure 26 reveals there is a marginal amount of hGH present in the insoluble fraction for Sec-targeted protein (equivalent to the amount in other fractions). For protein targeted to the Tat-pathway however there is a large amount of protein present in the insoluble fraction.

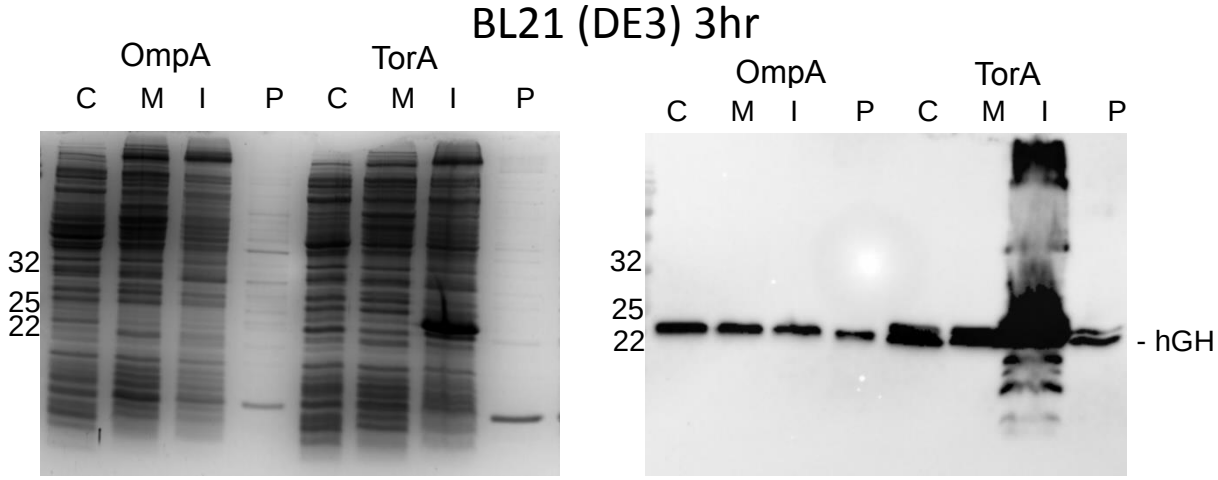


Figure 26: **Targeting hGH to the periplasm via the Tat pathway results in a larger insoluble fraction.** After 3 hrs induction, cultures were fractioned into C/M/P and Insoluble fraction (I). There is a large amount of insoluble hGH when targeted to the Tat pathway (TorA). Left: Coomassie stained gel, the predominant band in TorA 'I' lane is hGH. Right: Immunoblot of the gel shown on the left.

Sec and Tat transported human Growth Hormone are equally folded

Previous chapters have shown the Tat system more readily exports proteins that are in a less dynamic state (i.e. more tightly folded), therefore we thought it pertinent to assess the homogeneity and folding characteristics of Sec- and Tat-targeted hGH. We developed the hypothesis that as the Tat system more efficiently exports structurally rigid proteins, protein reaching the periplasm via this route would have a 'tighter' structure.

First, hGH was purified from the periplasm after 3 hrs induction for OmpA- and TorA-hGH by Nickel IMAC (section 2.4.1). The protein elutes immediately within the first 5 fractions, to a pure degree for protein exported via either pathway as shown by the absence of contaminating band in the Coomassie stained gel (left, Figure 27). Some TorA-hGH is lost in the wash fraction as evidenced by a band in the western blots (middle and right panels). Fractions containing pure protein were pooled (E1 – E5 in both cases) and were dialysed overnight into sodium phosphate buffer (section 2.4.3). The following day dialysed sample was spun at 4000 rpm in VIVASPIN 20 10, 000 MWCO spin columns (Sartorius, Gottingen, Germany) to a volume of 250 µL. The concentration determined by A_{280} and a calculated molar extinction coefficient of 17 670 M^{-1} (Expasy Protparam) as 2.1 mg/mL and 1.3 mg/mL for OmpA- and TorA-hGH respectively (Figure 28), and verified by NanoDrop.

One-dimensional proton (^1H) NMR spectroscopy was used to assess the tertiary structure of each of the hGH samples.

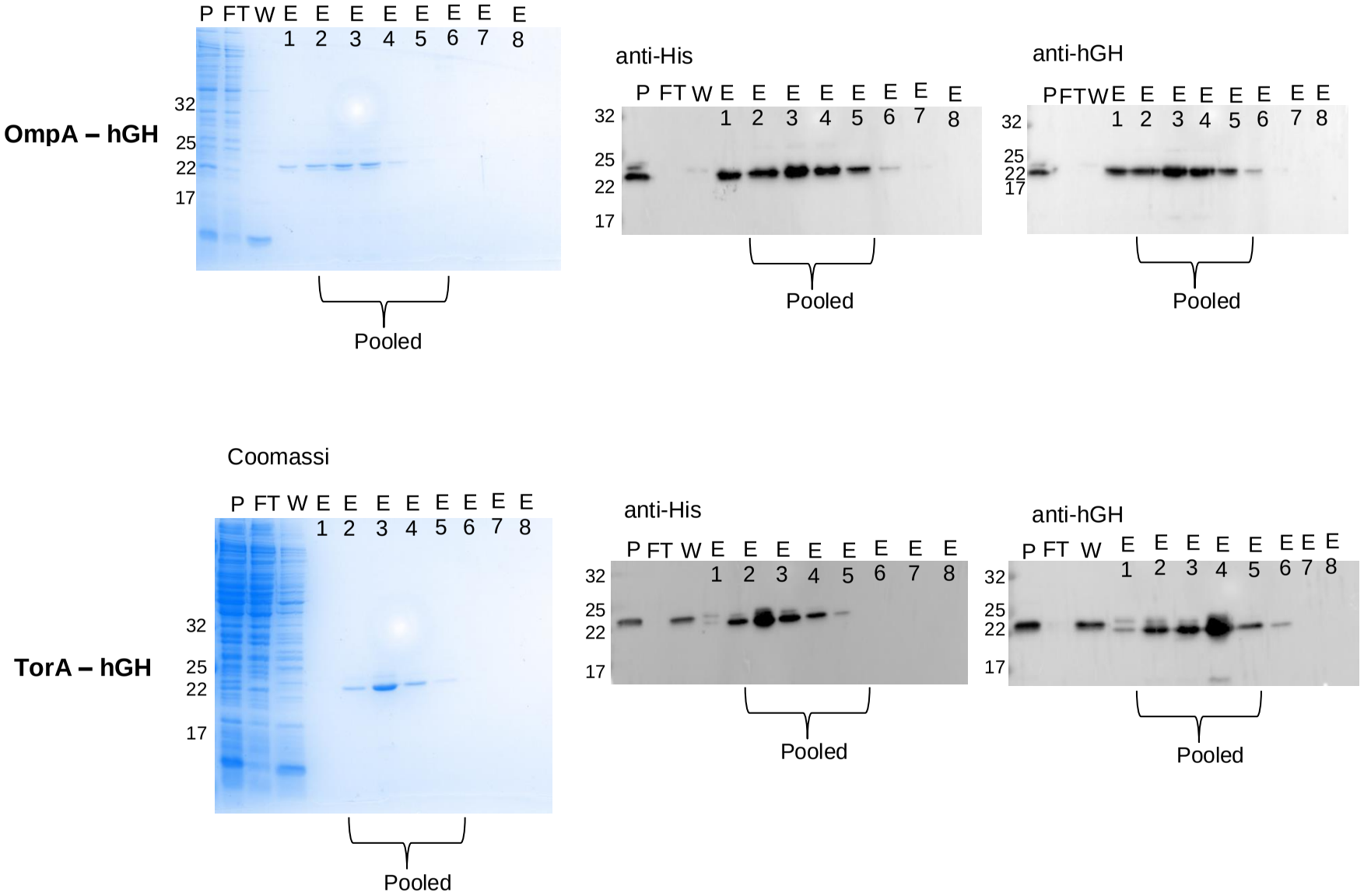


Figure 27: **Ni IMAC purification of hGH from the periplasm.** The top row shows purification from the periplasm for hGH via the General secretory pathway, while Tat hGH is shown in the bottom row. Left: Coomassie stained gel of the periplasm (P), flow through (FT), Wash (W) and elution fractions (E(n)). Pooled fractions are indicated. Note the lack of contaminating bands. Middle and Right: western immune blot to the C-terminal 6x histidine tag (middle) or anti-hGH (right).

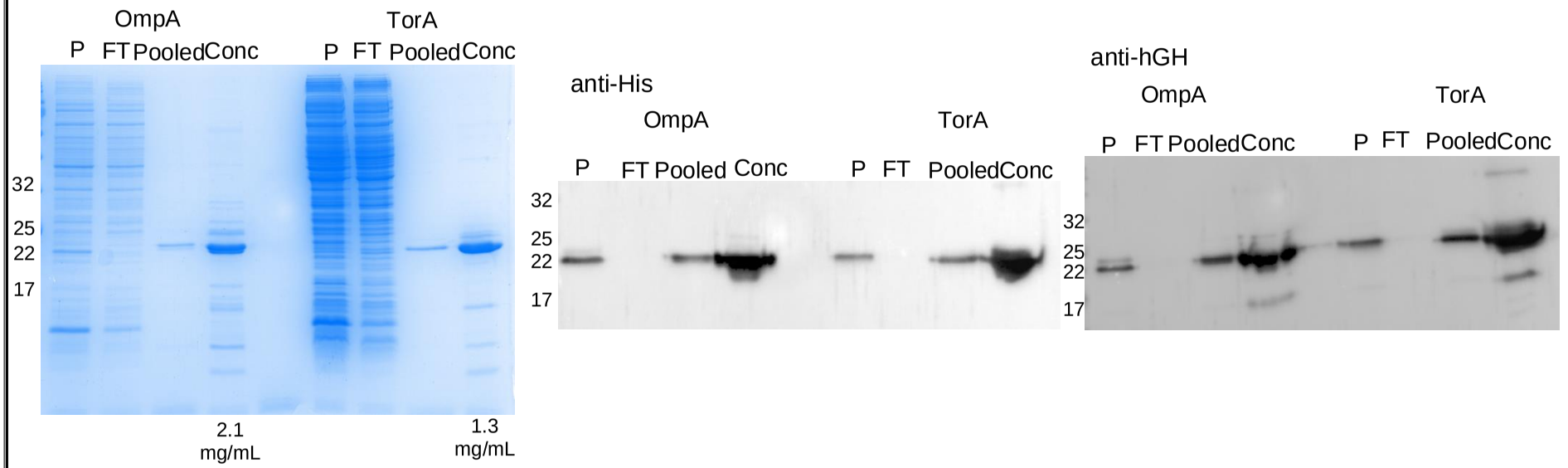


Figure 28: **Purified hGH.** Concentration of hGH in each sample was calculated by A_{280} and a molar extinction coefficient of 17 670 M^{-1} as 2.1 mg/mL and 1.3 mg/mL for Sec secreted and Tat secreted respectively.

Sec- and Tat-processed hGH were dissolved in a mixture of 5% D_2O/95%H_2O buffer containing 50 mM sodium phosphate, 150 mM sodium chloride at pH 7.4. These were analysed after extensive dialysis to reduce effects due to any systematic differences in sample preparation. Spectra were acquired at 298 K using a double pulse field gradient spin echo sequence (DPFGSE) to suppress water. Data were acquired on a Bruker Advance III spectrometer operating at 600 MHz with a He cooled cryogenic probe. Spectra were measured using 2048 complex data points over a sweep width of 9615 Hz using 1024 scans with an inter-scan relaxation delay of 1s. These data were processed and analysed using Topspin (Bruker) an exponential function of 8 Hz was applied to improve signal to noise, with zero filling being carried out to 8192 data points. A convolution based water suppression filter was used to suppress the residual HDO peak (omitted with ‘’//’’ in Figure 29).

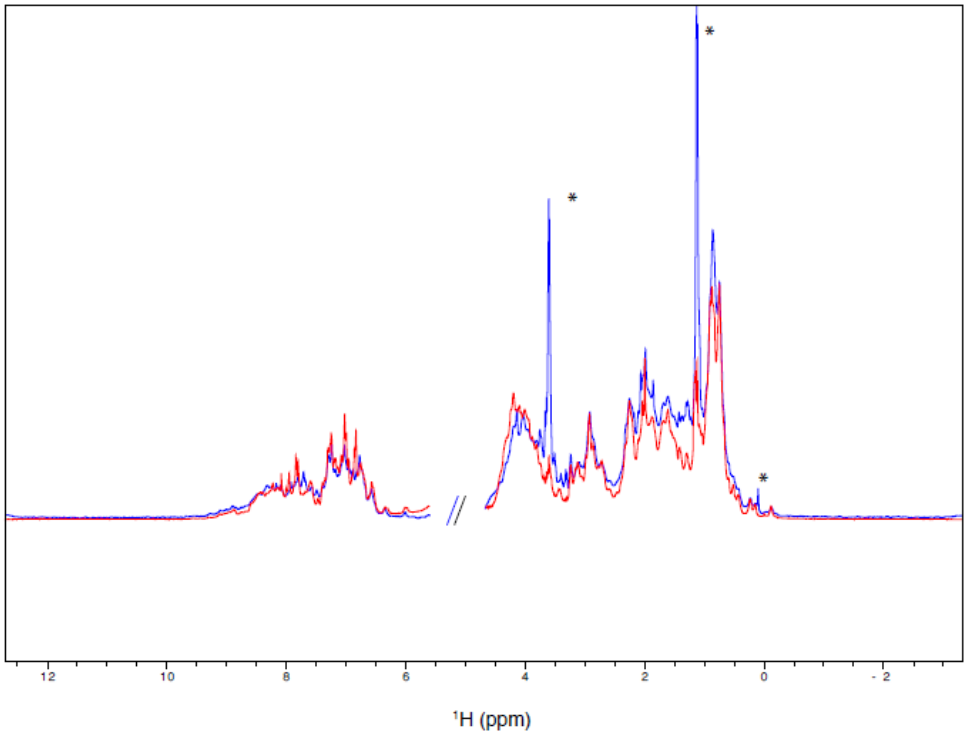


Figure 29: **Full spectrum proton NMR resonances for purified hGH processed by either the Sec or Tat pathway.** Overlay of the Sec-hGH (red) and Tat-hGH (blue) full spectrum (the large peaks at 1.2 ppm and 3.6 ppm are likely due to contaminating ethanol, asterisked (Fulmer *et al* 2010)). The HDO signal is suppressed with a double forward slash (//). NMR contributed by Dr Gary Thompson at the University of Kent.

Both Sec- (red) and Tat-hGH (blue) samples have spectra of a well-folded protein (Figure 29), as indicated by the broad dispersion of signal in the amide (Figure 30.B) and aromatic regions of the spectrum. Ring shifted methyl peaks, attributable to stacking of the methyl groups on aromatic rings in the core of a protein, are also observed (Kwan *et al* 2011) and these align between samples (Figure 30.A, indicated by black arrows). Overall the spectra are largely similar and subtle differences may be due to inconsistent salt concentrations between samples. The NMR work was carried out by Dr Gary Thompson at the University of Kent.

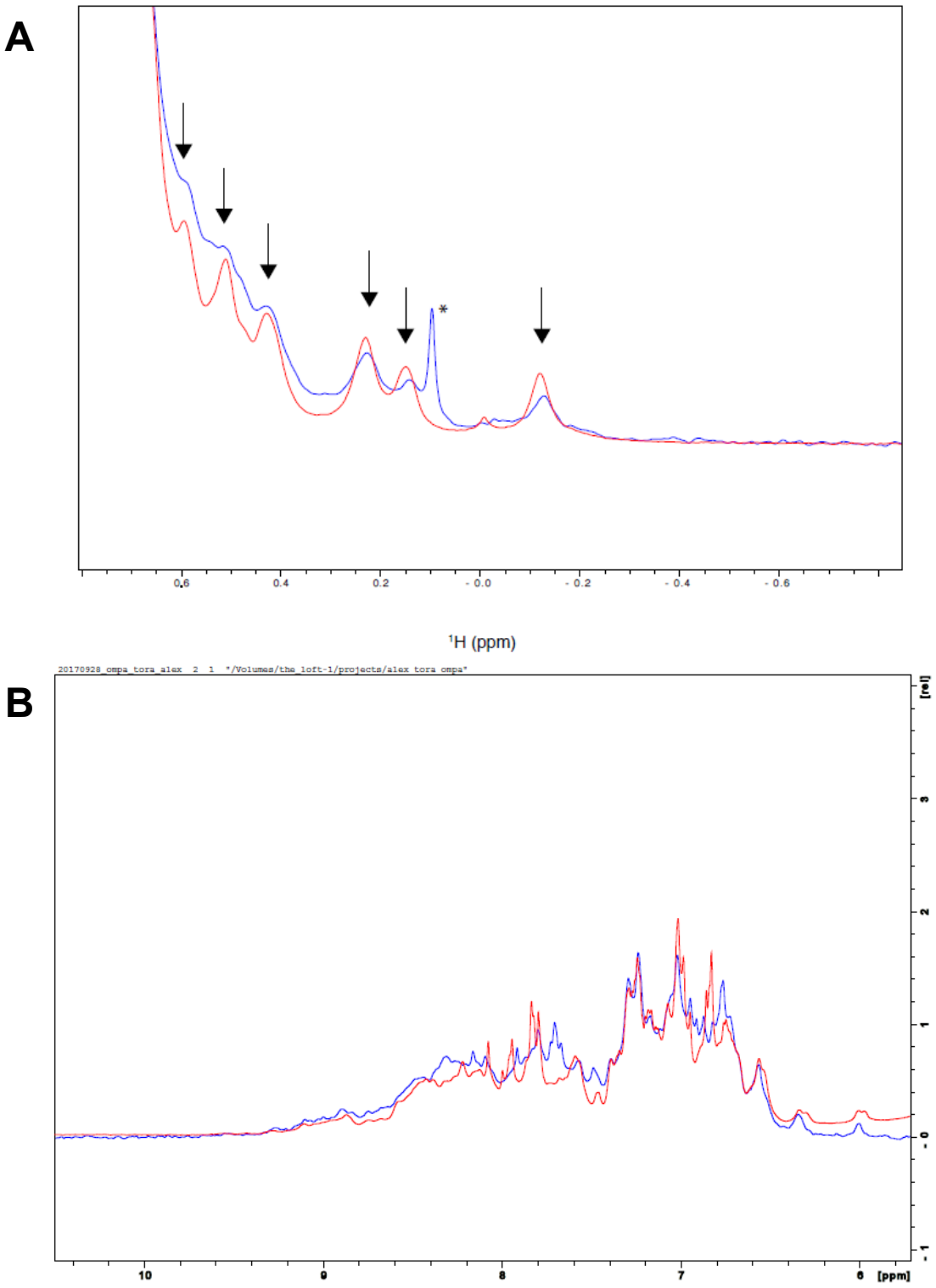


Figure 30: **Expansion of the methyl and aromatic regions for ^{1}H NMR spectra for Sec and Tat processed hGH. A**: Expansion of the methyl region with Sec (red) and Tat (blue) hGH overlaid. The black arrows indicate the ring-shifted methyl peaks indicative of a folded protein. The alignment of these peaks indicate protein processed via either pathway has the same fold. **B**: Expansion and overlay of the amide region. NMR contributed by Dr Gary Thompson at the University of Kent.

Discussion

The recombinant production of proteins in a bacterial host is the predominant platform for production of biotheraputics. Typically, these proteins are refolded from inclusion bodies or released from the periplasm, the latter being the less time-consuming and intense approach. Here we set out to compare the current mechanism for recombinant protein export to the periplasm via the Sec pathway with an alternative method – targeting to and export by the Tat pathway. To this end we secreted recombinant human Growth Hormone (rhGH) to the periplasm of *E. coli* via either pathway. While cultures expressing TorA-hGH grew to higher ODs, the amount of protein produced after 3 hrs was the same for both systems. Furthermore, targeting of hGH to the Tat system, resulted in a larger amount of insoluble product and this is likely due to the slow rate of export by the Tat system. Previous reports have shown if substrates are not committed to export within a certain time, they are Tat-dependently degraded. Purification and structural analysis by CD and Proton NMR of the products reveals that periplasmic rhGH is structurally similar and folded to an equal degree. The data also suggest the hGH produced by either means here are equivalent to a biopharmaceutical-grade hGH standard.

The Tat system is well known for employing a proofreading mechanism and previous chapters in this study have shown an increase in export efficiency for those substrates that are less structurally dynamic. The initial premise that export to the periplasm via the Tat pathway would produce a more structurally ridgid protein was not proven, in fact the proteins are of an equal quality. However, it has been shown that hGH is active and folded without formation of disulphide bonds (Bewley *et al* 1969, Youngman *et al* 1995) and previous studies with the Tat system have shown disulphide bonds are not necessary prior to export for hGH (Alanen *et al* 2015). Therefore, the cytoplasmic disulphide-forming system, CyDisCo, was not used in this chapter. When the Tat system is combined with

the CyDisCo components, a more structurally compact product may arise and this would be an intriguing course for further study.

Truncation analysis of Tat components effect on proofreading capacity.

Previous chapters have shown that the Tat system is capable of exporting two heterologous proteins; an scFv and a completely synthetic hemeB-binding maquette termed BT6. Moreover, the Tat proofreading mechanism was shown to be tolerant of significant variances in surface charge, both positive and negative, and hydrophobicity, while also being able to identify and reject substrates with disordered domains or lacking a cofactor. Given these data we assume the Tat proofreading mechanism favours rejection of substrate proteins based on conformational flexibility. For example, an scFv lacking native disulphide bonds is exported with circa 10-fold less efficiency than the oxidised form (chapter 3). Similarly, BT6 is only exported when heme B cofactors are able to bind imparting a significant increase on structural rigidity (chapter 4). In this chapter we wanted to address the proofreading process in the Tat components and attempt to restore export of these export-incompatible substrates.

The *E. coli* Tat system is composed of the three integral membrane proteins TatA, TatB and TatC (section 1.3 and reviewed in Berks 2015). TatA and TatB are members of the same family of proteins that comprise a short N-terminal domain, a TM domain, a hinge region, an amphipathic helix and finally a variable-length, unstructured C-terminal domain (section 1.3.1.1 and 1.3.1.2). Several studies have sought to determine the core domains and key residues necessary in TatA-family proteins for functional translocation by the Tat system (Lee *et al* 2002, Barrett *et al* 2003, Hicks *et al* 2005, Barnette *et al* 2008, Warren *et al* 2009, Maldonado *et al* 2011). Mutagenesis studies identified that while the Gly at position 21 in *E. coli* TatA and TatB is conserved, it is not essential for function and substitution with equally flexible residues (e.g. alanine) results in a slightly reduced, but still functional translocase (Barrett *et al* 2003). Additionally, altering the charged nature, but maintaining the hydrophilicity, of the cytoplasmic face of the TatA APH has little effect on translocase functionality, whereas substitutions within the

hydrophobic face generally reduce Tat functionality to minimal levels (Barrett *et al* 2003, Hicks *et al* 2005). The moderate importance of these residues is highlighted by the perhaps interesting phenomenon that certain substitutions block export when the Tat components are expressed at wild-type levels (i.e. from the chromosome) whereas overexpression from a plasmid leads to reduced, but still functional, translocase activity (Warren *et al* 2009).

TatA is 89 amino acids while TatB is 171. Several truncation studies in *E. coli* all agree that truncation of TatA by up to 40 amino acids and TatB by up to 130 amino acids maintains functional export of proteins utilising the Tat pathway (Lee *et al* 2002, Warren *et al* 2009). The case in TatB is a little more complex as wild-type levels of export were observed for truncations up to 30 amino acids whereas truncations between 40 and 130 aas maintain a reduced level of transport activity (Lee *et al* 2002). Interestingly, these truncations in TatA and TatB leave a minimal functional protein composed of the N-terminal periplasmic domain, the TM domain and the APH. Indeed, in TatA highly acidic residues at positions 45, 46 and 47 (*E. coli* numbering), denoted an ‘’acidic motif’’, immediately following the APH were shown to be important for a functional Tat translocase at both wild-type and high expression levels (Warren *et al* 2009), possibly due to destabilisation of the TatA pore-forming complex (section 1.3.4). Moreover, these truncated TatA and TatB proteins are able to complement *tatAE*, *tatB* and *tatAEB* deletion strains (Lee *et al* 2002).

The results of these studies have led to some confusion as to the purpose of the, largely unstructured, C-terminal regions of the TatA-like components, especially when it is taken into consideration that these proteins from the majority of organisms in which a Tat system is present possess this unstructured C-terminal ‘domain’. If this region was not essential, surely selection pressure would favour Tat-components of a more truncated

nature. Therefore, in this chapter we have laid the groundwork for probing the Tat components, where further inquiry will hopefully delineate domains contributing to proofreading.

Results

Truncation of TatA by up to 40 amino acids is able to complement the ΔtatABCDE *phenotype*

We first constructed C-terminal truncations in the TatA subunit of the *Escherichia coli* TatABC complex by substitution of residue-encoding codons with STOP codons (TAA or TAG) to generate TatA-20, TatA-30, TatA-40 and TatA-45 amino acids in plasmid pWM-Tat0 which encodes the *E. coli* TatABC components followed by a Strep-II tag at the C-terminal of TatC and is arabinose inducible. The Strep-II tag was necessary to confirm expression of all three Tat components (as each subunit has its promoter within the preceding gene), and previous studies have shown a C-terminal Strep-II tag in this position does not inhibit Tat function (Warren *et al* 2009). Plasmids were induced with 0.05 mM arabinose and expressed for 3 hrs before an equivalent number of cells were fractionated into cytoplasm, membrane and periplasm and immunoblotting to the Strep-II tag, shown in Figure 31. A band is apparent at the appropriate size for TatC (28.9 kDa) indicating transcription of the entire TatABC codon is occurring. Interestingly, despite an equivalent number of cells, band intensity decreases for increasing truncation length.

Inactivation of the *Escherichia coli* Tat pathway through deletion of *tatAE*, *tatB*, *tatC* or *tatABCDE* genes show defects in cell division. This manifests as formation of cell-chains

up to 10 cells long (Stanley *et al* 2001), which are visible under a light microscope. This chain formation is likely due to the inability to transport the native Tat-substrates AmiA and AmiC to the periplasm. AmiA, AmiB and AmiC are cell wall amidases involved in murein cleavage

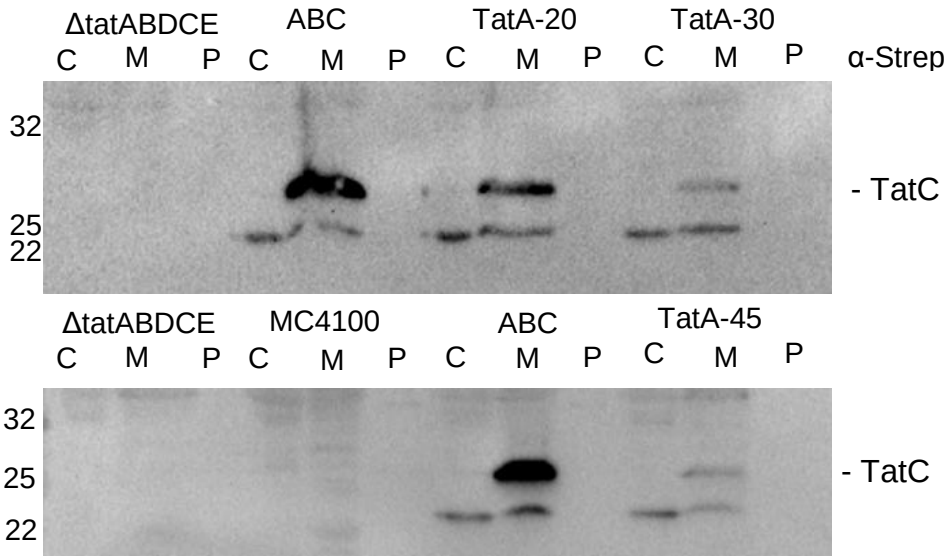


Figure 31: **Expression of truncated TatA from the arabinose-inducible Tat0 plasmid.** Wild-type and truncated TatA variants were expressed in a *tat* null strain. After 3 hrs induction equivalent cell densities were fractionated into C/M/P and immunoblotted to the C-terminal StrepII tag. Top: *ΔtatABCDE* strain with no plasmid is run alongside Tat0 expressing wild-type TatABC and TatA missing 20 or 30 aas from the C-terminal. Bottom: *ΔtatABCDE* and wild-type strain with no plasmid is shown against the non-functional TatA missing 45 aas. Plasmids were induced with 0.1 mM arabinose.

during cell division. While AmiB is a Sec substrate, it requires overexpression to complement the *ΔtatABCDE* phenotype (Ize *et al* 2003), and thus at native expression levels *ΔtatABCDE* strains have a filamentous phenotype.

Plasmid encoding either wild-type TatABC, TatA-20, TatA-30, TatA-40 or TatA-45 were transformed into MC4100 *ΔtatABCDE* strain and grown aerobically for 18 hrs at 30°C. The pWM-Tat0 plasmid is a pLysSBAD derivative and thus is leaky, therefore no inducer was used in an attempt to mimic wild-type TatABC expression levels (Warren *et al* 2009).

Images representative of cultures expressing these plasmids are shown in Figure 32 alongside *ΔtatABCDE* and wild-type MC4100 controls. Here the cell-chain phenotype characteristic of *ΔtatABCDE* strains is immediately obvious, with restoration of the wild-type phenotype observed upon expression of full-length TatABC (top row right and second row left respectively). A phenotype like that of the *ΔtatABCDE* strain is observed for TatA-45 with chains of up to 10 cells, consistent with a Tat system that is non-functional. Interestingly, a partial phenotype is observed for TatA-20, TatA-30 and TatA-40 which show progressively longer cell-chains, similar to that observed for *ΔtatA* but not *ΔtatE* or *ΔtatAE* strains (i.e. in strains with functional *tatBCDE*) (Stanley *et al* 2001).

To further test the functionality of the Tat systems with truncated TatA we screened for growth on LB agar plates supplemented with 2% SDS. Due to the cell separation defects, inherent in *E. coli* strains lacking a functional Tat system they are extremely sensitive to hydrophobic antibiotics (such as ampicillin) and the anionic detergent SDS (Stanley *et al* 2001, Ize *et al* 2003). Thus *E. coli* strains with a functional Tat system are able to grow on LB plates supplemented with up 4% SDS, while those with defective Tat apparatus are unable to survive when as little as 2% SDS is present (Ize *et al* 2003). For this reason, MC4100 *ΔtatABCDE* strains transformed with plasmids encoding TatABC, TatA-20, TatA-30, TatA-40 or TatA-45 were grown on LB agar plates supplemented with 2% SDS and 0.05 mM arabinose to induce Tat components and incubated at 37°C for 24 hrs. The left image in Figure 33 shows growth of *ΔtatABCDE* expressing TatABC, TatA-20, TatA-30, TatA-40 or TatA-45 on 2% SDS plates. Colony growth of TatA-20 and TatA-30 is comparable to the wild-type and *ΔtatABCDE* expressing full-length TatABC, whereas expression of TatA-40 only leads to partial growth and TatA-45 inhibits growth, mimicking the phenotype of *ΔtatABCDE*. These results reflect those observed for the filamentous phenotype, although it is surprising perhaps that TatA-20 and TatA-30

restore wild-type growth levels on SDS despite only partially complementing the filamentous phenotype.

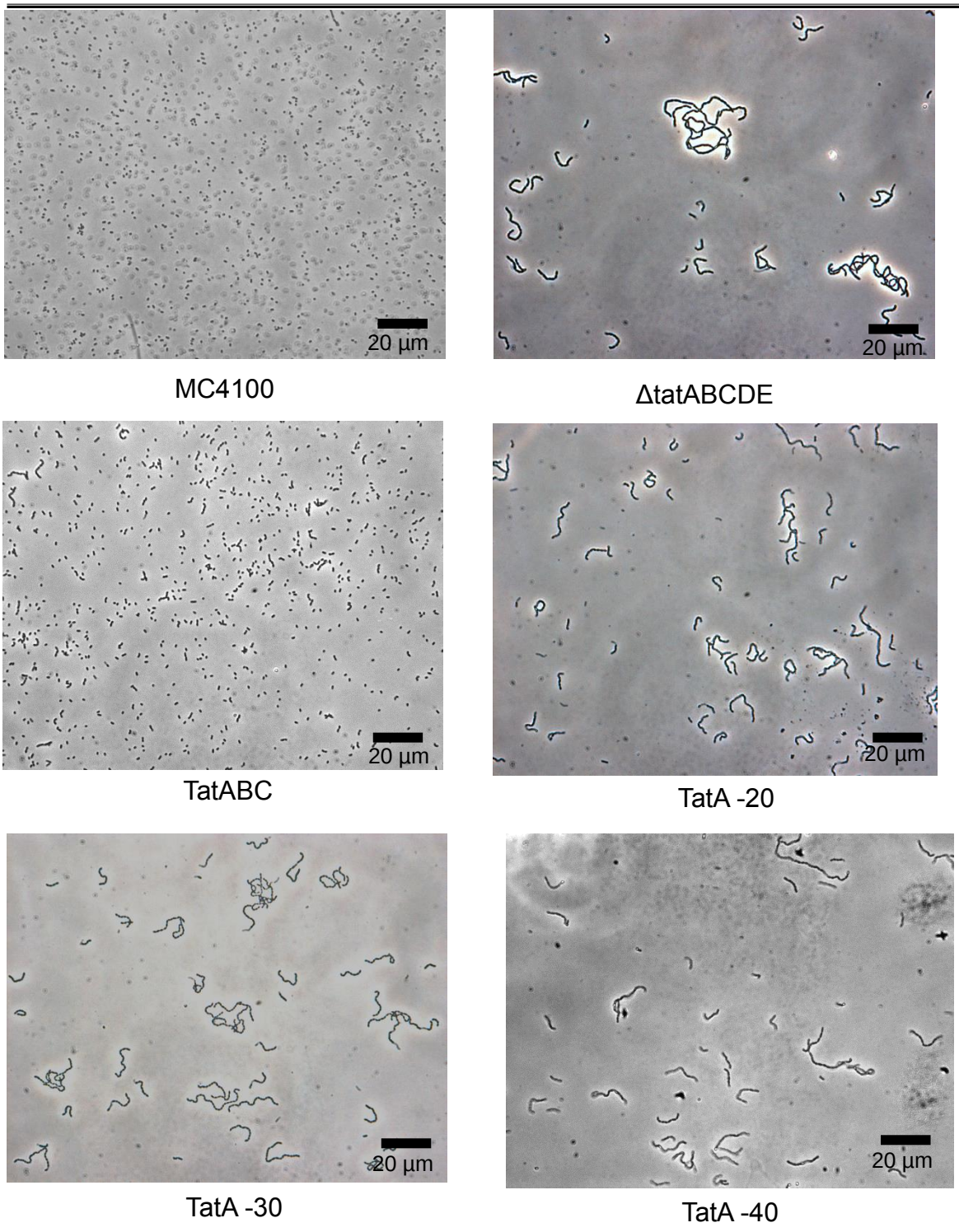


Figure 32: **Truncated TatA does not complement the filamentous phenotype of *ΔtatABCDE* cells.** Top: microscope images of wild-type and *ΔtatABCDE* cells. Middle: restoration of the wild-type phenotype in *ΔtatABCDE* cells through expression of TatABC from the Tat0 plasmid and plasmid encoding TatABC with TatA truncated by 20aas. Bottom: *ΔtatABCDE* cells expressing truncated TatA by 30 and 40aas does not complement the filamentous phenotype. Scale is shown in the bottom right. 400x magnification.

Truncation of TatB by up to 130 amino acids is able to complement the ΔtatABCDE *phenotype*

After assessing the impact of a truncated TatA component, we wanted to determine the impact of a truncated TatB component. Therefore, we constructed truncations in the TatB subunit of TatABC in the same manner as in the previous section for TatA – by substitution with STOP codons (TAA or TAG). TatB-20, TatB-30, TatB-50, TatB-75 and TatB-130 in pWM-Tat0 were thus generated. *ΔtatABCDE* cultures expressing these constructs were fractionated (C/M/P) and samples blotted against the Strep-II tag (A & B in Figure 34) A band at ~28 kDa is observed in the membrane fraction, the correct size for TatC, indicating successful induction of the plasmids. The intensity of the band appears greatly reduced in comparison to that of the full-length Tat components (TatABC lane), as observed for truncated TatA.

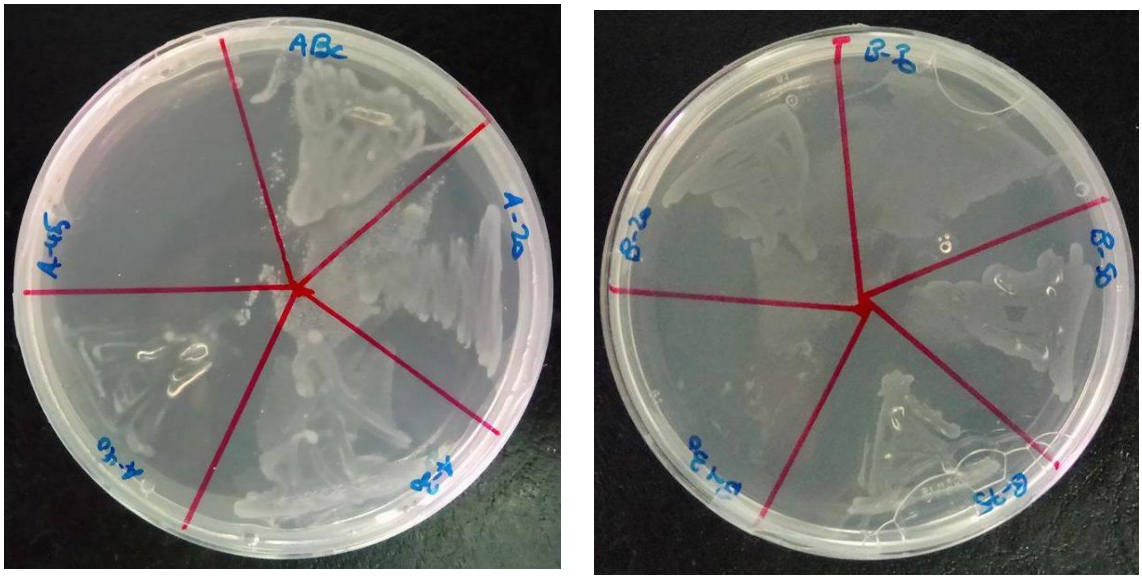


Figure 33: **Growth of truncated TatA or TatB on LB agar plates supplemented with 2% SDS and 0.05 mM arabinose.** Truncated TatA and TatB components enable growth on SDS plates except those that result in non-functional translocase i.e. TatA-45 and TatB-130. Left image, clockwise from top: *ΔtatABCDE* expressing TatABC, TatA-20, TatA-30, TatA-40 TatA-45 from pWM-Tat0. Right image, clockwise from top: *ΔtatABCDE* expressing TatB-30, TatB-50, TatB-75, TatB-130, TatB-20 from pWM-Tat0.

It was also possible to carry out anti-TatA blots for these constructs, a feat not possible in the previous section as the truncation of TatA interfered with binding of the TatA antibody (see TatA-45 lane, Figure 34.C). Western blots using antibodies generated to TatA were carried out for fractionation of cultures expressing the TatB truncation plasmids. These are shown in Figure 34 C&D. Here, low levels of TatA are observed in cultures of wild-type cells (MC4100 lane), with large amounts of TatA present in those cultures expressing Tat components from the modified pWM-Tat0 plasmids (ABC, A-45, B-20, B-30, B-50, B-75 and B-130 lanes). To act as negative controls, fractionation of *ΔtatABCDE* cells and TatA-45 expressing cells are loaded. No TatA is blotted in either of these cultures.

Blotting to TatA and the Strep-II tag of TatC give us significant confidence that all three components (TatA, TatB and TatC) are being translated, and help to assure the validity of these results.

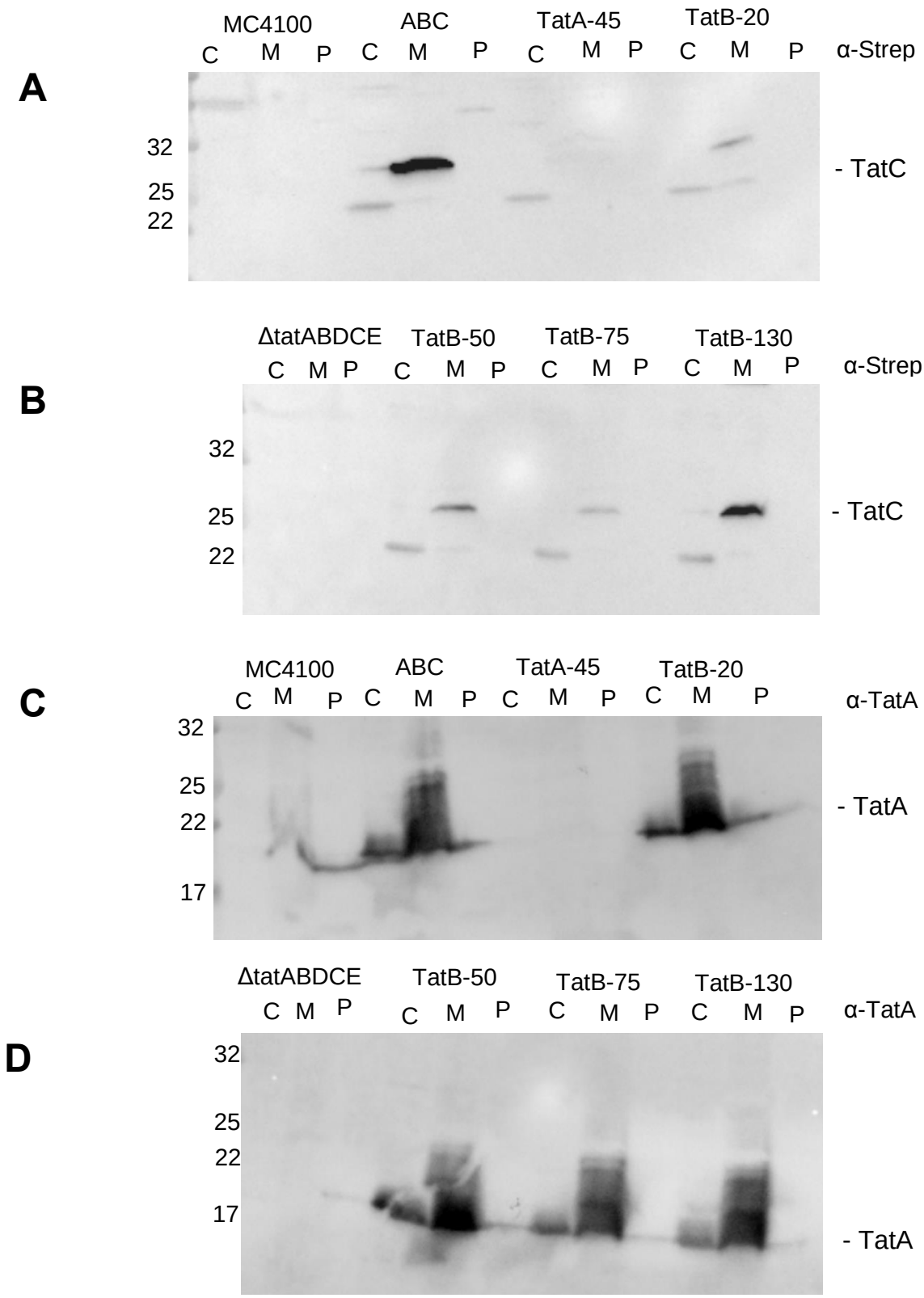


Figure 34: **anti-StrepII (A&B) and anti-TatA (C&D) immunoblot for expression of truncated TatB from the Tat0 plasmid.** Wild-type and truncated TatB variants were expressed in a *tat* null strain as in Figure 30. **A**: wild-type MC4100 cells with no plasmid then *ΔtatABCDE* expressing wild-type ABC, non-functional TatA and TatB missing 20 aas (left to right). **B**: empty *ΔtatABCDE* cells and the same expressing TatB truncated by 50, 75 or 130 aas from the C-terminal. **C** & **D** are the same layout as A and B respectively. The lack of signal in the TatA-45 lane indicates binding of antibody is likely in the C-terminal domain and thus not suitable for use in the TatA truncation studies. TatA is running as a dimer ~ 18 kDa. Molcular weight markers are shown on the left (in kDa) and TatC/TatA is indicated on the right.

Once the expression of truncated TatB was verified we observed the phenotype of *ΔtatABCDE* cells expressing these constructs under the microscope and growth upon 2% SDS plates. Figure 35 shows cultures at 400x magnification. The truncated TatB variants all complement the *tat* null chain-forming phenotype (Figure 35, top left panel), except the non-functional TatB-130. An interesting phenomenon when taking into consideration expression of the TatA truncations still formed cell-chains (Figure 32). The plates in the right-hand panel of Figure 33 show growth of TatB truncations on 2% SDS plates. All truncations allow growth except for the non-functional TatB-130.

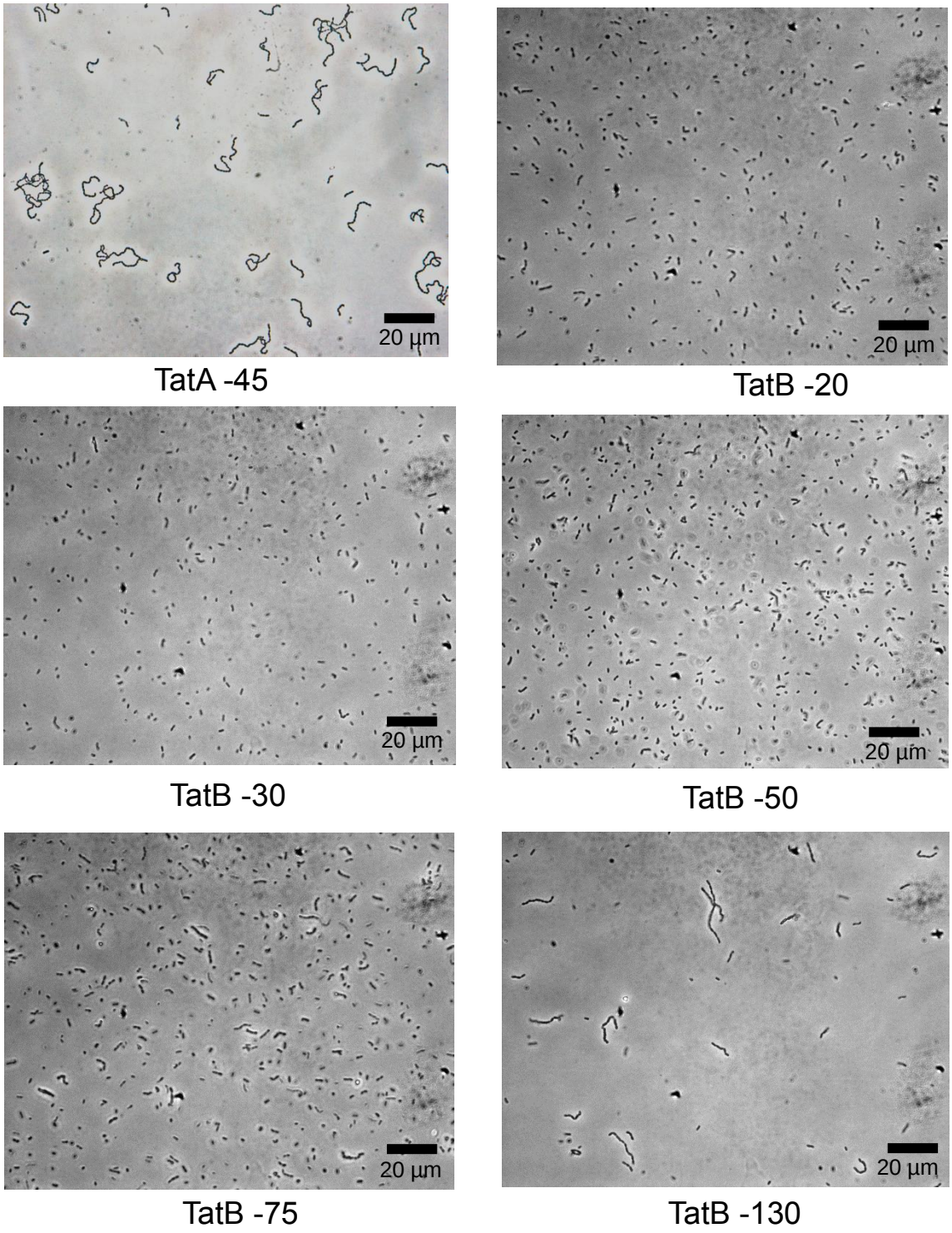


Figure 35: **Truncated TatB complements the filamentous phenotype.** Top: images of *ΔtatABCDE* expressing TatA-45 and TatB-20. TatA-45 encodes a non-functional Tat system and has a filamentous phenotype whereas TatB-20 has the wild-type phenotype of individual cells. Middle: TaB-30 and TatB-50. Bottom: TatB-75 and the non-functional TatB-130. Scale is shown in the bottom right of each image. 400x magnification.

Discussion

Whereas in previous chapters we have utilised substrates targeted to the Tat system in attempts to more accurately identify what is sensed by the proofreading mechanism in order to be rejected for export, here we have attempted to ascertain the region in the Tat components themselves that are responsible for proofreading taking place. Namely in the TatA-like components of *E. coli* TatABC system, TatA and TatB.

Building on previous work by our lab and others that indicated the unstructured tail of these components was not necessary for Tat functionality with native substrates, we thought this region could interact with substrate upon docking or translocation initiation and serve a role in proofreading. Therefore, we constructed multiple versions of an *E. coli* TatABC plasmid where TatA or TatB had blocks of residues removed from the C-terminal. We have shown through a variety of assays these are expressable in *E. coli*, although possibly to a lesser degree than full-length TatABC based on the decrease in signal for anti-StrepII blots.

Expression of full-length TatABC complements a chain-forming phenotype of a *ΔtatABCDE* strain. This is well known. We show however, that upon truncation of TatA or TatB by up to 40 or 130 aas respectively only partial complementation of this phenotype occurs. This could indicate limited Tat functionality as the cell wall amidases AmiA and AmiC, which are native Tat substrates, are not consistently being secreted to the periplasm. An interesting prospect as this could be due to one, or a combination, of two things: 1) truncation of TatA or TatB, while not abolishing Tat export, may reduce functionality. Previous studies have partially taken this into account and this may indeed be the case for truncations in TatB. 2) A loss or interruption of the proofreading mechanism, allowing mis-folded and therefore inactive amidases to be translocated.

Susceptibility to low levels of SDS is a hallmark of the *ΔtatABCDE* strain and Figure 33 shows these cannot grow on SDS supplemented plates. Restoration of growth on these plates is achieved through expression of full-length TatABC. Perhaps in contrast to complementation of the filamentous phenotype, truncation of TatA or TatB also enables growth on 2% SDS supplemented agar plates, at least until large sections from the C-terminal have been removed (40 aas and 130 aas for TatA and TatB respectively). This could indicate that while the Tat machinery is not functioning at 100% (as they still have a filamentous phenotype), it is still enough to stabilise the membrane, rendering cells more resistant to detergent.

Here we have laid the groundwork for attributing a functionality to the largely unstructured C-terminal domains of the TatA and TatB components of the *E. coli* TatABC machinery. While the findings presented in this chapter are interesting and hint at a role in either proofreading or complex stability for the C-terminal domain of these smaller Tat subunits, unfortunately time-constraints meant it was not possible to further asses these variants with more dynamic, folding deficient substrates. Further lines of enquiry would express these more dynamic substrates (e.g. reduced scFvM or BT6h0) alongside the truncated Tat variants and determine if export has been restored. Radioactive labelling and pulse-chase analysis of wild-type and truncated TatABC components would also be undertaken to identify a connection between stability and turnover of the Tat components and the C-terminal domain.

Final Discussion.

While lacking post translational modifications, *E. coli* encode a multitude of complex proteins. Some of these require folding in the cytoplasm, despite the periplasm being their site of function. To localise these folded proteins a secondary translocation pathway evolved, the Twin-Arginine translocase. The majority of Tat substrates are cofactor-containing proteins and several studies have shown the Tat system utilises a quality control (QC) and proofreading system to ensure only *correctly* assembled and folded proteins are translocated. Several studies have indicated the QC system takes place prior to interaction with the Tat machinery, particularly for those substrates that have specific chaperones. However, the exact mechanism of this QC process remains unclear. There is evidence to support an additional Tat-mediated proofreading mechanism, carried out at the translocase itself. The intricacies of how this proofreading mechanism functions are not well understood and the information presented in this study aims to give a clearer view as to what exactly Tat-mediated proofreading 'senses' in a substrate. We have also begun attempts to elucidate the domain(s) in the individual Tat components that perform the proofreading function. Here in this final chapter the major findings in this study will be summarised and future perspectives discussed.

The Tat proofreading mechanism is tolerant of surface charge and hydrophobicity but highly dependent on substrate flexibility.

The first chapter presented in this study identified important factors recognised by the Tat proofreading mechanism and gave a clearer view of what exactly is 'sensed' in a substrate, to more accurately identify what determines whether the substrate is accepted or rejected. This was achieved by taking an scFv for which we had structural data and generating a range of variants with altered surface charge or hydrophobicity, without

altering the folded or structural state of the substrate. None of these variants exhibited distinctly reduced export efficiency relative to the unmodified scFv (Figure 14), even those with a substantial increase in surface hydrophobicity (Figure 15). This is a perhaps surprising phenomenon when taking into consideration other studies which have suggested surface hydrophobicity may be a means by which the Tat proofreading system rejects substrates for export (Richter *et al* 2007). Conversely, this study went on to briefly assess whether the proofreading system sensed the structural flexibility in a substrate through targeting of the scFv with and without prior disulphide formation (Figure 9). Interestingly, the more structurally dynamic variant (i.e. without prior disulphide formation) was accepted for export ~10-fold less than when the scFv was expressed alongside the CyDisCo system and disulphide bonds were formed prior to translocation. Moreover, a variant with a highly disordered C-terminal 'tail' was completely blocked for export (Figure 16).

These results suggest the Tat proofreading mechanism is geared more towards a sense of the structural flexibility of a substrate rather than the surface charge, at least until a significant concentration is reached. This is despite evidence in the literature that long, unfolded polypeptides can be translocated when given a Tat signal peptide (Cline *et al* 2007, Lindenstraub *et al* 2009, Taubert *et al* 2014). However, these studies were carried out in the garden pea plant (*Pisum sativum*) thylakoid Tat system. Clustal omega (Goujon *et al* 2010, Sievers *et al* 2011, MCWilliam *et al* 2013) alignment of the TatA and TatB thylakoidal components of *P. sativum*, *A. thaliana* and *Z. mays* differ greatly from those of bacterial Tat systems. The TatA components have a much longer N-terminal domain of around 100 residues compared to 12 residues in *E. coli* with shorter ~20 residue C-terminal 'tails' compared to 40 residues in *E. coli* (Supplementary Figure 2). Moreover, a cladogram of the protein sequences indicates *E. coli* TatA and the thylakoid TatA of *P.*

sativum, *A. thaliana* and *Z. mays* share the most distant common ancestor (Supplementary Figure 3) and therefore may employ completely different proofreading mechanisms.

Probing the Tat proofreading mechanism with folding variants of a *de novo* designed protein.

After establishing the Tat proofreading mechanism was more responsive to rejecting a substrate based on structural flexibility, we further tested this through the use of a *de novo* designed protein maquette. Here, the sequential inhibition of cofactor binding which was linked to a decrease in tertiary structure through 1D NMR, resulted in a decrease in export efficiency (Figure 21). With the hemeB-deficient BT6 (BT6h0) almost completely blocked in export. A similar scenario to that for the more dynamic scFvM without prior disulphide formation and scFvM 26tail with a disordered domain.

We further went on to show that BT6h0 could be exported when linked behind the export-compatible BT6h2. This is a perhaps surprising and interesting phenomenon as it implies proofreading may take place at a single point – a checkpoint-like event in the translocation process, instead of continuously throughout. This would mean if a substrate was able to ‘’pass’’ this checkpoint, regardless of its folded state, it is committed to export. Such a scenario would account for the 10% export efficiency observed in scFvM without its disulphide bonds formed and the ability to export the linked BT6 constructs and possibly the export of the long, unfolded polypeptides reported by the Cline group.

Comparison of periplasmic human Growth Hormone translocated via the Sec or Tat system.

Given the presence of a proofreading mechanism, we thought it pertinent to address the folded state of substrates crossing the membrane via the Tat pathway and compare them to the same protein reaching the periplasm via the Sec pathway, which transports in an unfolded state and lacks a proofreading mechanism. Here it was shown that comparable amounts of hGH are present and able to be purified from either pathway. Moreover, this protein is folded to the same degree, as evidenced by the comparable ^{1}H NMR spectra (Figure 29 and Figure 30). Thus the Tat system is able to produce a product to at least the same quality as those currently available. In light of this it would be interesting to conduct a similar comparison with a more complicated or less-stable protein, such as bi-specifics or certain scFvs or in the presence of the CyDisCo system.

Effect on proofreading for truncation of TatA and TatB.

To round off this research, we wanted to define the domain responsible for carrying out Tat-mediated proofreading. While it was not possible to fully complete this work, the groundwork has been laid for attributing a level of functionality to the unstructured tail domain of TatA and TatB. We assessed expression via two phenotypic effects – growth on LBA plates supplemented with 2% SDS and formation of cell chains, visible under the microscope. Whilst expression of these TatA and TatB truncation variants was confirmed through immunoblotting, the cells demonstrate phenotypes of both the wild-type and *ΔtatABCDE* strains (Figure 32 and Figure 33). The phenotypic observations may be due to a less stable TatA/B/ABC complex or it could be due to impairment of the proofreading mechanism.

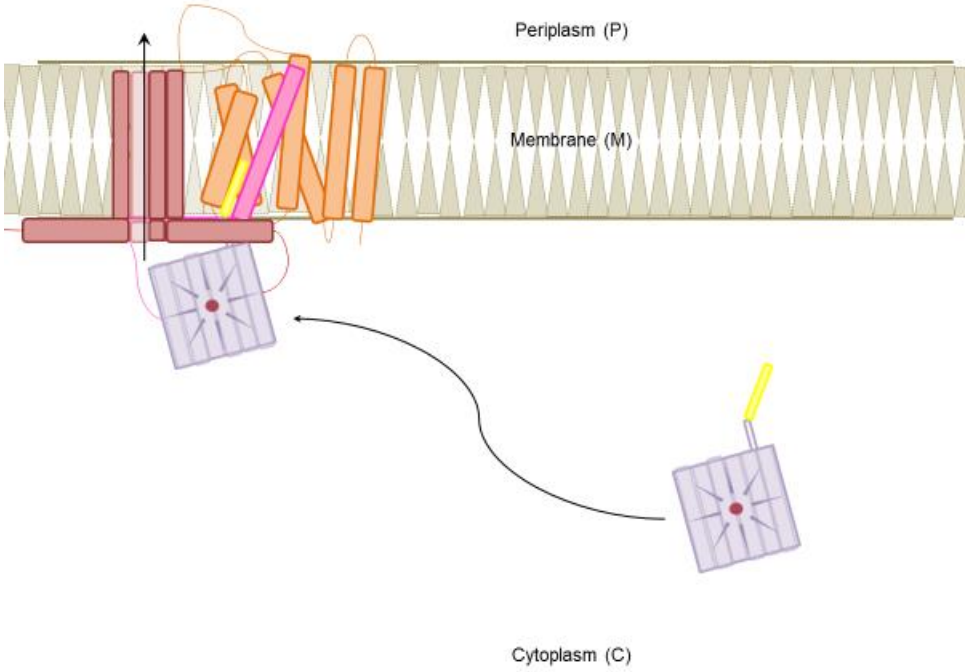
A model for Tat-mediated Quality Control.

Based on the data presented in this study it is possible to propose a simple mechanism by which Tat proofreads substrates. In the translocation process, substrates reach the TatBC substrate binding complex before TatA protomers are recruited to facilitate translocation (section 1.3.4). Although we were unable to identify whether TatA, TatB or TatAB play a role in proofreading and therefore if proofreading takes place before or after TatA nucleation, we have shown that a substrate is committed to transport based on its structural flexibility (chapter 3 and 4). This is a checkpoint-like event and once passed, substrate is committed to export (chapter 4). A summary model of this is shown in Figure 36.

This fits in with published data which suggests proofreading takes place only upon initial binding of substrate, after which a substrate is committed to export (Cline *et al* 2007). In combination with a mutagenesis study where residues were identified in the disordered domains of TatA and TatB and the cytoplasmic loop regions of TatC that possibly contributed to proofreading (Rocco *et al* 2012).

Combining these data, a more definitive proofreading scenario can be postulated. This proceeds as: substrate binds to TatBC and TatA protomers are recruited. At this point the structural flexibility of substrate is sensed through transient interactions with the disordered domains of TatA and TatB (Taubert *et al* 2014, Chapter 6) and then substrate is either committed to translocation or rejected. Therefore, the export efficiency observed for reduced scFvM and BT6h0 is due to those molecules that happen to be in their 'tightest' conformation upon docking to TatBC. A much reduced proportion compared to the more structurally stable scFvM produced in an oxidising cytoplasm or BT6 able to bind heme.

Allowing export of substrates that are slightly flexible – such as scFvM or hGH lacking disulphides (Alanen *et al* 2015) would make sense from an evolutionary standpoint. For example, a recent study has shown that the native Tat substrate CueO, which has copper ligands, is translocated in its apo-state. However, CD data of the mature and apo-forms were superimposable and indicated no change in structure upon Cu binding (Stolle *et al* 2016). Copper ions are toxic to cells as they displace the metal ligands many enzymes require to function. Thus it would not be practical to have mature CueO in the cytoplasm and therefore translocation in the more dynamic apo-form must occur.



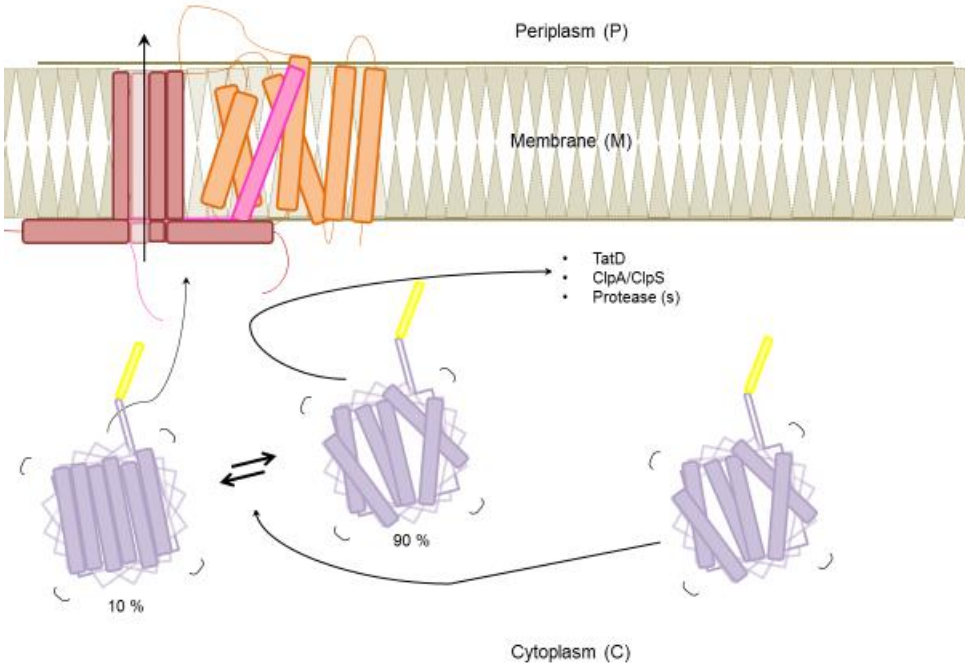


Figure 36: **Schematic representation for a possible mechanism for Proofreading by the TatABC machinery.** In the top panel a correctly folded substrate with bound cofactor/disulphide formation is assessed by the TatABC proofreading process (possibly the C-terminal domains of TatA/TatB), deemed folded and committed to translocation. Whereas in the bottom panel a substrate in a structurally dynamic state (e.g. lacking cofactor or not tightly folded) either happens to be in its 'tightest' conformation (10%) and is committed to translocation, or is in various 'loose' states (90%) and is directed to the degradation pathways.

Future Perspectives.

The data presented in this study has provided an initial insight into the means by which Tat-mediated proofreading broadly determines whether a substrate is accepted or rejected for translocation. It has also shown that simple substrates (i.e. hGH) are structurally identical in the periplasm when translocated by either the Sec or Tat pathway. Whilst the groundwork has been laid to try to identify domains in the Tat components that contribute to proofreading, several questions remains unanswered.

Here we have addressed Tat-mediated quality control and while high-resolution structures of the individual TatA and TatBC complexes are coming to light, a greater understanding of proofreading will come from structural data of these complexes interacting with substrate. These propose a challenge as the Tat components are membrane proteins, a notoriously difficult class of proteins to for structural studies. Expression of the truncated TatA and TatB components alongside a substrate with structural flexibility variants could help determine the domain(s) responsible for Tat-mediated proofreading. However, care must be taken in this line of study as previous reports show overexpression of the Tat components can void the proofreading mechanism and expression to this level is a possibility with a plasmid. An issue that could be countered by modifying genomic *tatABCE*.

A prime approach would be reconstitution of the TatABC system *in vitro* and the subsequent probing with a range of structurally flexible substrates. This would help to define Tat-mediated proofreading while eliminating Quality Control contributors and other components of a no doubt complicated and intertwined process.

From a biotechnological viewpoint, the Tat system is at least as good as the Sec system and further studies would do well to compare the quality of more complex proteins or oligomers.

Together these studies would ascertain the domain(s) responsible and the resolution of Tat-mediated proofreading. The work presented in this thesis provides a basic starting point for these studies and will hopefully provide for a more informed direction for further research to unravel the workings of this translocation system.

References

1. W Wickner, AJM Driessen, FU Hartl, The enzymology of protein translocation across the *Escherichia coli* plasma membrane, Annual Reviews of Biochemistry 60 (1991) 101 – 124.
2. P Natale, T Brüser, AJM Driessen, Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane – Distinct translocases and mechanisms, Biochimica et Biophysica Acta, 1778 (2008) 1735-1756.
3. D Simone, DC Bay, T Leach, RJ Turner, Diversity and evolution of bacterial twin arginine translocation protein, TatC, reveals a protein secretion system that is evolving to fit its environmental niche, PLOS One 8 (2013) 78742.
4. TA Rapoport, B Jungnickel, U Kutay, Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes, Annual Reviews of Biochemistry 65 (1996) 271 – 303.
5. J Yuan, JC Zweers, JM van Dijl, Protein transport across and into cell membranes in bacteria and archaea, Cellular and Molecular Life Sciences 67 (2010) 179 – 199.
6. G Blobel and B Dobberstein, Transfer of Proteins across membranes I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma, the Journal of Cell Biology 67 (1975) 835 – 851.
7. AR Osborne, TA Rapoport, B van den Berg, Protein Translocation by the Sec61/SecY channel, Annual Reviews Cell Developmental Biology 21 (2005) 529 – 550.

8. BC Berks, A common export pathway for proteins binding complex redox cofactors? Molecular Microbiology 22 (1996) 393-404.
9. H Mori, K Cline, A signal peptide that directs non-sec transport in bacteria also directs efficient and exclusive transport on the thylakoid delta pH pathway, The Journal of Biological Chemistry 273 (1998) 11405 – 11408.
10. E Spence, M Sarcina, N Ray, SG Moller, CW Mullneaux, C Robinson, Membrane-specific targeting of green fluorescent protein by the Tat pathway in the cyanobacterium *Synechocystis* PCC6803, Molecular Microbiology 48 (2003) 1482 – 1489.
11. C Aldridge, E Spence, MA Kirkilionis, L Frigerio, C Robinson, Tat-dependent targeting of Rieske iron-sulphur proteins to both the plasma and thylakoid membranes in the cyanobacterium *Synechocystis* PCC6083, Molecular Microbiology 70 (2008) 140 – 150.
12. S Cristóbal, JW de Gier, H Nielsen, G von Heijne, Competition between Sec- and Tat-dependent protein translocation in *Escherichia coli*, The EMBO Journal 18 (1999) 2982 – 2990.
13. M Kipping, H Lilie, U Lindenstrauss, JR Andreesen, C Griesinger, T Carlomango, T Brüser, Structural studies on a twin-arginine signal sequence, FEBS Letters 550 (2003) 18 – 22.
14. F Sargent, E Bogsch, NR Stanley, M Wexler, C Robinson, BC Berks, T Palmer, Overlapping functions of components of a bacterial Sec-independent protein export pathway, EMBO Journal 17 (1998) 3640 – 3650.
15. NR Stanley, T Palmer, BC Berks, The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in *Escherichia coli*, Journal of Biological Chemistry 275 (2000) 11591 – 11596.

16. AM Chaddock, A Mant, I Karnauchov, S Brink, RG Hermann, RB Klösgen, C Robinson, A new type of signal peptide: central role of a twin-arginine motif in transfer signals for the ΔpH-dependent thylakoidal protein translocase, EMBO Journal 14 (1995) 2715 – 2722.
17. D Halbig, T Wiegert, N Blaudeck, R Freudl, GA Sprenger, 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 binding, European Journal of Biochemistry 263 (1999) 543 – 551.
18. G Buchanan, F Sargent, BC Berks, T Palmer, 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 (2001) 107 – 112.
19. MP DeLisa, P Samuelson, T Palmer, G Georgiou, Genetic analysis of the twin arginine translocator secretion pathway in bacteria, Journal of Biological Chemistry 277 (2002) 29825 – 29831.
20. AP Hinsley, NR Stanley, T Palmer, BC Berks, A naturally occurring bacterial Tat signal peptide lacking one of the 'invariant' arginine residues of the consensus targeting motif, FEBS Letters 497 (2001) 45 – 49.
21. BC Berks, T Palmer, F Sargent, The Tat protein translocation pathway and its role in microbial physiology, Advances in Microbial Physiology 47 (2003) 187 – 254.
22. E Bogsch, S Brink, C Robinson, 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 (1997) 3851 – 3859.

23. D Tullaman-Ercek, MP DeLisa, Y Kawarasaki, P Iranpour, B Ribnicky, T Palmer, G Georgiou, Export pathway selectivity of *Escherichia coli* twin arginine translocation signal peptides, The Journal of Biological Chemistry 282 (2007) 8309 – 8316.
24. G von Heijne, Life and Death of a signal peptide, Nature: News and Views 396 (1998) 112 – 113.
25. M Paetzel, A Karla, NCJ Strynadka, RE Dalbey, Signal Peptidases, Chemical Reviews 102 (2002) 4549 – 4579.
26. C Milstein, GG Brwonlee, TM Hrarrison, MB Mathews, A possible precursor of immunoglobulin light chains, Nature News Biology 239 (1972) 117 – 120.
27. H Bedoulle, P Bassford, A Fwoler, I Zabin, J Beckwith, M Hofnung, Mutations which alter the function of the signal sequence of the maltose binding protein of *Escherichia coli*, Nature 285 (1980) 78 – 81.
28. SD Emr, S Hanley-Way, T Silhavy, Suppressor mutations that restore export of a protein with a defective signal sequence, Cell 23 (1981) 79 – 88.
29. AI Derman, JW Puziss, PJ Bassford, J Beckwith, A signal sequence is not required for protein export in *prlA* mutants of *Escherichia coli*, EMBO Journal 12 (1993) 879 – 888.
30. AKJ Veenendaal, C van der Does, AJM Driessen, The protein-conducting channel SecYEG, Biochimica et Biophysica Acta 1694 (2004) 81 – 95.
31. K Ito, PJ Bassford Jr, J Beckwith, Protein localization in *E. coli*: Is there a common step in the secretion of periplasmic and outer-membrane proteins? Cell 24 (1981) 707-717.

32. PJ Schatz, J Beckwith, Genetic analysis of protein export in *Escherichia coli*, Annual Review Genetics 24 (1990) 215 – 248.
33. PN Danese, TJ Silhavy, Targeting and Assembly of periplasmic and outer-membrane proteins in *Escherichia coli*, Annual Review Genetics 32 (1998) 59 – 94.
34. D Golrich, S Prehn, E Hartmann, K Kallies, T Rapoport, A mammalian homologue of Sec61p and SecYp is associated with ribosomes and nascent polypeptides during translocation, Cell 71 (1992) 489 – 503.
35. RJ Deshaies, R Schekman, A yeast mutant defective at an early stage in import of secretory protein precursors into the endoplasmic reticulum, Journal of Cellular Biology 105 (1987) 633 – 645.
36. DP Cerretti, D Dean, GR Davis, DM Bedwell, M Nomura, The *spc* ribosomal protein operon of *Escherichia coli*: sequence and cotranscription of the ribosomal protein genes and a protein export gene, Nucleic Acids Research 11 (1983) 2599 – 2616.
37. Y Akiyama, K Ito, Topology analysis of the SecY protein, an integral membrane protein involved in protein export in *Escherichia coli*, The EMBO Journal 6 (1987) 3465 – 3470.
38. B van den Berg, WM Clemons Jr, I Collinson, Y Modis, E Hartmann, SC Harrison, TA Rapoport, X-ray structure of a protein-conducting channel, Nature 427 (2004) 36 – 44.
39. I Mingarro, I Nilsson, P Whitley, G von Heijne, Different conformations of nascent polypeptides during translocation across the ER membrane, BioMed Central Cell Biology 1 (2000).

40. M Kowarik, S Kung, B Martoglio, A Helenius, Protein folding during cotranslational translocation in the Endoplasmic Reticulum, Molecular Cell 10 (2002) 769 – 778.
41. TV Kurzchalia, M Wiedmann, H Breter, W Zimmermann, E Bauschke, TA Rapoport, tRNA-mediated labelling of proteins with biotin. A nonradioactive method for the detection of cell-free translation products, European Journal of Biochemistry 172 (1988) 663 – 668.
42. K Tani, H Tokuda, S Mizushima, Translocation of ProOmpA possessing an intramolecular disulphide bridge into membrane vesicles of *Escherichia coli* effect of membrane energization, The Journal of Biological Chemistry 265 (1990) 17341 – 17347.
43. P Tian, I Andricioaei, Size, motion, and function of the SecY translocon revealed by molecular dynamics simulations with virtual probes, Biophysical Journal 90 (2006) 2718 – 2730.
44. PJ Schatz, PD Riggs, A Jacq, MJ Fath, J Beckwith, The *secE* gene encodes an integral membrane protein required for protein export in *Escherichia coli*, Genes & Development 3 (1989) 1035 – 1044.
45. PJ Schatz, KL Bieker, KM Ottemann, TJ Silhavy, J Beckwith, One of three transmembrane stretches is sufficient for the functioning of the SecE protein, a membrane component of the *E. coli* secretion machinery, The EMBO Journal 10 (1991) 1749 – 1757.
46. CK Murphy, J Beckwith, Residues essential for the function of SecE, a membrane component of the *Escherichia coli* secretion apparatus, are located in a conserved cytoplasmic region, PNAS 91 (1994) 2557 – 2561.

47. K Nishiyama, S Mizushima, H Tokuda, A novel membrane protein involoved in protein translocation across the cytoplasmic membrane of *Escherichia coli*, The EMBO Journal 12 (1993) 3409 – 3415.
48. K Nishiyama, M Hanada, H Tokuda, Disruption of the gene encoding p12 (SecG) reveals the direct involvement and important function of SecG in the protein translocation of *Escherichia coli* at low temperature, The EMBO Journal 13 (1994) 3272 – 3277.
49. K Nishiyama, T Suzuki, H Tokuda, Inversion of the membrane topology of SecG coupled with a SecA-dependent preprotein translocation, Cell 85 (1996) 71 – 81.
50. AM Flower, Molecular Machines of the Cell The SecY translocation complex: convergence of genetics and structure, TRENDS in Microbiology 15 (2007) 203 – 210.
51. KS Cannon, E Or, WM Clemans Jr, Y Shibata, TA Rapoport, Disulfide bridge formation between SecY and A translocating polypeptide localizes the translocation pore to the center of SecY, The Journal of Cell Biology 169 (2005) 219 – 225.
52. PCK Tam, AP Maillard, KKY Chan, F Duong, Investigating the SecY plug movement at the SecYEG translocation channel, The EMBO Journal 24 (2005) 3380 – 3388.
53. ED Gundelfinger, E Krause, M Melli, B Dobberstein, The organization of the 7SL RNA in the signal recognition particle, Nucleic Acids Research 11 (1983) 7363 – 7374.

54. DM Freymann, RJ Keenan, RM Stroud, P Walter, Structure of the conserved GTPase domain of the signal recognition particle, Letters to Nature 385 (1997) 361 – 364.

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

56. G Montoya, C Svensson, J Luirink, I Sinning, Crystal structure of the NG domain from the signal-recognition particle receptor FtsY, Letters to Nature 385 (1997) 365 – 368.

57. KR Legate, D Falcone, DW Andrews, Nucleotide dependent binding of the GTPase domain of the Signal Recognition Particle Receptor β-subunit to the α-subunit, Journal of Biological Chemistry 275 (2000) 27439 – 27446.

58. JD Miller, H Wilhelm, L Gierasch, R Gilmore, P Walter, GTP binding and hydrolysis by the signal recognition particle during initiation of protein translocation, Letters to Nature 366 (1993) 351 – 354.

59. JD Miller, HD Bernstein, P Walter, Interaction of *E. coli* Ffh/4.5S ribonucleoprotein and FtsY mimics that of mammalian signal recognition particle and its receptor, Letters to Nature 367 (1994) 657 – 659.

60. T Powers, P Walter, Reciprocal stimulation of GTP hydrolysis by two directly interacting GTPases, Science 269 (1995) 1422 – 1424.

61. K Romisch, J Webb, J Herz, S Prehn, R Frank, M Vingron, B Dobberstein, Homology of 54K protein of signal-recognition particle, docking protein and two *E. coli* proteins with putative GTP-binding domains, Letters to Nature 340 (1989) 478 – 482.

62. HD Bernstein, MA Poritz, K Strub, PJ Hoben, S Brenner, P Walter, Model for signal sequence recognition from amino-acid sequence of 54K subunit of signal recognition particle, Letters to Nature 340 (1989) 482 – 486.
63. V Ribes, K Romisch, A Giner, B Dobberstein, D Tollervey, *E. coli* 4.5S RNA is part of a ribonucleoprotein particle that has properties related to signal recognition particle, Cell 63 (1990) 591 – 600.
64. J Macfarlane, M Muller, The functional integration of a polytopic membrane protein of *Escherichia coli* is dependent on the bacterial signal-recognition particle, European Journal of Biochemistry 233 (1995) 766 – 771.
65. SA Ladefoged, G Christiansen, A GTP-binding protein of *Mycoplasma hominin*: a small sized homolog to the signal recognition particle receptor FtsY, Gene 201 (1997) 37 – 44.
66. J Luirink, CM ten Hagen-Jongmanm CC van der Weijden, B Oudega, S High, B Dobberstein, R Kusters, An alternative protein targeting pathway in *Escherichia coli*: studies on the role of FtsY, The EMBO Journal 13 (1994) 2289 – 2296.
67. EH Manting, C van de Does, H Remigy, A Engel, AJM Driessen, SecYEG assembles into a tetramer to form the active protein translocation channel, The EMBO Journal, 19 (2000) 852 – 861.
68. JF Menetret, A Neuhof, DG Morgan, K Plath, M Radermacher, TA Rapoport, CW Akey, The structure of ribosome-channel complexes engaged in protein translocation, Molecular Cell 6 (2000) 1219 – 1232.
69. I Collinson, C Breyton, F Duong, C Tziatzios, D Schubert, E Or, T Rapoport, W Kuhlbrandt, Projection structure and oligomeric properties of a bacterial protein translocase, The EMBO Journal 20 (2001) 2462 – 2471.

70. C Breyton, W Haase, TA Rapoport, W Kuhlbrandt, I Collinson, Three-dimensional structure of the bacterial protein-translocation complex SecYEG, Letters to Nature 418 (2002) 662 – 665.
71. K Mitra, C Schaffitzel, T Shaikh, F Tama, S Jenni, CL Brooks III, N Ban, J Frank, Structure of the *E. coli* protein-conducting channel bound to a translating ribosome, Nature 438 (2005) 318 – 324.
72. E de Leeuw, D Poland, O Mol, I Sinning, CM ten Hagen-Jongman, B Oudega, J Luirink, Membrane association of FtsY, the *E. coli* SRP receptor, FEBS Letters 416 (1997) 225 – 229.
73. B Misselwitz, O Staeck, K Matlack, T Rapoport, Interaction of BiP with the J-domain of the Sec63p component of the endoplasmic reticulum protein translocation complex, Journal of Biological Chemistry 274 (1999) 20110 – 20115.
74. KE Matlack, B Misselwitz, K Plath, T Rapoport, BiP acts as a molecular ratchet during posttranslational transport of prepro-α factor across the ER membrane, Cell 97 (1999) 553 – 564.
75. NG Haigh, AE Johnson, A new role for BiP: closing the aqueous translocon pore during protein integration into the ER membrane, Journal of Cell Biology 156 (2002) 261 – 270.
76. R Lill, K Cunningham, L Brundage, K Ito, D oliver, W Wickner, SecA protein hydrolyses ATP and is an essential component of the protein translocation ATPase of *Escherichia coli*, EMBO Journal 8 (1989) 961 – 966.
77. Y Papanikolau, M Papadovasilaki, RBG Ravelli, AA McCarthy, S Cusack, A Economou, K Petratos, Structure of dimeric SecA, the *Escherichia coli*

preprotein translocase motor, Journal of Molecular Biology 366 (2007) 1545 – 1557.

78. AJM Driessen, SecB, a molecular chaperone with two faces, Trends in Microbiology 9 (2001) 193 – 196.
79. CA Kumamoto, *Escherichia coli* SecB protein associates with exported protein precursors *in vivo*, PNAS 86 (1989) 5320 – 5324.
80. R Lill, W Dowhan, W Wickner, The ATPase activity of SecA is regulated by acidic phospholipids, SecY, and the leader and mature domains of precursor proteins, Cell 60 (1990) 271 – 280.
81. F Hartl, S Lecker, E Schiebel, J Hendrick, W Wickner, The binding cascade of SecB to SecA to SecY/E mediates preprotein targeting to the *E. coli* plasma membrane, Cell 63 (1990) 269 – 279.
82. P Natale, T den Blaauwen, C van der Does, AJM Driessen, Conformational state of the SecYEG-bound SecA probed by single tryptophan fluorescence spectroscopy, Biochemistry 44 (2005) 6424 – 6432.
83. E Schiebel, A Driessen, F Hartl, W Wickner, ΔμH+ and ATP function at different steps of the catalytic cycle of preprotein translocase, Cell 64 (1991) 927 – 939.
84. A Economou, W Wickner, SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion, Cell 78 (1994) 835 – 843.
85. S Matsuyama, Y Fujita, K Sagara, S Mizushima, Overproduction, purification and characterization of SecD and SecF, integral membrane components of the protein translocation machinery of *Escherichia coli*, BBA Protein Structure and Molecular Enzymology 1122 (1992) 77 – 84.

86. F Duong, W Wickner, Distinct catalytic roles of the SecYE, SecG, and SecDFyajC subunits of preprotein translocase holoenzyme, EMBO Journal 16 (1997) 2756 – 2768.
87. JA Pogliano, J Beckwith, SecD and SecF facilitate export in *Escherichia coli*, EMBO Journal 13 (1994) 554 – 561.
88. JW de Gier, J Luirink, The ribosome and YidC. New insights into the biogenesis of *Escherichia coli* inner membrane proteins, EMBO Reports 4 (2003) 939 – 943.
89. A Saaf, M Monne, JW de Gier, G von Heijne, Membrane topology of the 60-kDa Oxa1p homologue from *Escherichia coli*, Journal of Biological Chemistry 273 (1998) 30415 – 30418.
90. E van Bloois, S Nagamori, G Koningstein, RS Ullers, M Preuss, B Oudega, N Harms, HR Kaback, JM Herrmann, J Luirink, The Sec-independent function of *Escherichia coli* YidC is evolutionary-conserved and essential, The Journal of Biological Chemistry 280 (2005) 12996 – 13003.
91. PJ Hynds, D Robinson, C Robinson, The Sec-independent Twin-arginine translocation system can transport both tightly folded and malfolded proteins across the thylakoid membrane, Journal of Biological Chemistry 273 (1998) 34868 – 34874.
92. A Rodrigue, A Chanal, K Becks, M Müller, L Wu, Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial Tat pathway, Journal of Biological Chemistry 274 (1999) 13223 – 13228.
93. L Wu, A Chanal, A Rodrigue, Membrane Targeting and translocation of bacterial hydrogenases, Archives of Microbiology 173 (2000) 319 – 324.

94. RM Mould, C Robinson, A proton gradient is required for the transport of two luminal oxygen-evolving proteins across the thylakoid membrane, Journal of Biological Chemistry 266 (1991) 12189 – 12193.
95. TL Yahr, WT wickner, Functional reconstitution of bacterial Tat translocation in vitro, EMBO Journal 20 (2001) 2472 – 2479.
96. JH Weiner, PT Bilous, GM Shaw, SP Lubitz, L Frost, GH Thomas, JA Cole, RJ Turner, A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins, Cell 93 (1998) 93 – 101.
97. AM Settles, A Yonetani, A Baron, DR Bush, K Cline, R Martienssen, Sec-independent protein translocation by the maize Hcf106 protein, Science 278 (1997) 1467 – 1470.
98. H Mori, K Cline, A twin arginine signal peptide and the pH gradient trigger reversible assembly of the thylakoid ΔpH/Tat translocase, Journal of Cell Biology 157 (2002a) 205 – 210.
99. VJ Goosens, CG Monteferrante, JM van Dijl, The tat system of gram-positive bacteria, BBA Molecular Cell Research (2013).
100. RL Jack, F Sargent, BC Berks, G Sawers, T Palmer, Constitutive expression of the *Escherichia coli tat* genes indicates an important role for the twin-arginine translocase during aerobic and anaerobic growth, Journal of Bacteriology 183 (2001) 1801 – 1804.
101. UA Oschner, A Snyder, AI Vasil, ML Vasil, Effects of the Twin-arginine translocase on secretion of virulence factors, stress response, and pathogenesis, Proceedings for the National Academy of Sciences USA 99 (2002) 8312 – 8317.

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

103. JA McDonough, JR McCann, EM Tekippe, JS Silverman, NW Rigel, M Braunstein, Identification of functional Tat signal sequences in Mycobacterium tuberculosis Proteins, Journal of Bacteriology 190 (2008) 6428 – 6438.

104. E De Buck, E Lammertyn, J Anne, The importance of the twin-arginine translocation pathway for bacterial virulence, Trends in Microbiology 16 (2008) 442 – 453.

105. TH Walther, SI Grage, N Roth, AS Ulrich, Membrane alignment of the pore-forming component TatAd of the twin-arginine translocase from *Bascillus subtilis* resolved by solid-state NMR spectroscopy, Journal of the American Chemical society 132 (2010) 15945 – 15946.

106. S Koch, MJ Fritsch, G Buchanan, T Palmer, *Escherichia coli* TatA and TatB proteins have N-out, C-in topology in intact cells, Journal of Biological Chemistry 287 (2012) 14420 – 14431.

107. Y Hu, E Zhao, H Li, B Xia, C Jin, Solution NMR structure of the TatA component of the twin0arginine protein transport system from gram-positive bacterium *Bascillus subtilis*, Journal of the American Chemical Society 132 (2010) 15942 – 155944.

108. SD Müller, AA De Angelis, TH Walther, SL Grage, C Lange, SJ Opella, AS Ulrich, Structural characterisation of the pore forming protein TatAd of the twin-arginine translocase in membranes by solid-state 15N NMR, BBA Biomembranes 1768 (2007) 3071 – 3079.

109. C Lang, SD Muller, TH Walther, J Burck, AS Ulrich, Structure analysis of the protein translocating channel TatA in membranes using a multi-construct approach, BBA Biomembranes 1768 (2007) 2672 – 2634.

110. CML Barrett, JE Mathers, C Robinson, Identification of key regions within the *Escherichia coli* TatAB subunits, FEBS Letters 537 (2003) 42 – 46.

111. MG Hicks, E de Leeuw, I Porcelli, G Buchanan, BC Berks, T Palmer, The *Escherichia coli* twin-arginine translocase: conserved residues of TatA and TatB family components involved in protein transport, FEBS Letters 539 (2003) 61 – 67.

112. F Rodriguez, SL Rouse, CE Tait, J Harmer, A De Riso, CR Timmel, MSP Samsom, BC Berks, JR Schnell, Structural model for the protein-translocating element of the twin-arginine transport system, PNAS (2013) E1092 – E1101.

113. NP Greene, I Porcelli, G Buchanan, MG Hicks, SM Schermann, T Palmer, BC Berks, Cysteine scanning mutagenesis and disulphide mapping studies of the TatA component of the bacterial twin arginine translocase, Journal of Biological Chemistry 282 (2007) 23937 – 23945.

114. MG Hicks, PA Lee, G Georgiou, BC Berks, T Palmer, Positive selection for loss-of-function *tat* mutations identifies critical residues required for TatA activity, Journal of Bacteriology 187 (2005) 2920 – 2925.

115. Y Zhang, L Wang, Y Hu, C Jin, Solution structure of the TatB component of the twin-arginine translocation system, BBA Biomembranes 1838 (2014ba) 1881 – 1888.

116. MR Yen, YH Tseng, EH Nyugen, LF Wu, MH Saier Jr, Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system, Archives of Microbiology 177 (2002) 441 – 450.

117. F Sargent, NR Stanley, BC Berks, T Palmer, Sec-independent protein translocation on Escherichia coli. A distinct and pivotal role for the TaB protein, Journal of Biological Chemistry 51 (1999) 36073 – 36082.

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

119. N Bludeck, P Kreutzenbeck, M Muller, GA Sprenger, R Freudl, Isolation and characterization of bifunctional *Escherichia coli* TatA mutant preoteins that allow efficient Tat-dependent protein translocation in the absence of TatB, Journal of Biological Chemistry 280 (2005) 3426 – 3432.

120. J Behrendt, K Standar, U Lindenstrauß, T Bruser, Topological studies on the twin-arginine translocase component TatC, FEMS Microbiology Letters 234 (2004) 303 – 308.

121. C Punginelli, B Maldonado, S Grahl, R Jack, M Alami, J Schroder, BC Berks, T Palmer, Cysteine scanning mutagenesis and topological mapping of the *Escherichia coli* twin-arginine translocase TatC component, Journal of Bacteriology 189 (2007) 5482 – 5494.

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

123. X Ma, K Cline, Mapping the signal peptide binding and oligomer contact sites of the core subunit of the pea twin arginine protein translocase, Plant Cell 25 (2013) 999 – 1015.

124. SE Rollauer, MJ Tarry, JE Graham, M Jaaskelainen, F Jager, S Johnson, M Krehenbrink, SM Liu, MJ Lukey, J Marcoux, MA McDowell, F Rodriguez, P Roversi, PJ Stansfeld, CV Robinson, MSP Sansom, T Palmer, M Hogbom, BC Berks, S M Lea, Structure of the TatC core of the twin-arginine protein transport system, Nature 492 (2012) 210 – 214.

125. S Ramasamy, R Abrol, CJ Suloway, WM Clemons Jr, The glove-like structure of the conserved membrane protein TatC provides insight into the signal sequence recognition in twin-arginine translocation, Structure 21 (2013) 777 – 788.

126. G Buchanan, F de Leeuw, NR Stanley, M Wexler, BC Berks, F Sargent, T Palmer, Functional complexity of the twin-arginine translocase TatC component revealed by site-directed mutagenesis, Molecula Microbiology 43 (2002) 1457 – 1470.

127. F Sargent, EG bogsch, NR Stanley, M Wexler, C Robinson, BC Berks, T Palmer, Overlapping functions of components of a bacterial Sec-independent protein export pathway, EMBO Journal 17 (1998) 3640-3650.

128. E Elemr, J Fröbel, AS Blümmel, M Müller, TatE as a regular constituent of bacterial twin-arginine protein translocases, The Journal of Biological Chemistry 290 (2015) 29281 – 29289.

129. K Dilks, RW Rose, E Hartmann, M Pohlschroder, Prokaryotic utilization of the twin-arginine translocation pathway: a genetic survey, Journal of Bacteriology 185 (2003) 1478 – 1483.

130. H Tjalsma, H Antelmann, JDH Jongbloed, PG Braun, E Darmon, R Dorenbos, JYF Dubois, H Westers, G Zanen, WJ Quax, OP Kuipers, S Bron, M Hecker, JM van Dijl, Proteomics of protein secretion by *Bascillus subtilis*:

separating the ‘’secrets’’ of the secretome, Microbiology and Molecular Biology Reviews 68 (2004) 207 – 233.

131. T Palmer, F Sargent, BC Berks, The Tat protein export pathway, *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, American Society for Microbiology Press (2010).
132. JDH Jongbloed, U Grieger, H Antelmann, M Hecker, R Nijlan, S Bron, JM van Dijl, Two minimal Tat translocases in *Bascillus*, Molecular Microbiology, 54 (2004) 1319 – 1325.
133. JDH Jongbloed, R van der Ploeg, JM van Dijl, Bifunctional TatA subunits in minimal Tat protein translocases, Trends in Microbiology 14 (2006) 2 – 4.
134. Y Zhang, Y Hu, H Li, C Jin, Structural basis for TatA oligomerisation: an NMR study of *Escherichia coli* TatA dimeric structure, PLOS Ons 9 (2014b) c103157.
135. JP Barnette, RT Eijlander, OP Kuipers, C Robinson, A minimal Tat system from gram-positive organism: A bifunctional TatA subunit participates in discrete TatAC and TatA complexes, Journal of Biological Chemistry 283 (2008) 2534 – 2542.
136. BC Berks, F Sargent, T Palmer, The Tat protein export pathway, Molecular Microbiology 35 (2000) 260 – 274.
137. T Brüser, T Yano, DC Brune, F Daldal, Membrane targeting of a folded and cofactor-containing protein, European Journal of Biochemistry 270 (2003) 1211 – 1221.
138. RL Jack, G Buchanan, A Dubini, K Hatzixanthis, T Palmer, F Sargent, Coordinating assembly and export of complex bacterial proteins, EMBO Journal 23 (2004) 3962 – 3972.

139. A Bolhuis, JE Mathers, JD Thomas, CM Barrett, C Robinson, TatB and TatC form a functional and structural unit of the twin-arginine translocase from *Escherichia coli*, Journal of Biological Chemistry 276 (2001) 1135 – 1146.

140. F Lausberg, S Fleckenstein, P Kreutzenbeck, J Frobel, P Rose, M Muller, R Freudl, Genetic evidence for a tight cooperation of TatB and TatC during productive recognition of twin-arginine (Tat) signal peptides in *Escherichia coli*, PLOS One (2012) e39867.

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

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

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

144. BC Berks, SM Lea, PJ Stansfeld, Structural biology of Tat protein transport, Current Opinion in Structural Biology 27 (2014) 32 – 37.

145. AS Blummel, LA Haag, E Eimer, M Muller, J Frobel, Initial assembly steps of a translocase for folded proteins, Nature Communications 6 (2014).

146. C Aldridge, X Ma, F Gerard, K Cline, Substrate-gated docking of pore subunit Tha4 in the TatC cavity initiates Tat translocase assembly, Journal of Cellular Biology 205 (2014) 51 – 65.

147. C Maurer, S Panahandeh, AC Jungkamp, M Moser, M Muller, TatB functions as an oligomeric binding site for folded Tat precursor proteins, Molecular Biology Cell 21 (2010) 4151 – 4161.

148. MJ Tarry, E Schafer, S Chen, G Buchanan, NP Greene, SM Lea, T Palmer, HR Saibil, BC Berks, Structural analysis of substrate binding by the TatBC component of the twin-arginine protein transport system, PNAS 106 (2009) 13284 – 13289.

149. JM Celedon, K Cline, Stoichiometry for binding and transport by the twin arginine translocation system, Journal of Cellular Biology 197 (2012) 523 – 534.

150. NN Alder, SM Theg, Protein transport via the cpTat pathway displays cooperativity and is stimulated by transport-incompetent substrate, FEBS Letters 540 (2003a) 96 – 100.

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

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

153. GL Orriss, MJ Tarry, B Ize, F Sargent, SM Lea, T Palmer, BC Berks, TatBC, TatB and TatC form structurally autonomous units within the twin arginine transport system of *Escherichia coli*, FEBS Letters 581 (2007) 4091 – 4097.

154. S De Keersmaeker, LV Mellaert, E Lammertyn, K Vrancken, J Anne, N Geukens, Functional analysis of TatA and TatB in *Streptomyces lividans*, Biochemistry Biophysical Research Communications 335 (2005) 973 – 982.

155. M Westermann, OI Pop, R Gerlach, TR Appel, W Schlormann, S Schreiber, JP Muller, The TatAd component of the *Bascillus subtitlis* twin-arginine protein transport system forms homo-multimeric complexes in its cytosolic and membrane embedded localisations, BBA Biomembranes 1758 (2006) 443 – 451.

156. S Frielingsdorf, M Jakob, RB Klosgen, A stromal pool of TatA promotes Tat-dependent protein transport across the thylakoid membrane, Journal of Biological Chemistry 283 (2008) 33838 – 33845.

157. J Oates, CML Barrett, JP Barnett, KG Bryne, A Bolhuis, C Robinson, 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 (2005) 295 – 305.

158. MC Leake, NP Greene, RM Godun, T Granjon, G Buchanan, S Chen, RM Berry, T Pamer, BC Berks, Variable stoichiometry of the TatA component of the twin-arginine protein transport system observed by in vivo single-molecule imaging, PNAS 105 (2008) 15376 – 15381.

159. C Dabney-Smith, K Cline, Clustering of C-terminal stromal domains of Tha4 homo-oligomers during translocation by the Tat protein transport system, Molecular Biology of the Cell 20 (2009) 2060 – 2069.

160. GF White, SM Schermann, J Bradley, A Roberts, NP Greene, BC Berks, AJ Thomson, Subunit organization in the TatA complex of the twin arginine protein translocase: a site-directed EPR spin labelling study, Journal of Biological Chemistry 285 (2010) 2294 – 22301.

161. BC Berks, The twin-arginine protein translocation pathway, Annual Reviews Biochemistry 84 (2015) 843 – 864.

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

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

164. J Frobel, P Rose, F Lausberg, AS Blummel, R Freudl, M Muller, Transmembrane insertion of twin-arginine signal peptides is driven by TatC and regulated by TatB, Nature Communications (2012) 1311.

165. K Gouffi, F Gerard, CL Santini, LF Wu, Dual topology of the *Escherichia coli* TatA protein, Journal of Biological Chemistry 279 (2004) 11608 – 11615.

166. U Gohlke, L Pullman, CA McDevitt, I Porcelli, F de Leeuw, T Palmer, HR Saibil, BC Berks, The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter, PNAS 102 (2005) 10482 – 10486.

167. TH Walther, C Gottselig, SL Grage, M Wolf, AV Vargiu, MJ Klein, S Vollmer, S Prock, M Hartmann, S Afonin, E Stockwald, H Heinzmann, OV Nolandt, W Wenzel, P Ruggerone, AS Ulrich, Folding and self-assembly of the TatA translocation pore based on a charge zipper mechanism, Cell 152 (2013) 316 – 326.

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

169. D Beck, N Vasisht, J Baglieri, CG Monteferrante, JM van Dijl, C Robinson, CJ Smith, Ultrastructural characterisation of *Bascillus subtilis* TatA complexes suggests they are too small to form homooligomeric translocation pores, BBA Molecular Cell Research 1833 (2013) 1811 – 1819.

170. C Aldridge, A Storm, K Cline, C Dabney-Smith, The chloroplast twin arginine transport (Tat) component, Tha4, undergoes conformational changes leading to Tat protein transport, Journal of Biological Chemistry 287 (2012) 34752 – 34763.

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

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

173. J Taubert and T Bruser, Twin-arginine translocation-arresting protein regions contact TatA and TatB, Biological Chemistry 395 (2014) 827 – 836.

174. RM Mould and C Robinson, A proton gradient is required for the transport of two luminal oxygen-evolving proteins across the thylakoid membrane, The Journal of Biological Chemistry 19 (1991) 12189 – 12193.

175. SM Musser and SM Theg, Proton transfer limits protein translocation rate by the thylakoid ΔpH/Tat machinery, Biohcemistry 39 (2000) 8228 – 8233.

176. NN Alder, SM Theg, Energetics of protein transport across biological membranes: a study of the thylakoid ΔpH-dependent/cpTat pathway, Cell 112 (2003b) 231 – 242.

177. M Alami, D Trescher, LF Wu, M Muller, Separate analysis of Twin-Arginine Translocation (Tat)-specific membrane binding and translocation in

Escherichia coli, The Journal of Biological Chemistry 277 (2002) 20499 – 20503.

178. R Patel, SM Smith, C Robinson, Protein transport by the bacterial Tat pathway, BBA Molecular Cell Research 1843 (2014) 1620 – 1628.
179. MP DeLisa, D Tullman, G Georgiou, Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway, PNAS 100 (2003) 6115 – 6120.
180. CFRO Matos, C Robinson, A Di Cola, The Tat system proofreads FeS protein substrates and directly initiates the disposal of rejected molecules, EMBO Journal, 27 (2008) 2055 – 2063.
181. N Ray, J Oates, RJ Turner, C Robinson, DmsD is required for the biogenesis of DMSO reductase in *Escherichia coli* but not for the interaction of the DmsA signal peptide with the Tat apparatus, FEBS Letters 534 (2003) 156 – 160.
182. H Li, L Chang, JM Howell, RJ Turner, DmsD, a Tat specific chaperone, interacts with other general chaperones and proteins involved in the molybdenum cofactor biosynthesis, BBA Proteins and Proteomics 1804 (2010) 1301 – 1309.
183. CS Chan, X Song, SJ Qazi, D Setiaputra, CK Yip, TC Chao, RJ Turner, Unusual pairing between assistants: interaction of the twin-arginine system-sprcific chaperone DmsD with the chaperonin GroEL, Biochemical and Biophysical Research Communications 456 (2015) 841 – 846.
184. JM Dow, F Gabel, F Sargent, T Palmer, Characterisation of a pre-export enzyme-chaperone complex on the twin-arginine transport pathway, Biochemical Journal 452 (2013) 57 – 66.

185. D Guymer, J Maillard, MF Agacan, CA Brearley, F Sargent, Intrinsic GTPase activity of a bacterial twin-arginine translocation proofreading chaperone induced by domain swapping, The FEBS Journal 227 (2010) 511 – 525.

186. LC Potter, JA Cole, Essential roles for the products of the *napABCD* genes, but not *napFGH*, in periplasmic nitrate reduction by *Escherichia coli* K-12, Biochemical Journal 344 (1999) 69 – 76.

187. CS Chan, L Chang, KL Rommens, RJ Turner, Differential interactions between Tat-specific redox enzyme peptides and their chaperones, Journal of Bacteriology 191 (2009) 2091 – 2101.

188. C Maurer, S Panahandeh, M Moser, M Müller, Impairment of twin-arginine-dependent export by seemingly small alterations of substrate conformation, FEBS Letters 583 (2009) 2849 – 2853.

189. C Robinson, CFRO Matos, D Beck, C Ren, J Lawrence, N Vasisht, S Mendel, Transport and proofreading of proteins by the twin-arginine translocation (Tat) system in bacteria, BBA Biomembranes 1808 (2011) 876 – 884.

190. VJ Goosens, CG Monteferrante, JM van Dijl, Co-factor insertion and disulphide bond requirements for the Twin-arginine translocase-dependent export of the *Bascillus subtilis* rieske protein QcrA, The Journal of Biological Chemistry 289 (2014) 13124 – 13131.

191. HI Alanen, Kl Walker, MLV Suberbie, CFRO Matos, S Bonisch, RB Freedman, E Keshavarz-Moore, LW Ruddock, C Robinson, 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

disulphide bond formation, BBA Molecular Cell Research 1853 (2015) 756 – 763.

192. TA Bewley, J Brovetto-Cruz, CH Li, Human pituitary growth hormone. Physiochemical investigations of the native and reduced-alkylated protein, Biochemistry 8 (1969) 4701 – 4708.
193. KM Youngman, DB Spencer, DN Brems, MR DeFilippis, Kinetic analysis of the folding of human growth hormone, Journal of Biological Chemistry 270 (1995) 19816 – 19822.
194. P Stolle, B Hou, T Brüser, The Tat substrate CueO is transported in an incomplete folding state, Journal of Biological Chemistry 291 (2016) 13520 – 13528.
195. S Richter, U Lindenstrauss, C Luke, R Bayliss, T Bruser, Functional Tat transport of unstructured, small, hydrophilic proteins, Journal of Biological Chemistry 282 (2007) 33257 – 33264.
196. K Cline and M McCaffery, Evidence for a dynamic and transient pathway through the Tat protein transport machinery, EMBO Journal 26 (2007) 3039 – 3049.
197. RA Roffey, SM Theg, Analysis of the import of carboxyl-terminal truncations of the 23-kilodalton subunit of the Oxygen-Evolving complex suggests that its structure is an important determinant for thylakoid transport, Plant Physiology 111 (1996) 1329 – 1338.
198. MA Rocco, D Waraho-Zhmayev, MP DeLisa, Twin-arginine translocase mutations that suppress folding quality control and permit export of misfolded substrate proteins, PNAS 109 (2012) 13392 – 13397.

199. LL Randall, SJS Hardy, 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 (1986) 921 – 928.

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

201. JI Austerberry, R Dajani, S Panova, D Roberts, AP Golovanov, A Pluen, CG van der Walle, S Uddin, J Warwicker, JP Derrick, R Curtis, The effect of charge mutations on the stability and aggregation of a human single chain Fv fragment, European Journal of Phamaceutics and Biopharmaceutics 115 (2017) 18 – 30.

202. F Hatahet, VD Nguyen, KEH Salo, LW Ruddock, Disruption of reducing pathways is not essential for efficient disulphide bond formation in the cytoplasm of *E. coli*, Microbial Cell Factories 9 (2010).

203. VD Nguyen, F Hatahet, KEH Salo, E Enlund, C Zhang, LW Ruddock, Pre-expression of a sulfydryl oxidase significantly increases the yields of eukaryotic disulphide bond containing proteins expressed in the cytoplasm of *E. coli*, Microbial Cell Factories 10 (2011).

204. BR Miller, SJ Demarest, A Lugovskoy, F Huang, X Wu, WB Snyder, LJ Croner, N Wang, A Amatucci, JS Michaelson, SM Glaser, Stability engineering of scFvs for the development of bispecific and multivalent antibodies, Protein Engineering, Design & Selection 23 (2010) 549 – 557.

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

206. JD Thomas, RA Daniel, J Errington, C Robinson, Export of active green fluorescent protein to the periplasm by the twin-arginine translocase (Tat) pathway in *Escherichia coli*, Molecular Microbiology 39 (2001) 47 – 53.

207. MS Lawrence, KJ Phillips, DR Liu, Supercharging proteins can impart unusual resilience, JACS Communications 129 (2007) 10110 – 10112.

208. LA Kelley, S Mezulis, CM Yates, MN Wass, MJE Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis, Nature Protocols 10 (2015) 845 – 858.

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

210. M Zavodszky, CW Chen, JK Huang, M Zolkiewski, L Wen, R Krishnamoorthi, Disulfide bond effects on protein stability: Designed variants of *Cucurbita maxima* trypsin inhibitor-V, Protein Science 10 (2001) 149 – 160.

211. CFRO Matos, HI Alanen, P Prus, Y Uchida, LW Ruddock, RB Freedman, E Keshavarz-Moore, C Robinson, Efficient export of prefolded, disulphide-bonded recombinant proteins to the periplasm by the Tat pathway is *Escherichia coli* CyDisCo strains, Biotechnology Progress 30 (2014) 281 – 290.

212. WF DeGrado, ZR Wasserman, JD Lear, Protein Design, a minimalist approach, Science 243 (1989) 622 – 628.

213. GM Edelman, JA Gally, Degeneracy and complexity in biological systems, Proceedings of the National Academy of Sciences of the United States of America, 98 (2001) 13763 – 13768.

214. JM Shifman, CC Moser, WA Kalsbeck, DF Bocian, PL Dutton, Functionalized *de novo* designed proteins: Mechanism of proton coupling to Oxidation/Reduction in heme protein maquettes, Biochemistry 37 (1998) 16815 – 16827.

215. AM Grosset, BR Gibney, F Rabanal, CC Moser, PL Dutton, Proof of Principle in a *de novo* designed protein maquette: An allosterically regulated, charge-activated conformational switch in a tetra-α-helix bundle, Biochemistry 40 (2001) 5474 – 5487.

216. BR Lichtenstein, TA Farid, G Kodali, LA Solomon, JLR Anderson, MM Sheehan, NM Ennist, BA Fry, SE Chobot, C Bialas, JA Mancini, CT Armstrong, Z Zhao, TV Esipova, D Snell, SA Vinogradov, BM Discher, CC Moser, PL Dutton, Engineering oxidoreductases: maquette proteins designed from scratch, Biochemical Society Transactions 40 (2012) 561 – 566.

217. TA Farid, G Kodali, LA Solomon, BR Lichenstein, MM Sheehan, BA Fry, C Bialas, NM Ennist, JA Siedlecki, Z Zhao, MA Stetz, KG Valentine, JLR Anderson, AJ Wand, BM Discher, CC Moser, PL Dutton, Elemenntary tetrahelical protein design for diverse oxidoreductase functions, Nature Chemical Biology 9 (2013) 826 – 834.

218. DM Dykxhoorn, R St Pierre, T Linn, A set of compatible tac promoter expression vectors, Gene 177 (1996) 133 – 136.

219. U Lindenstrauß, CFRO Matos, W Graubner, C Robinson, T Brüser, Malfolded recombinant Tat substrates are Tat-independently degraded in *Escherichia coli*, FEBS Letters 584 (2010) 3644 – 3648.

220. G Walsh, Biopharmaceutical benchmarks 2014, Nature Biotechnology 32 (2014) 992 – 1000.

221. G Gerogiou and L Segatori, Preparative expression of secreted proteins in bacteria: status report and future prospects, Current Opinion in Biotechnology 16 (2005) 538 – 545.

222. JJ Pierce, C Turner, E Keshavarz-Moore, P Dunnill, Factors determining more efficient large-scale release of a periplasmic enzyme from *E. coli* using lysozyme, Journal of Biotechnology 58 (1997) 1 – 11.

223. B Balasundaram, S Harrison, DG Bracewell, Advances in product release strategies and impact on bioprocess design, Trends in Biotechnology 27 (2009) 477 – 485.

224. AH Kwan, M Mobli, PR Gooley, GF King, JP Mackay, Macromolecular NMR spectroscopy for the non-spectroscopist, the FEBS Journal 278 (2011) 687 – 703.

225. GR Fulmer, AJM Miller, NH Sherden, HE Gottlieb, A Nudelman, BM Stoltz, JE Bercaw, KI Goldberg, NMR chemical shifts of trace impurities: Common laboratory solvents, organics, and gases in deuterated solvents relevant to the organometallic chemist, Organometallics 29 (2010) 2176 – 2179.

226. PA Lee, G Buchanan, NR Stanley, BC Berks, T Palmer, Truncation analysis of TatA and TatB defines the minimal functional units required for protein translocation, Journal of Bacteriology 184 (2002) 5871 – 5879.

227. G Warren, J Oates, C Robinson, AM Dixon, Contributions of the transmembrane domain and a key acidic motif to assembly and function of the TatA complex, Journal of Molecular Biology 388 (2009) 122 – 132.

228. B Maldonado, H Kneuper, G Buchanan, K Hatzixanthis, F Sargent, BC Berks, T Palmer, Characterisation of the membrane-extrinsic domain of the

TatB component of the twin arginine protein translocase, FEBS Letters 585 (2011) 478 – 484.

229. NR Stanley, K Findlay, BC Berks, T Palmer, *Escherichia coli* strains blocked in Tat-dependent protein export exhibit pleiotropic defects in the cell envelope, Journal of Bacteriology 183 (2001) 139 – 144.

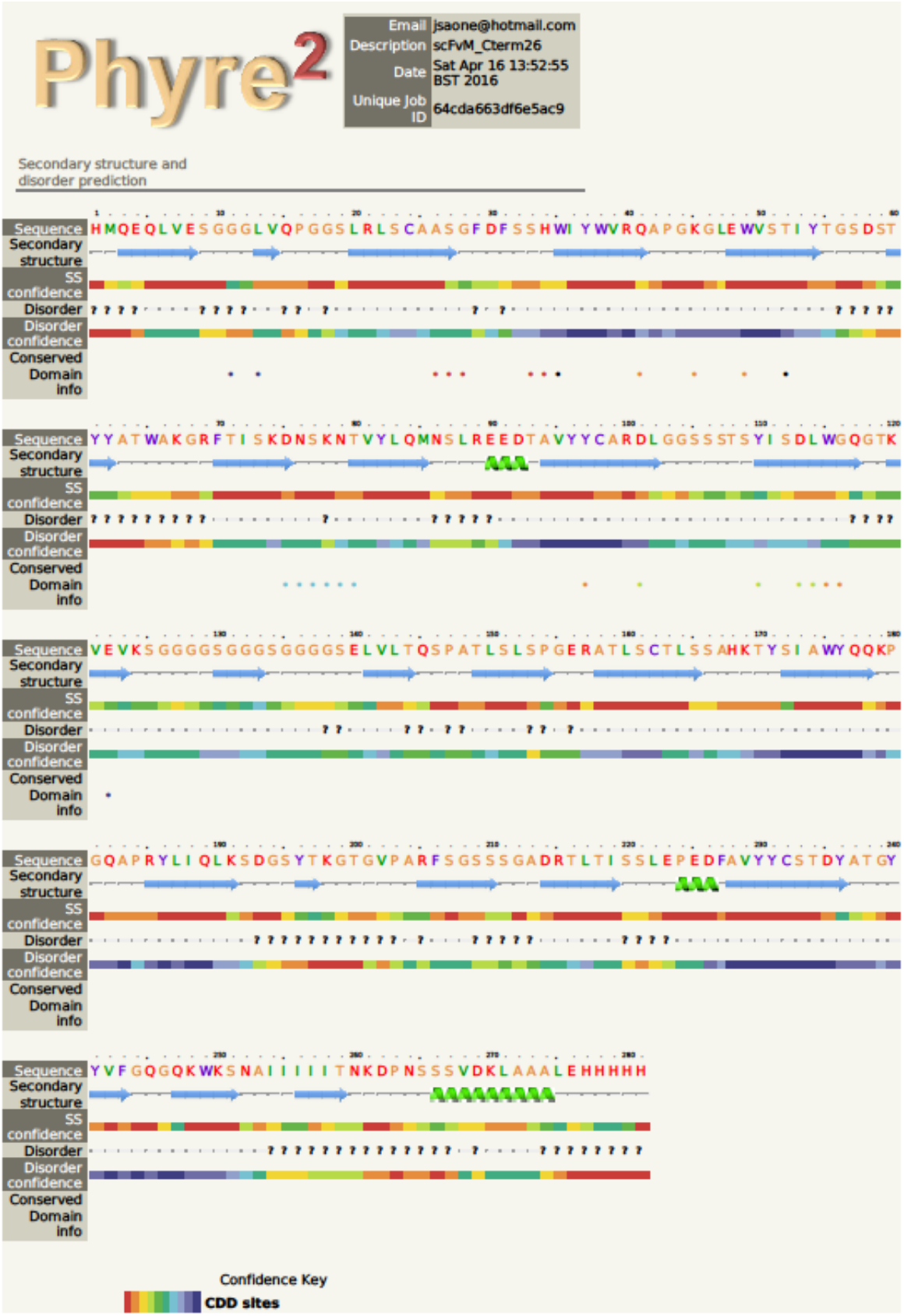
230. B Ize, NR Stanley, G Buchanan, T Palmer, Role of the *Escherichia coli* Tat pathway in outer membrane integrity, Molecular Microbiology 48 (2003) 1183 – 1193.

231. M Goujon, H McWilliam, W Li, F Valentin, S Squizzato, J Paern, R Lopez, A new bioinformatics analysis tools framework at EMBL-EBI, Nucleic Acids Research 38 (2010) W695 – W699.

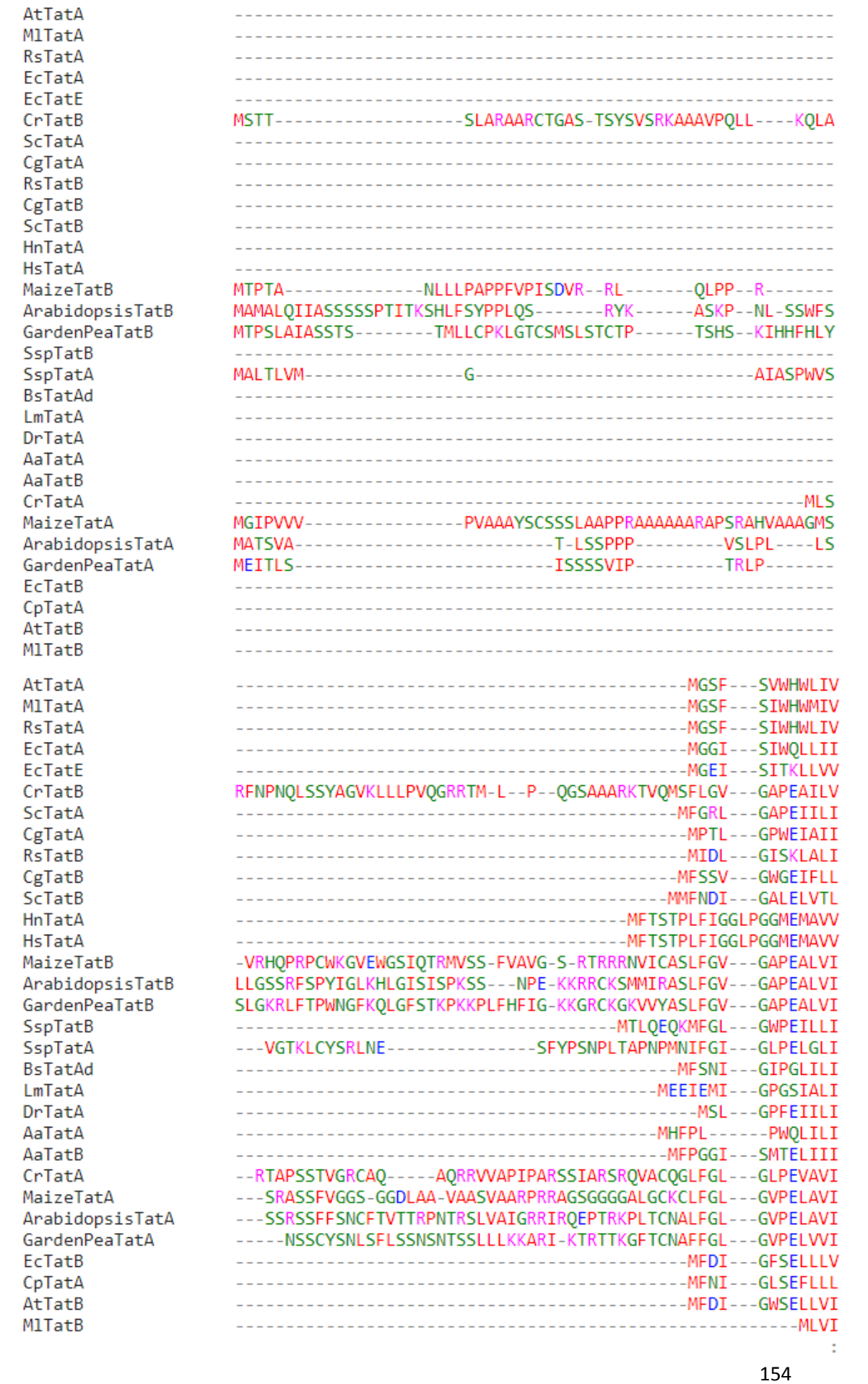
232. F Sievers, A Wilm, D Dineen, TJ Gibson, K Karplus, W Li, R Lopez, H McWilliam, M Remmert, J Söding, JD Thompson, DG Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Molecular Systems Biology 7 (2011) 539.

233. H McWilliam, W Li, M Uludag, S Squizzato, YM Park, N Buso, AP Cowley, R Lopez, Analysis tool web services from the EMBL-EBI, Nucleic Acids Research 41 (2013) W597 – W600.

Supplementary

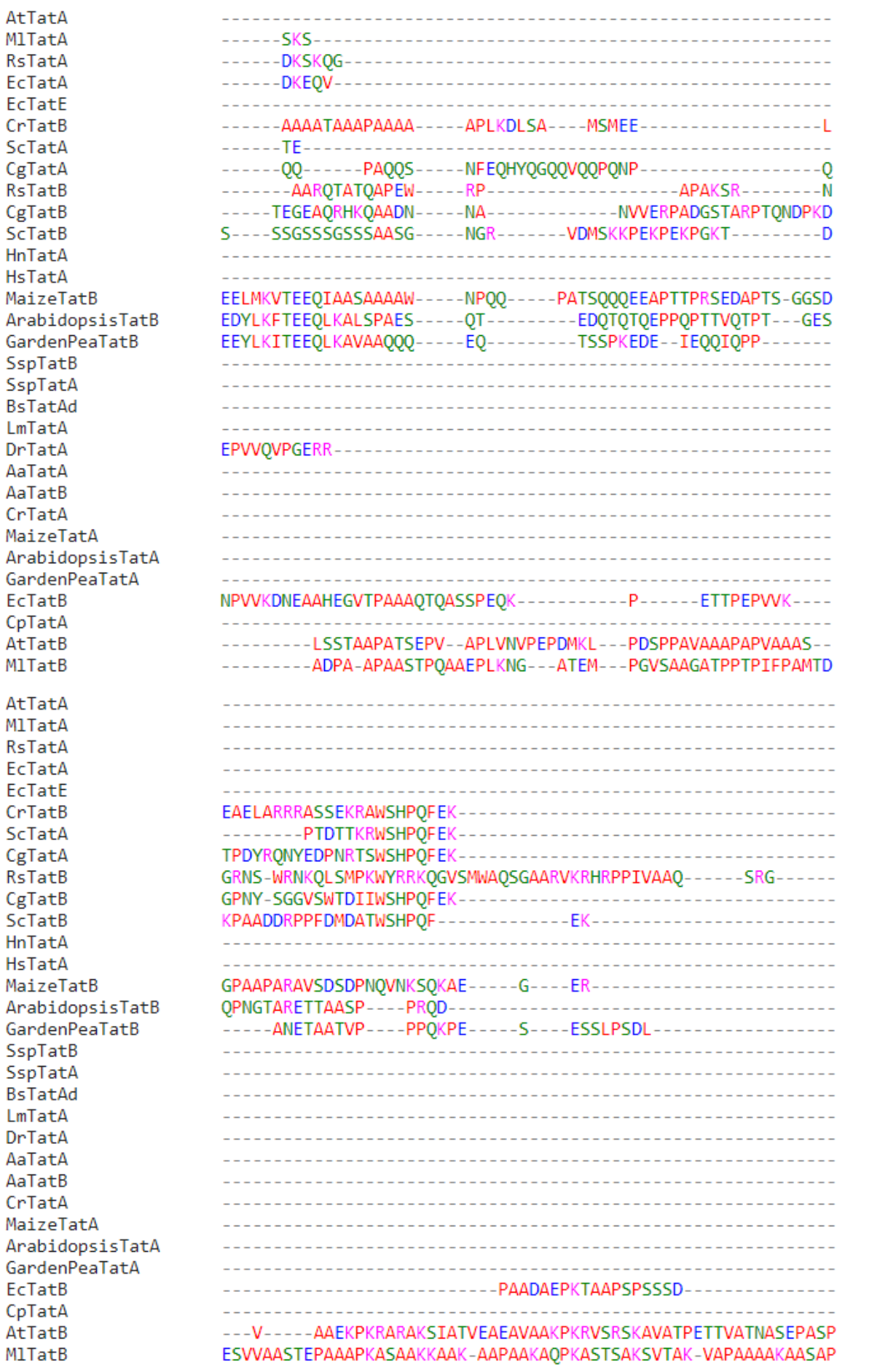


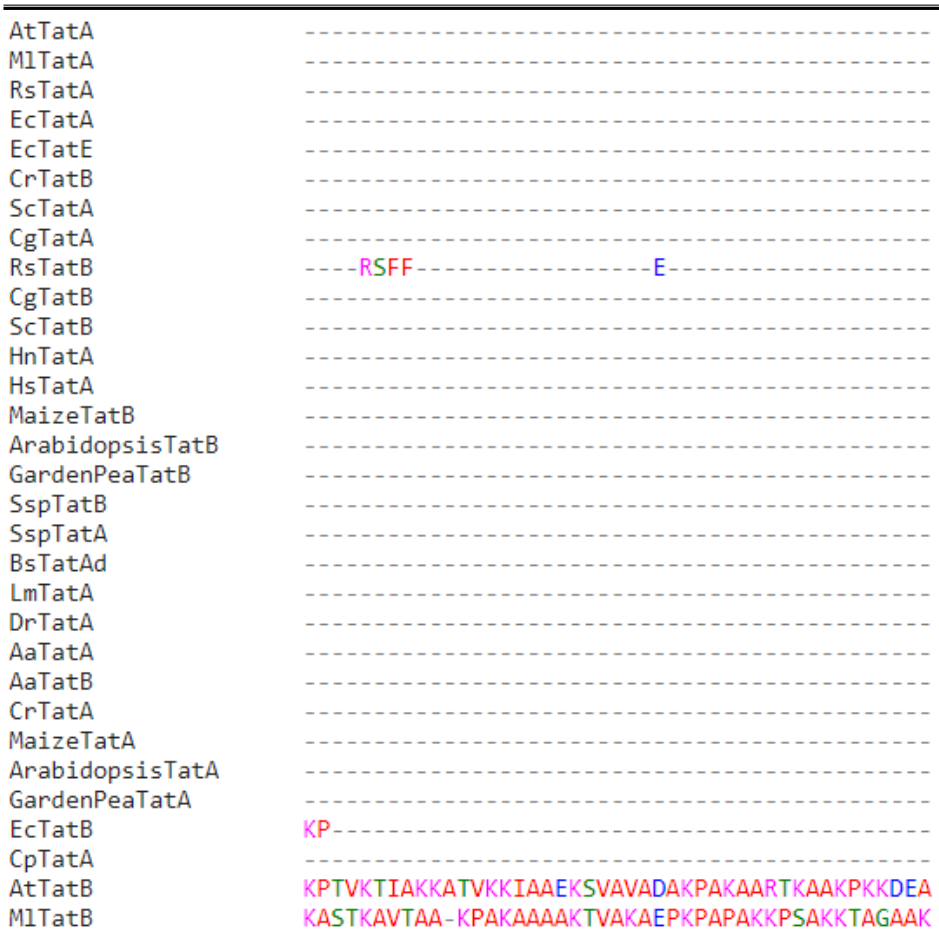
Supplementary Figure 1. Phyre2 structural prediction of scFvM 26tail.


AtTatA
MlTatA
RsTatA
EcTatA
EcTatE
CrTatB
ScTatA
CgTatA
RsTatB
CgTatB
ScTatB
HnTatA
HsTatA
MaizeTatB
ArabidopsisTatB
GardenPeaTatB
SspTatB
SspTatA
BsTatAd
LmTatA
DrTatA
AaTatA
AaTatB
CrTatA
MaizeTatA
ArabidopsisTatA
GardenPeaTatA
EcTatB
CpTatA
AtTatB
MlTatB
AtTatA
MlTatA
RsTatA
EcTatA
EcTatE
CrTatB
ScTatA
CgTatA
RsTatB
CgTatB
ScTatB
HnTatA
HsTatA
MaizeTatB
ArabidopsisTatB
GardenPeaTatB
SspTatB
SspTatA
BsTatAd
LmTatA
DrTatA
AaTatA
AaTatB
CrTatA
MaizeTatA
ArabidopsisTatA
GardenPeaTatA
EcTatB
CpTatA
AtTatB
MlTatB

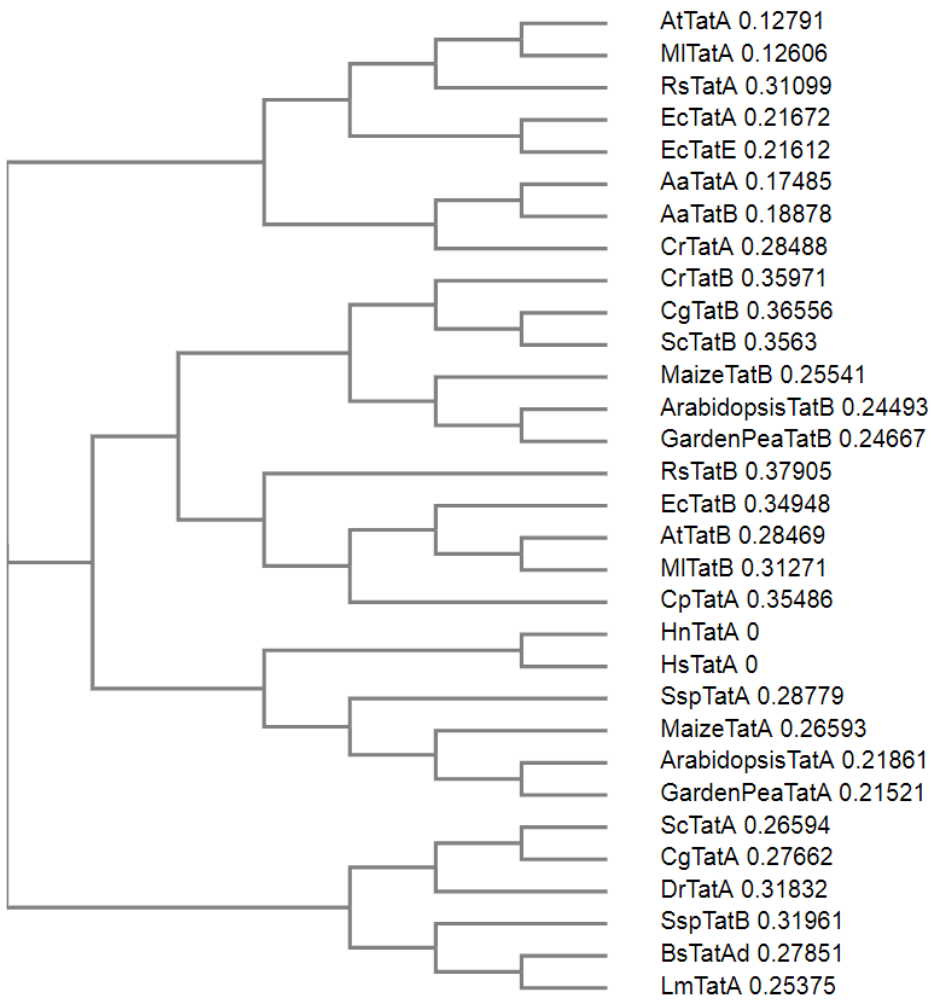
```
AtTatA                LVIVLVLFGRGKIPELMGDVAKGIKSFKKGMADEDQTPPPADA-------NA--NAKTVD
MlTatA                LVIVLLVFGRGKIPELMGDMAKGIKSFKKGMADDDVADDK----------------RTVE
RsTatA                LVIIMMVFGTKKLRNIGSDLGSAVKGFKEGMREGSEDKPAGSQ-------QGQQAAGQPP
EcTatA                AVIVVLLFGTKKLGSIGSDLGASIKGFKKAMSDDEPKQDKTSQ-------DADFTAKTIA
EcTatE                AALVVLLFGTKKLRTLGGDLGAAIKGFKKAMNDDDAAAKKG-A-------DVDLQAEKLS
CrTatB                AVVALVVFGPKGLADAARSVGSALRTFQPTIKEVVQVSQELKGTLESELGINELREAARP
ScTatA                LVVIILLFGAKKLPDMARSLGKSARILKSEAKAMKSEAKADDA----------APADPP
CgTatA                VLLIILLFGAKKLPDAARSIGRSMRIFKSEVKEMNKDGDTPEQ----------QQQQPQ
RsTatB                GAVALVVIGPERLPKVARTAGALIGRAQRYIADVKAEVSREI-------ELEELRKMRTE
CgTatB                VVVGLVVIGPERLPRLIQDARAALLAARTAIDNAKQSLDSDFGSEFDEIRKPLTQVAQYS
ScTatB                VVLAVLVFGPDKLPKVIQDVTRTIRKIREFSDSAKQDIRQELGPEFKDFEFEDLNPKTFI
HnTatA                LLIAILLFGANKIPKLARSSGEAIGEFQKGREEVEQELQEIKSESAPDASADATADT---
HsTatA                LLIAILLFGANKIPKLARSSGEAIGEFQKGREEVEQELQEIKSESAPDASADATADT---
MaizeTatB             GVVALLVFGPKGLAEVARNLGKTLRAFQPTIRELQDVSREFRSTLEREIGIDEVSQSTNY
ArabidopsisTatB       GVVALLVFGPKGLAEVARNLGKTLRTFQPTIRELQDVSRDFKSTLEREIGLDDISTPNVY
GardenPeaTatB         GVVALLVFGPKGLAEVARNLGKTLREFQPTIREIQDVSREFKSTLEREIGIDDITNPLQS
SspTatB               LGVVIIIFGPKRIPELGGALGKTLRGFKEELQTQDVDSTEV----------------VD
SspTatA               FVIALLVFGPKKLPEVGRSLGKALRGFQEASKEFETELKREA-------QNLEK-SVQIK
BsTatAd               FVIALIIFGPSKLPEIGRAAGRTLLEFKSATKSLVSGDEKEE-------KSAEL-TA-VK
LmTatA                VGAALVIFGPKKLPELGRAAGDTLREFKNATKGMMDDSKEET-------KKEDS-RP---
DrTatA                LVVIALLFGAKKLPELMKGMGQGIREFKNEVKDPAAPVPPLQ-------PGQPLQPGQTA
AaTatA                LLVILVIFGASKLPEVGKGLGEGIRNFKKALSGEEE-EK----------------------
AaTatB                LAVILLLFGAGRLPEAGRALGEGIRNFRKALSGETE-VKEVK-------AE-DVKTEERK
CrTatA                AGVAALIFGPSKLPELGKSLGKTVKSFQTAASEFNDELKAGM-------AADDAKKPAAV
MaizeTatA             AGVAALVFGPKQLPEIGRSIGKTVKSFQQAAKEFETELKKEP-------GEGGDQPPPAT
ArabidopsisTatA       AGVAALLFGPKKLPEIGKSIGKTVKSFQQAAKEFESELKTEP-------EESVAESSQVA
GardenPeaTatA         AGVAALVFGPKKLPEVGRSIGQTVKSFQQAAKEFETELKKEP-------NPTEEISVAS-
EcTatB                FIIGLIVLGPQRLPVAVKTVAGWIRALRSLATTVQNELTQELK-------LQEFQDSLKK
CpTatA                LLIAFLVVGPKDLPKIARAIGKGVRYLGNLKEEFGEILNI----------EEEINAVKND
AtTatB                AVVLIVVVGPKDLPPMIRAFGKTMAGLRKMAGDFRTQFDEALK-------EADMDDVRQT
MlTatB                AIVMIVVVGPKDLPNMLRTFGRTTAKLRSMAADFQKQFNEALK-------EAELDDVKKS
                           ::.*   :

AtTatA                ------------------HKADEIK-------------------------------------
MlTatA                ------------------HRADETVSA----------------V---K-EKA---------
RsTatA                ------------------RELHD--------------------ATTIDV-EAR---------
EcTatA                ------------------DKQADTNQEQ-------------AKTEDA-KRH---------
EcTatE                ------------------HKE-----------------------------------------
CrTatB                APRP-VSE-----ATGVSD-TM--AKAS--------------AAPAAP-AAAPAAAPA---
ScTatA                --NP-EQS-----AA--QR-T----------------------IQAAP-GDVTSSRPV---
CgTatA                --Q-----------------Q----------------------QQIAP-NQIEAPQPV---
RsTatB                FEEA-ARNVEQTIHQEVSKHTSEINERLN-------------EALGDP-------------
CgTatB                R----M---SP--KTAITKALFDNDSSF--------------LDDFDPKKIMAEG------
ScTatB                RKQL-DNE-EL--GLKEIRNGFDLKKEM--------------AEVTDAVHGRD----AES
HnTatA                -----TS--DT--TTDATSSAD---------------------DTATN--------------
HsTatA                -----TS--DT--TTDATSSAD---------------------DTATN--------------
MaizeTatB             RPTT-MN-------------------NNQ--------------QPAADP-NVKPEPAPYTS
ArabidopsisTatB       NQNR-TNPVQP--PPPPPPPSVPSTEAP--------------VTANDP-NDSQSPKAYTS
GardenPeaTatB         TYSS-NVR--N--TTPT-PSATEITNNS--------------QTAVDPNGKVDESKAYSS
SspTatB               -VEL-GEN-----DPAAGKK-------------------------------------------
SspTatA               -AEL-EES-----KTPESSSSS---------------------EKA---------------
BsTatAd               -QDK-NAG-------------------------------------------------------
LmTatA                -------------------------------------------------------------
DrTatA                -QTT-VTA-QP--VVDVASRQL--------------------DPMIDPVT----GAPVVN
AaTatA                -----GKEVKK--EGEWSHPQF--------------------EK----------------
AaTatB                -EEK-KEEKEK--VEAWSHPQF--------------------EK----------------
CrTatA                -E-----EKKE--ETKWSHPQF--------------------EK----------------
MaizeTatA             -PTA-VSGGEE--KGLEASSSK--------------------ESA---------------
ArabidopsisTatA       -TS--NKEEEK--KTEVSSSSK--------------------ENV---------------
GardenPeaTatA         -------EQEK--QEIKVSSTK--------------------DNV---------------
EcTatB                VEKASLTN----LTPE-------LKASMDELRQAAESMKRSYVAND-PEKASDEAHTI-H
CpTatA                FEE--IKD----ITKEAEQASRDIIDVMNEAKKDLKDIDGAISEAKSVVE----------
AtTatB                ISD--VRN----LNPTNS-----LRDAMNPLRQLGNEIKSDLQKATSPSDG---------
MlTatB                VDE--LRG----LSPVAE-----IRKQLNPFEQAAADVRAGVDAAMKPKPA---------
```


AtTatA
MlTatA
RsTatA
EcTatA
EcTatE
CrTatB
ScTatA
CgTatA
RsTatB
CgTatB
ScTatB
HnTatA
HsTatA
MaizeTatB
ArabidopsisTatB
GardenPeaTatB
SspTatB
SspTatA
BsTatAd
LmTatA
DrTatA
AaTatA
AaTatB
CrTatA
MaizeTatA
ArabidopsisTatA
GardenPeaTatA
EcTatB
CpTatA
AtTatB
MlTatB
AtTatA
MlTatA
RsTatA
EcTatA
EcTatE
CrTatB
ScTatA
CgTatA
RsTatB
CgTatB
ScTatB
HnTatA
HsTatA
MaizeTatB
ArabidopsisTatB
GardenPeaTatB
SspTatB
SspTatA
BsTatAd
LmTatA
DrTatA
AaTatA
AaTatB
CrTatA
MaizeTatA
ArabidopsisTatA
GardenPeaTatA
EcTatB
CpTatA
AtTatB
MlTatB



Supplementary Figure 2. Clustal Omega multiple sequence alignment of TatA and TatB amino acid sequences from a range of organisms. At – *Agrobacterium tumafaciens*. Ml – *Mesorhizobium loti*. Rs – *Ralstonia salanacearum*. Ec – *Escherichia coli*. Cr – *Chlamydomonas reinhardti*. Sc – *Streptomyces coelicar*. Hn – *Halobavterium sp*. NRC-1. Ssp – *Synechocystis sp* PCC6803. Bs – *Bascillus subtillus*. Lm – *Listeria monocytes*. Dr – *Deinococcus radiodurans*. Aa – *Aquifex aeolicus*. Maize – *Zea mays*. Arabidopsis – *Arabidopsis thaliana*. Garden Pea – *Pisum sativum*. * = fully conserved residue, : = similar properties. Red = small, Blue = acidic, Pink = Basic, Green = hydroxyl/sulfhydryl/amine.



Supplementary Figure 3. Clustal Omega cladogram of TatA and TatB amino acid sequences from a range of organisms. At – *Agrobacterium tumafaciens*. Ml – *Mesorhizobium loti*. Rs – *Ralstonia salanacearum*. Ec – *Escherichia coli*. Cr – *Chlamydomonas reinhardti*. Sc – *Streptomyces coelicar*. Hn – *Halobavterium sp*. NRC-1. Ssp – *Synechocystis sp* PCC6803. Bs – *Bascillus subtillus*. Lm – *Listeria monocytes*. Dr – *Deinococcus radiodurans*. Aa – *Aquifex aeolicus*. Maize – *Zea mays*. Arabidopsis – *Arabidopsis thaliana*. Garden Pea – *Pisum sativum*.