Quality control of substrate conformation in the *Escherichia coli* Twin Arginine protein-targeting pathway

Daphne Maria Johanna Mermans MSc

A thesis submitted for the degree of Doctor of Philosophy

University of Kent

Department of Biosciences

August 2018

Table of Contents

List of figures

List of Tables

Acknowledgements

My PhD journey would not have been possible with the kind support and help of many great people. Therefore, I would like to take the opportunity to express my sincere thanks to all of them. First, I would like to thank my supervisor Professor **Colin Robinson** for the opportunity to work as a PhD student in his lab, for the time and effort that he has committed and the useful discussions we have had.

I would like to thank **Jo Roobol** and **Ally Walters** for their effort in keeping the lab running smoothly. I would also like to acknowledge other past and present members of the CR-lab, such as **Wayne Miller, Andrew Dean, Sarah Bisschof** and **Chi Jawara** for their assistance, which has been very helpful and an important contribution to my project. You all made the CR-lab such a great place to work. I will miss all the biscuits, cake, chocolate and coffee. A huge thank you to **Kelly Walker, Doris Gangl** and **Kirsty Richards** for teaching me everything during the first months of my PhD. I would like to thank **Julie Zedler** and **David Russo** for all their advices on a scientific career and their encouraging words. I also enjoyed our time together on sunny Sicily. Thank you **Emi Nemoto-Smith** for being such a great company and for all the fun nights out in Canterbury. I would like to thank **Serena Lima** for sharing her honest opinion and all your Italian food lecturers**.** Big thanks to **Amber Peswani** and **Conner Webb,** it was a pleasure to explore Amsterdam with you! I am looking forward to more adventures. A special thank you to **Kelly Frain** (KMF) and **Alex Jones** (Dr. Jones) for helping brainstorm ideas, mental support and for great company and entertainment throughout the time of my project. It was great to go through the whole process of completing a PhD together. We have collected some amazing memories during my time in Canterbury and I will miss you guys!

I would also like to thank the Marie Curie ITN consortium that I was proud to be part of. Thank you for all the fruitful discussions and happy memories during these years! It was great to be part of such a nice group with inspiring scientists. Thanks to **Isabel Guerrero-Montero** for being my CR-buddy within the ITN consortium. Thank you to **Jan-Maarten van Dijl** and **Sierd Bron** for coordinating all the great meetings.

Last but not least a very special thanks to my family and my friends back home for their great support and visits to Canterbury. In particular, I would like to thank my parents, **Nicolle** and **Frank Mermans**, for providing me with opportunities, support and encouragement during my PhD and life in general. Thank you to my sister **Anouk Mermans** and brother-in-law **Jasper Bracké** for making coming home special every time. We truly had some good laughs and I am looking forward to many more. Big thanks to the cutest dog **Maks** for saying welcome home every single time with all your cuddles, kisses and snuggles. Thank you to my aunt **Jose Spijkers** for adding nice memories to my time in Canterbury by all your nice visits. It was great to have you here and to go shopping, eat scones and laugh a lot. Thousand thanks to **Felix Nicolaus** for being a massive support for me during my PhD. Thank you for always being available to listen to me and to check my sequencing results with me at 3am in the morning so I can sleep without having to worry. Thank you for making me laugh and always seeing the positive in everything. Your encouragement and optimism have made me a better person and scientist. Special thanks go to my dear grandpa (**Opa**) for being an immense support. I could always count on you. Your support has been very valuable to me and I am very thankful for that. You were looking so much forward to witnessing me defending my thesis and you will be missed on this day.

Thank you!

Declaration

The work presented in this thesis is original, and was conducted by myself (unless otherwise stated) 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.

This research was funded by a Marie Curie Initial Training Network grant (Horizon 2020, ProteinFactory, 642863).

Collaborative work has been conducted contributing to producing data used in this thesis. I would like to thank the following collaborators:

George Sutherland, **Professor Neil Hunter**, and **Professor Leslie Dutton** for the Maquette BT6 NMR work and the useful advices.

Professor **Gunnar von Heijne** for inviting me to carry out experiments in his lab and the useful discussions.

Kevin Howland and **Gary Thompson** for the mass spectrometry analysis and NMR analysis.

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

Sutherland GA, Grayson KJ, Adams NBP, Mermans DMJ, Jones AS, Robertson AJ, Auman DB, Brindley AA, Sterpone F, Tuffery P, Derreumaux P, Dutton PL, Robinson C, Hitchcock A, Hunter CN (2018) Probing the quality control mechanism of the Escherichia coli twin-arginine translocase with folding variants of a de novo-designed heme protein. *J Biol Chem.* doi: 10.1074/jbc.RA117.000880

Daphne M.J. Mermans, Alexander S. Jones, Gary S. Thompson, C. Neil Hunter, P. Leslie Dutton, Colin Robinson (2018) The twin-arginine proofreading ability of *Escherichia coli* can sense localized unfolded regions of a man-made maquette protein variant.

Summary

In *Escherichia coli*, the twin arginine translocase (Tat) is one of the major protein translocation mechanisms. The Tat system has the ability to transport folded proteins across the inner membrane. Therefore, it has the ability to discriminate between folding states. However, it is not well understood how the Tat system senses the folding state of a substrate. In this study we probed the Tat proofreading mechanism and we investigated whether Tat substrates in *E. coli* are translocated by the Tat system due to their rapid folding kinetics.

We demonstrate that the *E. coli* Tat machinery can process a *de-novo* designed substrate (BT6 maquette). Moreover the Tat proofreading mechanism can discriminated between different folding states of this substrate. This data and the fact that this simple four helix artificial substrate offers a lot of engineering freedom, suggests that BT6 is an ideal candidate to study the Tat proofreading mechanism (chapter 3).

In chapter 4, we focussed on the Tat system's proofreading ability by substituting substrate surfaces of BT6 maquette. Mutants with substituted surface properties were expressed in order to understand what Tat senses as folded. Expression assays showed whether the mutants were accepted or rejected by Tat. We propose that the proofreading system does not sense a global unfolded state of the substrate but has the ability to sense localised unfolded regions.

Finally, we tested whether Tat substrates fold co- or post-translationally to determine the speed of the folding kinetics by using an arrest peptide-mediated force measurements assay (chapter 5). This study was to increase our understanding about the rationale for using the Tat system.

List of abbreviations

°C	degrees Celsius
µL	microlitre
µM	micromolar
A	alanine (A)
A_c	arrested-length control
AxA	alanine – any residue – Alanine
AP	arrest peptide
APH	amphipathic helix
APS	ammonium persulphate
ATP	adenosine triphosphate
bp	base pair
C	cytoplasmic fraction
C-terminus	carboxy-terminus
D	Aspartic acid
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DTT	dithiothreitol
E. coli	*Escherchia coli*
ECL	enhanced chemo luminescence
EDTA	ethylenediaminetetraacetic acid
ER	endoplasmic reticulum
FeS	iron-sulphur
Ffh	fifty-four homolog

f_{FL}	fraction full-length
FL_c	full-length control
FT	flow through
g	grams
GFP	green Fluorescent Protein
GTP	guanosine triphosphate
H	histidine (His)
hGH	human growth hormone
HRP	horseradish peroxidase
IFN	human interferon α2b
IMAC	immobilised metal affinity chromatography
IMP	inner membrane protein
IPTG	isopropulthiogalactoside
ISO	isothermal reaction buffer
kDa	kilodalton
KR	lysine (Lys), arginine (Arg)
LB	Luria Bertani broth
LBA	Luria Bertani agar
M	membrane fraction
M	molar
mA	milliampere
MBP	maltose binding protein
mg	milligram
MGD	molybdenum guanine dinucleotide
min	minute
mL	millilitre

mM	millimolar
NBS	nucleotide-binding site
ng	nanogram
N-terminus	amino-terminus
OD	optical density
P	periplasmic fraction
PAGE	poly-acrylamide gel electrophoresis
PBS	phosphate buffered saline
PBS-T	phosphate buffered saline and Tween20
PCC	protein conducting channel
PhoA	alkaline phosphatase
PMF	proton motif force
RNA	ribonucleic acid
REMP	redox enzyme maturation protein
rpm	revolutions per minute
RNC	ribosome-nascent chain
RR	twin-arginine motif
SDS	sodium dodecylsulphate
Sec	general secretory pathway
Sec	seconds
SP	signal peptide
SPase	signal peptidase
SR	SRP receptor
SRP	single Recognition Particle
Tat	Twin Arginine Translocase
TCA	trichloroacetic acid

TEMED	tetramethylethylenediamine
TMAO	trimethylamine N-oxide
TMH	transmembrane helix
TMS	transmembrane segment
TorA	trimethylamide-N-oxide reductase
TorA-	torA signal peptide fused to following protein
Tris	tris (hydroxymethyl) aminomethane
Tween20	polyoxyethylenesorbitan monolaurate
UV	ultraviolet
V	volts
v/v	volume per volume
w/v	weight per volume
WT	wild-type strain

Chapter 1:

General introduction

Cells are the building blocks of life. Within each individual cell, various essential biochemical processes take place to maintain the cell. Cellular life can be divided into three domains: archaea, bacteria and eukaryotes. Eukaryotes have a more complex construction with different organelles, such as a cell nucleus. Archaea and bacteria both are prokaryotes, which don't have a cell nucleus. However, archaea differ from bacteria in many structural, biochemical and physiological characteristics, such as the lack of peptidoglycan in their cell wall. Bacteria can be divided into two groups: Gram-positive and Gram-negative. The difference between Gram-negative and Gram-positive is caused by a difference in the structure of the cell wall. Gram-positive bacteria have a thick peptidoglycan layer that is less impenetrable than the relatively thin peptidoglycan layer (inner membrane) of Gram-negative bacteria. However, the structure of Gram-negative cells is more complex compared to Gram-positive cells. A second membrane is located outside the thin peptidoglycan layer. The outer membrane contains lipopolysaccharides (LPS), which offers the cell an extra protection. The two membranes form an extracellular compartment called the periplasm.

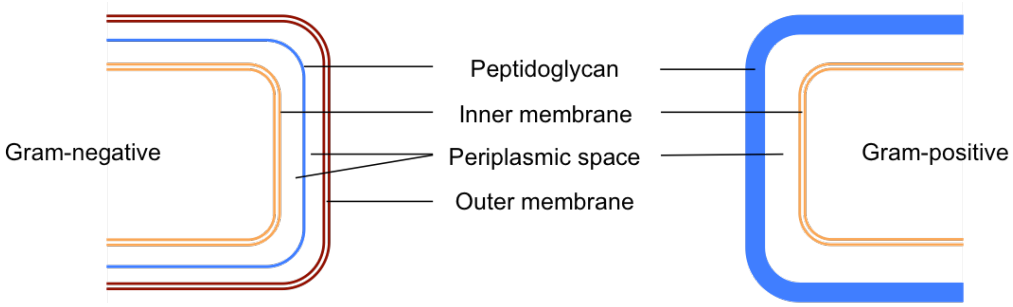


Figure 1. Schematic comparison of Gram-negative and Gram-positive cells

This figure demonstrates the difference between the cell walls of Gram-negative and Gram-positive cells. The peptidoglycan layer is shown in blue, the inner membrane in orange and the outer membrane in red. Both type of cells contain an inner cell membrane and a peptidoglycan layer. However, the peptidoglycan layer of Gram-positive cells is much thicker. Gram-negative cells have an additional outer membrane build out of lipopolysaccharides (LPS).

The periplasmic compartment contains many periplasmic proteins that are involved in different biochemical pathways, such as peptidoglycan synthesis, electron transport and cell division. Optimal protein trafficking is essential as more than one third of the proteome of Gram-negative cells are secreted across or inserted into the inner and outer membrane (Economou, 2010a). There are three major pathways to transport proteins from the cytoplasm to the periplasm or to insert proteins into the inner membrane; Tat translocase (Tat), secretory (Sec) translocase and YidC insertase.

1.1 Protein targeting

Proteins are synthesized by ribosomes in the cytosol and many need to be transported across or into the inner membrane. Therefore, they need to be recognised and targeted by the Tat or Sec translocase embedded in the inner membrane (Driessen and Nouwen, 2008, Rapoport, 2007).

1.1.1 The signal peptide

Precursor proteins possess an amino-terminal (N-terminus) signal peptide (SP) to direct the translocation or membrane insertion (Blobel, 1975). Although the sequence and length of signal peptides are highly variable between different substrates and organisms, they do have a homologous tripartite structure consisting of a positively charged hydrophilic N-terminus, a hydrophobic core (h-region) and a polar carboxyl-terminal (C-terminus) domain containing the signal peptidase cleavage site (von Heijne, 1987). Figure 2 shows a schematic comparison of the Tat and Sec N-terminal signal peptides. Hydrophobic as well as electrostatic interactions take place between the signal peptide and the membrane components of the translocase. Previous studies demonstrated the importance of the signal

peptide's hydrophobicity. The hydrophobic amino acids span the membrane as an α-helix loop (Briggs *et al.*, no date). Moreover, alterations of single hydrophobic residues to charged residues in the signal peptide blocked the export of the secretory maltose binding protein (MBP) (Bedouelle *et al.*, 1980). The positively charged N-terminal domain has been suggested to interact with the negatively charged phospholipid head groups of the inner membrane (Nesmeyanova *et al.*, 1997). During or shortly after translocation, the C-terminal domain is recognized by type I signal peptidase (SPase) that cleaves the signal peptide from the mature secretory protein (Paetzel *et al.*, 2002). Type I SPase binds small, uncharged residues at the -1 and -3 position of the C-terminus, typically AxA (von Heijne, 1983).

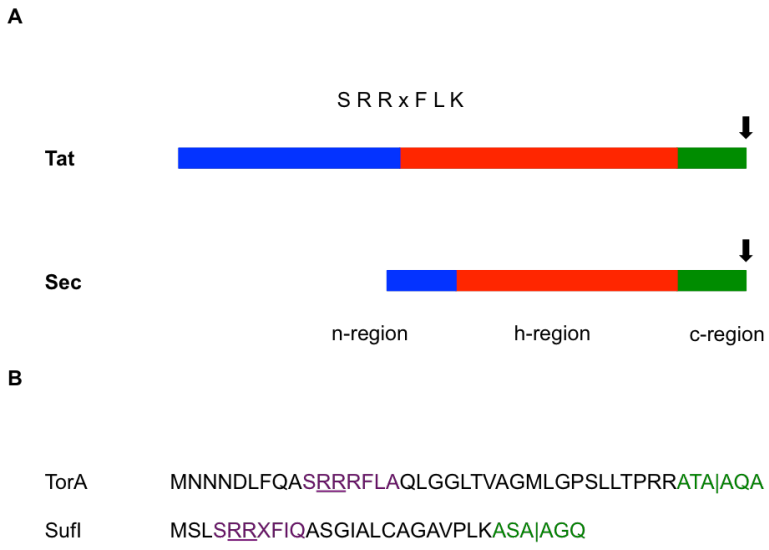


Figure 2. Schematic overview of Tat signal peptides vs. Sec signal peptides

(A) The N-terminal Tat and Sec signal peptides are displayed. Both signal peptides have a homologous tripartite structured consisting of a hydrophilic n-region (blue), a hydrophobic h-region (red) and a polar c-region (green) containing the signal peptidase cleavage site (typically AxA). (B) The sequences of TorA and SufI, the two most commonly used signal peptides in studies on the Tat pathway. The Tat specific RR-motif is demonstrated in purple and underlined. The type I signal peptidase cleavage site is displayed in green.

Sec precursor proteins do not contain a Sec specific signal sequence motif except for the cleavage site. The sec components recognise the tripartite structure of the Sec precursor protein. The h- and n- regions are critical structural elements. SRP (Signal Recognition Particle) binds to the h-region via hydrophobic interactions. Studies have demonstrated that increased hydrophobicity promotes SRP binding (Peterson *et al.,* 2003, Lee and Bernstein, 2001). The binding affinity of the motor protein SecA for signal peptides increases with the number of positive charges in the n-region (Cunningham and Wickner, 1989).

1.1.2 The Tat signal sequence

Like most other bacterial secretory proteins, Tat substrates possess an N-terminal signal sequence to direct the precursor proteins to the periplasm. Tat signal peptides possess the previously described tripartite structure with the type I signal peptidase cleavage site on the polar c-region (figure 2). Although the structures of Sec and Tat signal peptides share similarities, Tat signal peptides differ from Sec signal peptides.

Tat signal peptides (figure 2) are 26 – 58 amino acids long, whereas Sec signal peptides are 18 – 26 amino acids long. The n- and h-region of Tat signal peptides are larger than Sec signal peptides (von Heijne, 1985, Berks, 1996). The overall hydrophobicity of the Tat signal peptide plays an important role in Tat dependent translocation. The h-region of Tat signal peptides is in general less hydrophobic than h-regions of Sec signal peptides. It has been demonstrated that secretory proteins can be redirected to the Sec translocase when the hydrophobicity of the Tat signal peptide increases (Cristóbal *et al.*, 1999).

Besides the differences in length and hydrophobicity, the Tat signal peptides are recognised by a conserved Twin-Arginine motif. The name giving twin arginine motif was defined as S-R-R-x-F-L-K and is located at the n-region of the Tat signal peptide (Berks, 1996). Although the RR-motif was initially thought to be invariant, paired arginine residues are not an absolute requirement. Numerous studies have shown that the first arginine residue can be substituted for a lysine residue without blocking tat dependent export (Stanley, Palmer and Berks, 2000). Moreover, a natural substrate is known were the signal peptide is harbouring a KR-motif. The signal peptide of the TtrB subunit of the tetrathionate reductase of *Salmonella enterica* is Tat dependent but contains a single arginine motif (Hinsley *et al.*, 2001). Other studies demonstrated that the substitution of a single arginine either by a Lys, Asn or Gln is tolerated by the Tat translocase (DeLisa *et al.*, 2002). The substitution of the two arginine residues to lysines completely abolishes translocation (Buchanan *et al.*, 2002).

In addition to these Tat specific signal peptide features, the Tat signal peptides usually possess one or more positively charged residues in the c-domain. These residues are rarely found in Sec signal peptides and prevent interactions with the Sec translocase (Bogsch *et al.*, 1997). The presence of these positively charged residues functions as the so-called 'Sec-avoidance signal'. The charge of the c-region of Tat signal peptides is on average +0.5, whereas the average charge of Sec signal peptides is +0.03 (Berks *et al.,* 2003).

Other studies have suggested that a positive charge in the first amino acids of the mature protein is also important in 'Sec avoidance'. This suggests that several Tat substrates that do not contain cofactors or form multimers are Tat dependent because they require positively charged N-terminal residues to function which abolishes recognition by the Sec

translocase (Tullman-Ercek *et al.*, 2007). It thus appears that, the twin arginine motif, overall hydrophobicity and charge are important determinants for Tat specific targeting.

1.2 The Secretory translocase

The Sec translocase is the most described translocon and is located in the inner membrane. It is present in bacteria, archaea, the thylakoid membrane of plant chloroplasts and in the endoplasmic reticulum (ER) of eukaryotic cells (Muñiz *et al.*, 2001). The sec translocase acts as a transporter of unfolded proteins to the periplasm as well as a mechanism to insert proteins into the lipid bilayer of the inner membrane. In bacteria, a multimeric protein complex, known as SecYEG, forms the protein-conducting channel (PCC). SecYEG interacts with the peripherally associated motor domain, ATPase SecA, which is the motor of the translocation reaction. Another multimeric membrane protein complex, SecDFyajC, interacts with SecYEG to facilitate protein translocation (Veenendaal *et al.,* 2004). In addition, YidC is an integral membrane protein that is involved in the insertion of proteins into the inner membrane. Figure 3 represents an overview of protein transport by the Sec machinery.

1.2.1 The Sec components

The two integral membrane proteins SecY and SecE form the core of the protein-conducting channel. Although, SecG is not essential for protein translocation it interacts with the SecYE complex, resulting in the SecYEG complex formation (Figure 3).

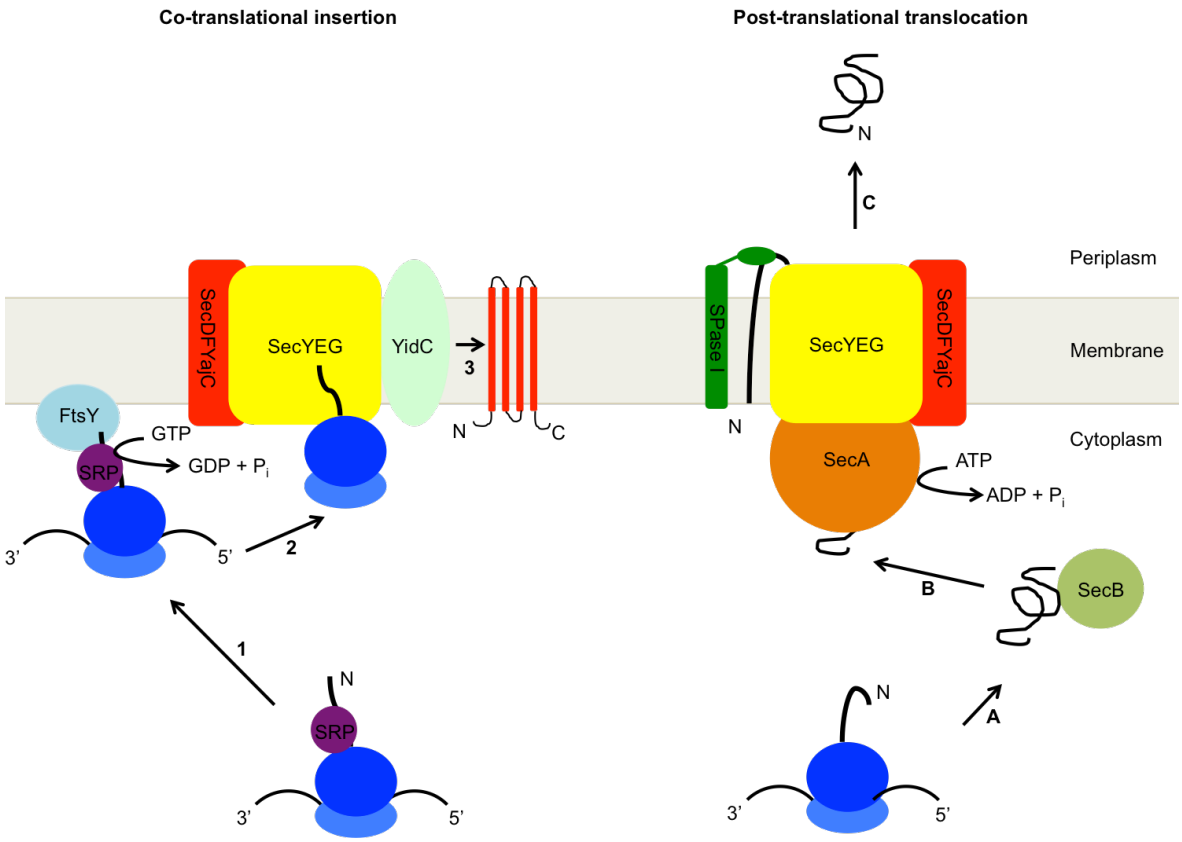


Figure 3. Schematic overview of the Secretory (Sec) pathway

The co-translational protein insertion pathway is listed on the left. (1) SRP, in bacteria Ffh, binds the N-terminal signal peptide, resulting in an SRP-bound ribosome complex. This complex than binds to the bacterial SRP receptor (SR), denoted as FtsY. (2) The Ribosome-SR complex associates with the SecYEG complex. (3) In combination with the hydrolysis of GTP, the insertion of the nascent membrane protein into the inner membrane takes place. During protein insertion, YidC can cooperate with the Sec system. The post-translational protein translocation pathway is displayed on the right. (A) Following translation, the N-terminal signal sequence of the unfolded precursor protein binds to the cytosolic chaperone SecB. (B) SecB mediates the precursor protein to the ATPas SecA. Once the precursor protein and SecA bind to the SecYEG complex, the protein can be transported across the inner membrane by repeated cycles of ATP hydrolysis. (C) Shortly after translocation, SPase I cleaves the signal peptide from the protein. The protein will fold in the periplasm to form a mature protein.

1.2.1.1 SecY

The largest SecYEG component is SecY, which is a highly hydrophobic 48 kDa protein and it spans the membrane with 10 transmembrane segments (TMSs) of which five are exposed in the periplasm and six in the cytoplasm including the N- and C-terminal regions (Akiyama and Ito, 1987). The fifth cytoplasmic loop (C5) plays an important role in SecA-dependent translocation. To allow SecA insertion and protein translocation, SecY activates the ATPase activity by interacting with SecA (Matsumoto *et al.,* 1997) Mutagenesis studies on the C5 loop resulted in the disruption of the SecY-SecA interaction (Mori and Ito, 2001)

1.2.1.2 SecE

In most bacteria, SecE is a single membrane spanning protein of which the third transmembrane segment (TMS3) and the second cytoplasmic loop (C2) are homologous to *E. coli* SecE (Schatz *et al.*, 1989). In *E. coli,* SecE is a 14 kDa clamp-like protein composed of three TMSs and an amphipathic helix (Frauenfeld *et al.*, 2012). Despite the three TMSs, only one of the three is sufficient for cell survival. Truncation of the first two TMSs does not affect cell viability and protein translocation (Schatz *et al.*, 1991). Highly conserved residues of TMS3 and C2 are essential for protein translocation as it contributes to the stability of the SecYE complex. C2 forms an amphipathic helix structure which might be important in the formation of a pore and for the stability of the SecYE complex. TMS3 serves as a membrane anchoring domain to C2 (Pohlschröder *et al.,* 1996).

1.2.1.3 SecG

In *E. coli*, SecG is a 12 kDa protein and acts as a core-associated and stimulatory component of the PCC. It compasses two TMSs linked by a weaker cytosolic exposed linker (Nishiyama *et al.*, 1996). SecG plays an important role in SecA translocation ATPase activity and stimulates SecYE mediated protein translocation. Although SecG is not essential for cell viability, deletion of SecG results in a cold-sensitive phenotype. Studies have shown that SecG plays a critical role in protein translocation as deletion of the gene impairs translocation at temperatures below 37 °C (Nishiyama *et al.*, 1994). Furthermore, *in vitro* assays revealed that secG is essential for the stimulation of protein translocation in the absence of the proton motive force (PMF). These observations suggest that SecG plays an important role in protein translocation in the absence of PMF and/or at low temperatures (Hanada *et al.*, 1996).

1.2.1.4 SecDFyajC

SecYEG interacts with an accessory hetrotrimeric membrane complex, called SecDFyajC, which comprises SecD, SecF and YajC (Duong and Wickner, 1997). In *E. coli*, SecD and SecF are two separate 67 and 35 kDa proteins, which share a similar topology. Both Sec components contain 6 TMSs and an additional large periplasmic domain. In *E. coli*, yajC is a 12 kDa single spanning protein with a large cytosolic domain. It has been suggested that the SecDFyajC complex is thought to play a role in the biogenesis of membrane proteins as well as the translocation and folding of secreted proteins (Schlegel *et al.*, 2014).

1.2.2 Post-translational targeting

Secretory proteins that are translocated after they have been fully synthesised are targeted post-translationally (Figure 3). Most secreted proteins are associated with the cytosolic chaperone SecB in a post-translational manner as they are targeted to the membrane after synthesis has been completed (Driessen and Nouwen, 2008, Rapoport, 2007). SecB is a cytosolic chaperone, which is involved in the translocation of secretory proteins. It is a 17 kDa acidic homotetrameric soluble protein (Xu *et al.*, 2000). The precursor protein is targeted and stabilised by the SecB after most residues of the nascent precursor protein have emerged form the ribosome exit tunnel (Driessen, 2001). SecB has two functions: it stabilises the precursor and it passes the precursor protein on to the ATPase SecA. Although SecB maintains precursor proteins in an unfolded and translocation competent state, it is not essential for protein translocation (Tsirigotaki *et al.*, 2017). Protein folding of substrates that require an oxidative environment will also be prevented in the reducing environment of the cytoplasm (Economou, 2010). Moreover, it has been demonstrated that the N-terminal signal peptide delays the folding of fast folding mature domains (Beena *et al* ., 2004).

Translocation of secretory proteins from the cytoplasm to the periplasm requires the input of an energy source. Protein translocation via the Sec system can be driven by chemical energy (ATP, GTP) and electrochemical energy (proton motive force, PMF). GTP is a driving force for polypeptide elongation during translation by the ribosome. ATP is the main energy source during post-translational translocation and initiates the translocation event. ATP facilitates the movement of substrates through the protein conducting channel (PCC) and coordinates the opening and closure of the PCC (Wickner, 1994). SecA is the ATP-dependent force generator, however soluble SecA exhibits a low ATPase activity

(Van Der Wolk *et al.*, 1997). Once precursor proteins and the SecYEG complex bind to SecA, the activity is stimulated (Lill *et al.*, 1990). SecA consists of two nucleotide-binding sites (NBS), one is located in the N-terminal domain and has a high ATP affinity (NBS-I) and one has a low ATP affinity (NBS-II) and is located in the C-terminal domain (Papanikolau *et al.*, 2007). ATP binding facilitates translocation of the signal sequence and the N-terminal mature region of the precursor protein. Once the first residues (20 – 30 residues) of the precursor protein are inserted the protein will be released from SecA by hydrolysis of the ATP molecule at NBS-I (Schiebel *et al.*, 1991). Following the release the precursor protein, SecA will be released from the membrane by hydrolysis of the second ATP molecule at NBS-II. The release of SecA allows the translocation intermediate to bind a new SecA molecule. In the absence of rebinding to SecA, the precursor protein will move backwards. Binding to a new SecA molecule and hydrolysis of ATP drives the translocation of another 20 – 30 residues. Multiple cycles take place to complete the translocation of the whole protein (Natale *et al.*, 2008). The PMF can prevent the precursor protein to move backwards and complete the translocation (Driessen & Wickner, 1991).

1.2.3 Co-translational targeting

The biogenesis of inner membrane proteins (IMP) in Gram-negative bacteria follows a co-translational pathway (Figure 3). The pathway can be divided in three steps: targeting, insertion and folding/quality control. Membrane-integrated proteins are targeted during their translation as ribosome- bound nascent chains (RNCs) in a different pathway involving the signal recognition particle (SRP). SRP is a ribonucleoprotein that is involved in co-translational targeting of integral membrane and secretory proteins to the ER membrane in eukaryotes and the inner membrane in prokaryotes. The *E. coli* SRP is the simplest SRP variant and is composed of one Ffh (fifty-four homolog) protein and a

relatively small 4.5S RNA component. SRP binds with high selectivity to the first hydrophobic segment of the nascent protein chain (signal anchor sequence), which subsequently is mediated to the Sec translocon (Luirink *et al.*, 2005). Cleavable signal peptides of most proteins translocated across the inner membrane are in general less hydrophobic than signal anchor sequences. Therefore they are targeted via the SecB pathway as there is less affinity for SRP binding (Randall and Hardy, 2002). Despite the specificity for IMPs, some secretory proteins containing a cleavable signal peptide can be targeted via the SRP pathway. The signal peptide of the periplasmic disulfide bond oxidoreductase (DsbA) is significantly more hydrophobic than other Sec signal peptides. Furthermore, the post-translational pathway is relatively slow compared to the co-translational pathway. It has been suggested that DsbA might be directed to the SRP pathway because of its rapid folding in the cytoplasm (Schierle *et al.*, 2003). Other studies have shown that secretory proteins can be redirected to the SRP pathway when the hydrophobicity of the signal peptides is increased (Bowers and Lau, 2003, Lee and Bernstein, 2001). Surprisingly, two secretory proteins (Hbp and SecM), which do not contain an exceptional hydrophobic signal peptide, are targeted via the SRP pathway. However, they both contain an unusually long signal peptide (Sijbrandi *et al.*, 2003, Nakatogawa and Ito, 2001). The SRP interacts with the RNC in a relatively early stage when the first trans membrane segments (TMs) are exposed outside the ribosome (Houben *et al.*, 2005).

Once the SRP targets the RNC, the *E. coli* SRP receptor (SR) binds to the ribosome-SRP complex. The *E. coli* SR is known as FtsY, and can be found both in the cytosol as in the inner membrane (Luirink *et al.*, 2004). FtsY targets proteins to the membrane, which are required for cell division. Therefore, depletion of FtsY leads to a cell division defect.

Membrane associated FtsY supports targeting to the inner membrane through its affinity for lipids and the sec translocon component SecY (Angelini *et al.,* 2005).

The SRP-mediated co-translational protein targeting is regulated by GTP binding and hydrolysis in both the SRP and FtsY (Luirink and Sinning, 2004). Both the SRP and FtsY have to be in a GTP-bound state for SRP-SR complex formation. Once the SRP interacts with the ribosome the GTP binding affinity is increased (Bacher *et al.*, 1996). At this stage, the binding of the signal anchor sequence inhibits GTP hydrolysis. The GTP bound SRP and FtsY stimulate each other to form a SRP-SR complex formation (Powers and Walter, 1995). GTP hydrolysis dissociates the SRP-SR complex after the signal anchor sequence has been released (Connolly *et al.*, 1991).

The insertion and assembly of integral membrane proteins have been studied intensely in the past decades, however the entire process is highly complex and has still not been fully understood. Once the RNC is targeted to the translocon, the first transmembrane segment (TMS) of the nascent membrane protein enters the interior of the channel of the heteromeric membrane complex SecYEG (Veenendaal *et al.*, 2004). Through a lateral gate each TMS can leave the channel into the lipid bilayer one by one or in pairs (Luirink *et al.*, 2005). It is assumed that the gate continuously opens and closes, which brings the TMS in equilibrium between the hydrophilic environment in the channel and the lipid. The TMS could first intercalate between the TMSs forming the lateral gate and then be released in the lipid bilayer phase (Rapoport *et al.*, 2004). By gaining more insight in the insertion mechanisms research has also focused on how TMSs are actually recognized by the translocon (White and von Heijne, 2008). An *in vitro* expression system has already explored the hydrophobicity requirements for the insertion of transmembrane helices (TMHs) (Hessa *et al.*, 2005). Based on the integration propensity of a designed TMH, the

free energy for the insertion of each naturally occurring amino acid was calculated, resulting in a biological hydrophobicity scale. Furthermore, it could be shown that the free energy was dependent on the position of the amino acids within a test TMH (Hessa *et al.*, 2007). These observations suggest that translocon recognition requires a direct interaction between the emerging TMH and the surrounding lipid. It has been shown that some membrane proteins can also integrate via the membrane-integrated YidC insertase (Van Der Laan *et al.*, 2005, Dalbey and Kuhn, 2004).

1.3 The Twin Arginine Translocase

As previously described, the Sec translocase transports unfolded proteins across the inner membrane. In contrast, the twin arginine translocase (Tat) transports folded proteins across membranes. Tat is located on the inner membrane of bacteria, archaea, thylakoid membranes of plant plastids and mitochondria of plants (Bogsch *et al.*, 1998). The energy force that is required for Tat dependent translocation is the previously mentioned PMF. Transport of folded proteins is mainly carried out by three integral membrane proteins: TatA, TatB, and TatC. In Gram-positive bacteria and archaea, TatA and TatC form a TatAC substrate-binding complex. TatA, TatB and TatC form a TatABC substrate-binding complex in Gram-negative bacteria and plant chloroplasts (Figure 4).

Although the rationale for using Tat remains unclear, three main reasons have been identified. Many Tat dependent substrates fold around a co-factor in the cytoplasm which bypasses an extra mechanism to export the cofactor to the periplasm (Berks, 1996). Another reason is avoidance of competition between metal ions that compete for insertion into the active site of Tat dependent substrates. The periplasmic metalloprotein MncA binds Mn^{2+} to avoid competition of periplasmic metal ions (Tottey *et al.*, 2008). The third

rationale for the use of Tat is that some tat dependent proteins form hetero-oligomeric complexes in the cytoplasm. This so-called hitchhiker co-translocation occurs when a subunit binds a twin arginine containing subunit to form a dimeric protein. The bacterial periplasmic nickel-binding hydrogenase 2 (HYD2) is composed of a 30 kDa small subunit (hybO) and a 60 kDa large subunit (hybC). The small subunit contains an N-terminal Tat specific signal peptide, whereas the large subunit does not contain an N-terminal signal peptide. However, it does contain additional C-terminal residues that are involved in keeping the subunit in a co-factor insertion competent conformation. This additional C-terminal tail is cleaved upon co-factor insertion by a cytoplasmic protease. This event results in a conformational change, which allows interaction with hybO to form the hetero-oligomeric HYD2 complex. The formation of HYD2 triggers translocation of the complex to the periplasm (Rodrigue *et al.*, 1999). In archaea, Tat exports monomers and cofactor-less proteins because of their fast folding kinetics due to a high cytoplasmic salt concentration (Rose *et al.*, 2002).

Tat substrates play an important role in energy metabolism, formation of the cell envelope, iron and phosphate acquisition, cell division and cell motility. Because they are involved in biogenesis of the cell envelope, the phenotype of *Δtat* deletion strains can be characterised by the formation of long cell chains (Hinsley *et al.*, 2001).

Since Tat exports fully folded proteins, it consists of a proofreading and a quality control mechanism. Due to this unique capacity it is likely that the Tat system has a natural tendency to produce high quality, active products. This is particularly relevant for biotechnological exploitation. Various studies have shown the Tat system's ability to reject unfolded proteins (Robinson *et al.*, 2011, Maurer *et al.*, 2009). Emerging evidence suggests that quality control involves more than the proofreading as misfolded substrates

rapidly degrade after interaction with the Tat components (Matos *et al.,* 2008). Other studies have shown that the Tat system tolerates minor structural changes (Alanen *et al.*, 2015). Furthermore, studies seem to suggest a hydrophobicity sensor employed by the Tat apparatus rather than assessment of the folded state of substrates (Richter *et al.*, 2007).

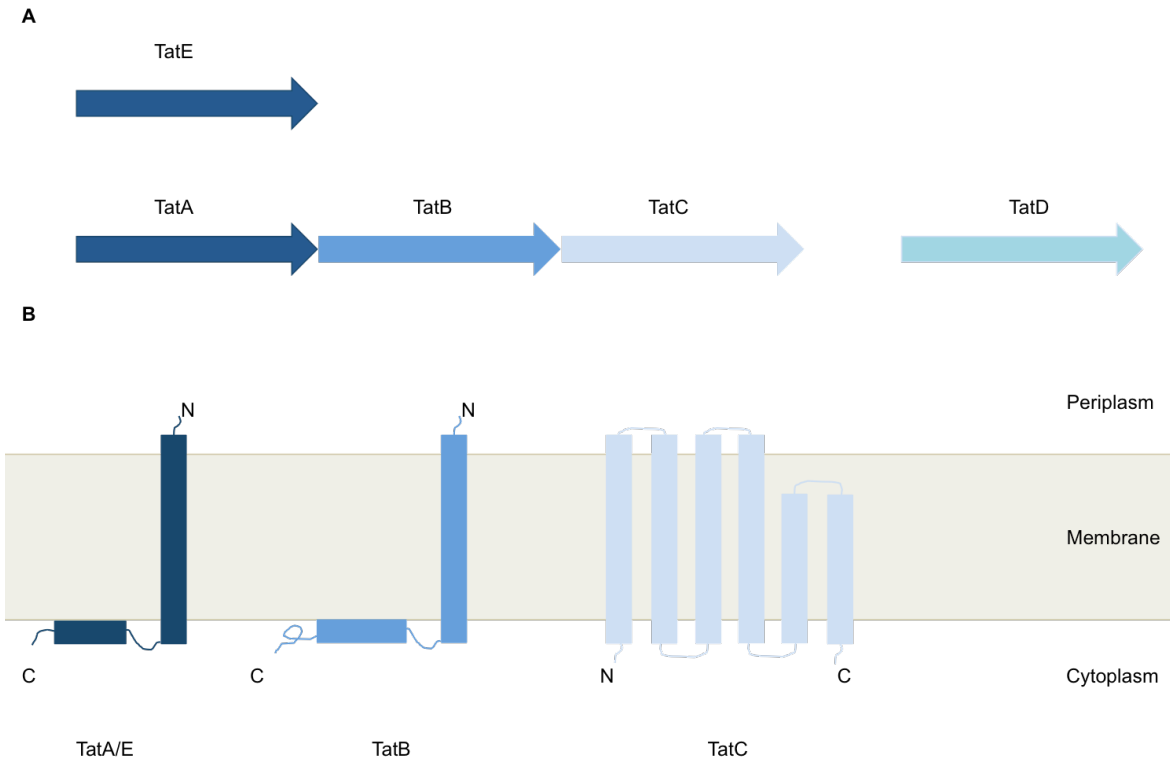


Figure 4. Organisation of the Tat translocase system in Gram-negative bacteria

(A) The integral membrane proteins TatA/E, TatB and TatC are expressed from the *tatABC* operon. A fourth gene is expressed downstream of the *tatABC* operon, *tatD*. The monocistronic *tatE* gene encodes TatE. (B) TatA, TatE and TatB are single spanning integral membranes that consist of an N-terminal TMS, a amphipathic helix and an unstructured C-terminal region. Compared to TatA and TatE, TatB has a longer amphipathic helix and C-terminal region. TatC contains 6 TMSs with both the N- and C-termini located in the cytoplasm. TMS5 and TMS6 are too short to completely span the membrane.

1.3.1 The Tat components

Two types of membrane proteins assemble the Tat translocase: a hexahelical TatC-type protein and TatA-type single spanning proteins, TatA, TatB and TatE. In *E. coli*, three integral membrane proteins TatA, TatB and TatC are expressed form the *tatABC* operon, which contains a downstream *tatD* gene encoding the cytosolic TatD component (Figure 4) (Wexler *et al.*, 2000). The TatA component forms a pore or a channel in the inner membrane to allow translocation of folded substrates from the cytoplasm to the periplasm. TatB and TatC form a stable TatBC complex and is the initial receptor of the Tat translocase pathway (Patel *et al.*, 2014). Translocation of Tat substrates is initiated by the interaction of the TatBC complex with Tat substrates via their twin arginine signal peptides (Cline and Mori, 2001). Binding of the signal peptide to the TatBC complex requires the PMF across the inner membrane and allows the recruitment of TatA (Frobel *et al.,* 2012).

1.3.1.1 TatA

Compared to other Tat components, TatA is synthesised at the highest level (Palmer and Berks, 2001). TatA is a 9.6 kDa single spanning integral membrane protein and it consist of an N-terminal TMS, a partially membrane-embedded amphipathic helix (AMP) and an unstructured C-terminal region (Figure 4)(Chan *et al.*, 2011). It has the ability to form a pore in the inner membrane by homo-oligomerization with other TatA components. Two regions of high sequence conservation are essential for TatA activity: (i) the interdomain hinge between the N-terminal TMS and the amphipathic helix and (ii) an invariant phenylalanine residue within the amphipathic helix (Hicks *et al.*, 2003). Furthermore, studies showed that the truncation of 40 C-terminal amino acids did not affect the TatA activity (Lee *et al.*, 2002). NMR studies have determined the structure of the orthologous

$TatA_d$ in *Bacillus subtilus* in bicelles containing phospholipids as a membrane-mimicking environment. The relatively short 14 amino acids long N-terminal TMS was deeply inserted into the inner membrane and is tilted by 13°. Because the TMS is relatively short, it pulls the N-terminus of the APH into the membrane. This causes the hinge region and the N-terminal part of the APH to be immersed into the lipid bilayer in a tilt angle and was thought to posses a higher mobility than the N-terminal TMS (Torsten *et al.,* 2010). Single-particle electron microscopy studies support the hypothesis that TatA forms the PCC of the Tat translocase. TatA is facing a challenging task since Tat dependent substrates range in size from less than 10 kDa to 150 kDa. TatA forms homo-oligomeric complexes to form pores of variable diameters to allow translocation of Tat dependent substrates. A lid formed by the APHs to maintain the ionic permeability barrier of the membrane gates the cytoplasmic end of the pores (Gohlke *et al.*, 2005, Palmer and Berks, 2012). It has been suggested that the walls of the pores are formed by interactions between the TMSs. Ile^{12} and Val^{14} arrangements between TMSs were observed in homo-oligomeric TatA complexes. The Ile^{12} residue, located on the side of the TatA TMS, was found to contact Val^{14}, located on the other side of the TMS form another TatA TMS (White *et al.*, 2010).

1.3.1.2 TatB

TatB functions as a mediator between substrate recognition and translocation events. It is a 18.5 kDa single spanning TatA-like Tat component and has evolved into a functionally independent protein due to a gene duplication of *TatA* (Yen *et al.*, 2001). Despite sharing 20% sequence identity and being structural homologous to TatA, there are functional and structural differences. Both TatA and TatB comprise an N-terminal TMS, a partially membrane-embedded amphipathic helix (AMP) and an unstructured C-terminal region. However, TatB is larger than TatA and has a longer APH and a longer C-terminal tail

(Figure 4) (Hicks *et al.*, 2003). Like TatA, the region between the TMS and the APH are important for the function of TatB. Moreover, site-directed mutagenesis studies demonstrate the importance of glutamic acid residues in the APH (Hicks *et al.*, 2003).

Although no single mutation has been identified to inactivate the TatB function, it is essential for the translocation of endogenous Tat substrates in *E. coli* (Frobel *et al.*, 2012). In contrast, studies have demonstrated low levels of export of an artificial Tat substrate containing a twin arginine signal peptide in *ΔTatB* mutant stains (Ize *et al.*, 2002). These findings might be an indication for a Tat-B-like activity of *E. coli* TatA. In addition, Gram-positive bacteria posses TatAC-systems, which lacks TatB. Therefore, Gram-positive bacteria possess a bifunctional TatA component, which has the ability to mediate between substrate recognition and translocation events, and it fulfils the pore-forming function of TatA. Furthermore, it has been demonstrated that substitutions in TatA of Gram-negative bacteria can compensate for the absence of TatB, suggesting that the bifunctionality of ancestral orthologues of TatA can be restored in *E. coli* (Blaudeck *et al.*, 2005).

1.3.1.3 TatC

TatC is the largest and most conserved Tat component and it is the only Tat component with multiple transmembrane domains (Figure 4). The primary site of interaction between a twin arginine signal peptide and the Tat translocase is TatC. This component is 28.9 kDa and exhibits 6 TMSs with the N- and C- termini located in the cytoplasm (Behrendt *et al.*, 2004). 3D crystallisation studies on TatC from the thermophilic bacterium *Aquifex aeolicus* have revealed that TatC resembles a concave-shaped structure that faces the periplasm because TMS5 and TMS6 are too short to span the membrane (Rollauer *et al.*, 2013). They are relatively shorter than the other TMSs resulting in the membrane to be

slanted inwards at the ends of TMS5 and TMS6. This causes thinning of the membrane around TatC, which might be required for translocation of Tat dependent substrates across the membrane or to accommodate TatB. The two periplasmic loops form a structured periplasmic cap that stabilises the position and tilted orientations of the TMSs. Site-directed mutagenesis studies demonstrated that the N-terminus, the first cytoplasmic loop and the two periplasmic loops are essential for TatC activity (Allen *et al.*, 2002, Buchanan, *et al.*, 2002). In addition, site-directed mutagenesis of TMS5 also inactivates TatC, probably because there is a specific contact site between TMS5 of TatC and the TMS of TatB (Kneuper *et al.*, 2012).

1.3.1.4 TatD

The *tatABC* operon expresses a fourth gene downstream, *tatD,* encoding the soluble cytoplasmic 28.9 kDa TatD component. Although the function of TatD is largely unknown, it is not essential for Tat dependent translocation. However, the co-transcription of *tatD* with three essential genes for Tat translocation strongly suggests an involvement of TatD in the Tat translocase system. TatD has a high affinity for metal ion-binding and functions as a magnesium-dependent DNase (Wexler *et al.*, 2000).

1.3.1.5 TatE

Some bacteria, including *E. coli*, express a TatA like component, TatE, that is encoded by a monocistronic *tatE* gene (Figure 4). Whilst TatB emerged from TatA by an relatively early gene duplication event, TatE evolved by a late gene duplication event (Yen *et al.*, 2002). Despite TatE sharing 53% sequence identity with TatA, it is 7 kDa and much shorter (Sargent *et al.*, 1998). Beside the structural similarities, previous studies suggested

that TatE is a functional paralog of TatA. In *ΔtatA* strains, overexpressed TatE can facilitate efficient translocation of several Tat substrates (Sargent *et al.*, 1999). However, this was thought to be insufficient to persist TatE during evolution. Despite 3D density maps of purified TatE demonstrating ring-shaped structures, the TatE complexes are much smaller and discrete (Baglieri *et al.*, 2012). Recent studies revealed that TatE interacts with TatA, TatB and TatC, suggesting that TatE is a regular constituent of the Tat (A)BC-complex. In contrast to TatE being a functional surrogate of TatA it is now thought to cooperate with TatA. TatE and TatA assembling into hetero-oligomeric complexes and it was confirmed that TatE cross-links with a common binding site of TatA to TatC (Eimer *et al.*, 2015). Recently, it has been demonstrated that TatE and TatB share functional similarities.

Unlike TatA, TatB can counteract premature processing of translocation- incompetent substrates (Fröbel *et al.*, 2012). TatE shares the ability to counteract premature processing, albeit with less efficiency (Eimer *et al.*, 2018).

1.3.2 The Tat (A)BC-complex

The Tat signal peptide contains a twin arginine signal peptide (1.1.2), which is recognised by a TatBC complex (figure 5). This event triggers the proton motive force-dependent recruitment and oligomerization of TatA and TatB to form the Tat(A)BC-complex (Rose *et al.*, 2013, Alami *et al.*, 2003). As discussed in section 1.3.1, the individual Tat components have the tendency to form homo-oligomeric complexes. For the assembly of the Tat translocase in *E. coli* and other Gram-negative bacteria, the TatA, TatB and TatC proteins form the hetero-oligomeric TatABC-complex (figure 5). TatB and TatC are the most stable components and form the substrate receptor complex. Both components associate in a 1:1

ratio and due to homo-oligomerisation a functional TatBC complex consists of several TatBC dimers. TatA is less associated and interacts with the complex by binding to TatC (Bolhuis *et al.*, 2001).

The TatBC complex recognises and binds to Tat dependent twin arginine signal peptides. The cytosolic N-terminus and the first cytosolic loop of TatC play a role in recognition and binding to the RR-motif of the signal peptide (Buchanan, *et al.*, 2002). Crosslinking between the twin arginine motif of the signal peptide with residues on the N-terminus and first cytosolic loop were observed in a PMF independent manner. In addition, several single residues that cross-links with the RR-motif of the signal peptide also cross-link to TatB (Zoufaly *et al.*, 2012). Other studies suggested that the K residue within the SRRxFLK consensus motif, interacts with a conserved EE motif within the TatC cavity (Ramasamy *et al.*, 2013). Once the signal peptide is recognised by the TatC complex it inserts into the central cavity of TatC in a hairpin loop-like conformation. (Fincher *et al.*, 1998, Berks *et al.*, 2014). The insertion of the signal peptide triggers the TMS of TatB to bind to TMS5 of TatC by the formation of a site-specific disulphide bond (Kneuper *et al.*, 2012).

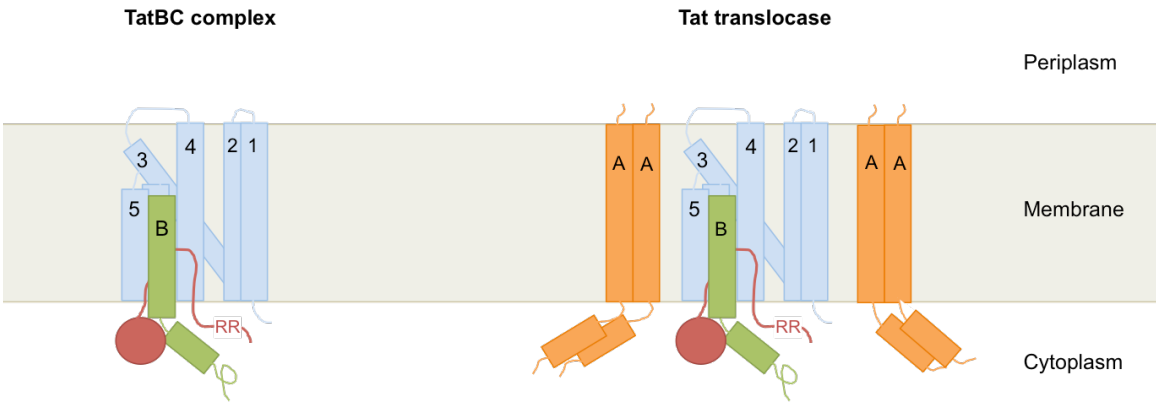


Figure 5. The topology of the Tat(A)BC complex

The TatC components are displayed in blue, TatB components in green, TatA components in orange and the substrate in red. The binding of the twin arginine signal peptide to the TatBC receptor complex is displayed on the left. The signal sequence is embedded between helices of TatC and TatB. Insertion of the signal peptide into the inner membrane by TatC triggers TatB to interact to the TatC TMS5. The formation of the precursor-TatBC-complex recruits TatA components resulting in the hetero-oligomeric TatABC-complex.

In the presence of the PMF, the signal peptide is accommodated in a concerted TatBC-binding groove. A deep hairpin loop-like insertion between TatB and TatC, results in a TatBC-binding pocket that prevents premature cleavage by signal peptidase. Studies have demonstrated that a Tat translocase lacking TatB, allows premature removal of the signal sequence of a translocation incompetent precursor substrate (Fröbel *et al.*, 2012). Besides the twin arginine consensus motif, TatB also contacts the hydrophobic h-region of the signal peptide (Maurer *et al.,* 2010). Recent studies provided evidence that the cytosolic domains of TatB interact with the first residues of the folded mature protein (Ulfig and Freudl, 2018). These findings suggest that TatB supports the binding of Tat substrates to the TatBC substrate receptor complex. Substrate binding to TatBC triggers the recruitment

of TatA in a PMF dependent manner (Mori and Cline, 2002). TatA oligomers facilitate protein translocation by forming a size-fitting translocation pore (trap-door model) or by weakening the membrane (Gohlke *et al.*, 2005, Rodriguez *et al.*, 2013).

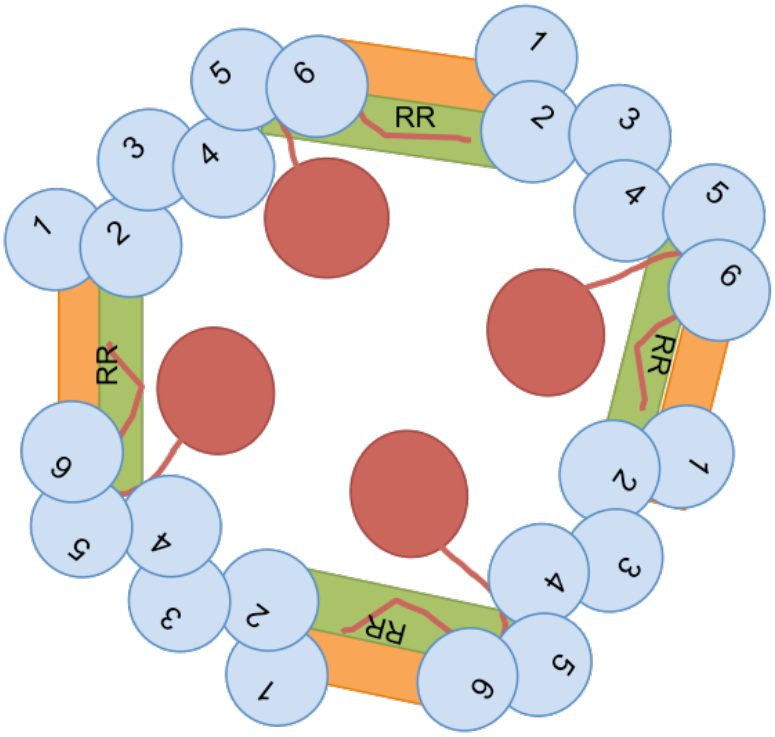


Figure 6. Cytoplasmic overview of the Tat(A)BC-complex

A schematic overview of the TatABC complex in the inner membrane after substrate binding. The TatC components are displayed in blue, TatB components in green, TatA components in orange and the substrate in red. TatA inserts between two TatC components to increase the diameter of the TatBC complex. TatB forms a bridge between two TatC components by interacting with TMS5 and TMS2.

1.3.3 The Tat system of Gram-positive bacteria

Although this study is focussed on the most frequently studied Tat system in *E. coli*, the understanding of the Tat system in Gram-positive bacteria, such as *Bacillus subtilis* and *Staphylococcus aureus* has largely increased. This has gained insight in some strong similarities and potential differences between Gram-positive and Gram-negative bacteria (Goosens *et al.*, 2014).

A major difference is the lack of a TatB component, suggesting that the TatA-like components of Gram-positive bacteria perform the functions of both TatA and TatB. Gram-positive bacteria possess a minimal Tat system, composed of two TatA-like components and one TatC component. In *B. subtilis*, three TatA-like components and two TatC variants have been described, denoted TatAd, TatAy, TatAc, TatCd and TatCy. Previous studies revealed that *B. subtilis* contains two active TatAC- systems named TatAdCd and TatAyCy (Jongbloed *et al.*, 2004). The TatAdCd system only exports a single substrate, PhoD, to the periplasm under conditions of phosphate starvation (Jongbloed *et al.*, 2000). The TatAyCy system is constitutively expressed and exports various substrates including, EfeB (YwbN), QcrA, and YkuE, in all conditions (Jongbloed *et al.*, 2004). Despite the fundamental differences between the TatABC and TatBC-systems, the TatAd component of the *B. subtilis* Tat-system is able to compensate for the absence of components of the *E. coli* Tat system (Barnett *et al.*, 2011). TatAc is the third TatA-like component of *B. subtilis* and its function remains unknown. However, it has been suggested that it can form active Tat translocase systems with TatCd and TatCy (Monteferrante *et al.*, 2012). In addition, TatAc can functionally replace the *E. coli* TatA and TatE components, suggesting that TatAc has a relevant role in protein translocation (Beck *et al.*, 2013).

To ensure correct folding and co-factor insertion, protein interactions between substrate specific chaperones take place. However, the understanding of the *B. subtilis* Tat system's quality control and proofreading remains unknown, protein interactions between substrate specific chaperones have been observed. Proteomics studies showed that the protease WprA is important for the translocation of YwbN by TatAyCy and that HemAT and CsbC are overexpressed during the translocation of PhoD by TatAdCd (Monteferrante *et al.*, 2013). Moreover, the chaperones DnaJ and SufS that are potentially associated with quality control were highlighted during proteomics studies (Goosens *et al.*, 2013). Unfortunately, no Gram-positive specific chaperones that are associated with quality control and proofreading have been identified.

1.3.4 The Tat translocase mechanism

While the understanding of protein targeting and association of Tat substrates with Tat(A)BC-complexes has significantly increased, the actual translocase mechanism remains to be the least understood step of the Tat pathway. Two translocation mechanism models have been proposed: the trap-door model and the membrane-weakening model (figure 7) (Patel *et al.*, 2014). The two proposed models are listed in figure 7.

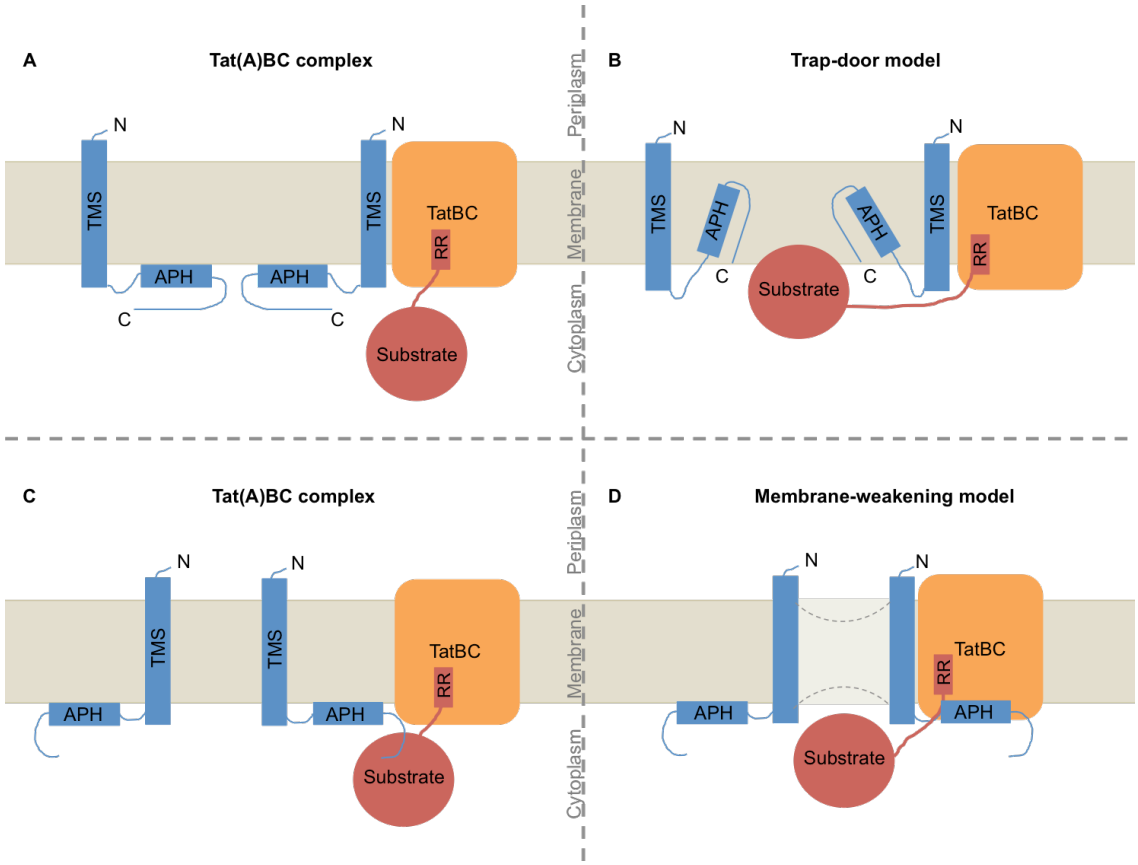


Figure 7. The Tat translocation mechanisms

The figure shows the two proposed models for protein translocation by the TatA complex before and during translocation. (A/C) A substrate containing a twin arginine signal peptide is recognised by the TatBC complex. (B) The trap-door model is based on the hypothesis that the amphiphatic helix of the TatA component has a dual topology. The amphipathic helix flips into the membrane resulting in a hydrophilic pore to translocate substrates in the presence of the proton motive force. (D) The membrane-weakening model hypothesises that the formation of a precursor-TatBC-complex induces the APH to disrupt the lipid bilayer, which permits translocation of the substrate.

1.3.4.1 Trap-door model

Although the translocation mechanism is not well understood, it is known that TatA permits translocation of the substrate by the formation of an integral pore (Gohlke *et al.*, 2005). The TatA forms pores of variable diameters to allow translocation of substrates. The walls of the pores are formed by the TMSs whereas the APHs form a cytoplasmic lid to gate access to the translocase whilst maintaining the ionic permeability barrier of the membrane (White *et al.*, 2010).

The trap-door model is based on the ability of the APHs of TatA to adopt different topologies (figure 7B) (Gouffi *et al.*, 2004). It has been suggested that the APHs form a hydrophilic pore by obtaining a membrane spanning topology. Another study suggested a 'charge zipper mechanism' (Walther *et al.*, 2013). This mechanism is based on the fact that the sequence of charges of the C-terminal region of TatA is complementary to the charge pattern on the APH. The two complementary sequences of charges can assemble via a 'charge zipper' resulting in the formation of a hairpin. Interaction of the substrate with the TatA hairpins triggers the hairpins to flip into the membrane to form a hydrophilic pore.

1.3.4.2 Membrane weakening model

In contrast to the 'trap-door model', the 'membrane-weakening model' is based on the idea that TatA destabilises the lipid bilayer (figure 7D). It has been hypothesised that TatA components accumulate in an unordered manner resulting in destabilisation of the membrane (Brüser and Sanders, 2003). This hypothesis was supported by NMR studies suggesting that the hinge between the APH and TMS lacks the flexibility to allow the formation of a hairpin. Furthermore, the APH is partially embedded in the membrane

which might have a lipid disrupting effect (Torsten *et al.*, 2010). Electron microscopy studies of the *B. subtilis* TatAd component showed that TatAd can not form the translocation channel. In contrast to TatA pore complexes of 7.5 – 9 nm diameter in *E. coli*, TatAd forms relatively small pores of 2.5 – 3 nm diameter. Furthermore, TatAd does not exhibit the heterogeneity of TatA but can only form homogeneous pores (Beck *et al.*, 2013). The fact that the TatA TMS and TatC TMS5 and TMS6 are too short to completely span the membrane might also cause disruption of the lipid bilayer (Rodriguez *et al.*, 2013, Rollauer *et al.*, 2013). During translocation TatC pulls the substrate across the destabilised membrane in a PMF dependent manner (Kudva *et al.*, 2013).

The phage shock protein PspA in *E. coli* is believed to be involved in protein translocation (DeLisa *et al.*, 2004). It has been suggested that PspA maintains export competence and enhances translocation efficiency. In addition, it plays an important role in maintaining PMF efficiency under cellular stress conditions (Michiel Kleerebezem, 1996). It has been suggested that cellular stress induced by membrane destabilisation recruits PspA to counteract membrane stress (Mehner *et al.*, 2012).

1.3.5 Proofreading and quality control of Tat substrates

Tat dependent substrates are exported via the Tat pathway because they require the enzymatic insertion of cofactors in the cytoplasm or form a hetero-oligomeric complex in the cytoplasm (Patel *et al.*, 2014). The Tat pathway exports correctly folded substrates by a remarkable proofreading and quality control mechanism. Various studies have shown the Tat system's ability to reject unfolded proteins (Robinson *et al.*, 2011, Maurer *et al.*, 2009). In addition, emerging evidence suggests that quality control involves more than the proofreading as misfolded substrates rapidly degrade after interaction with the Tat

components (Matos *et al.*, 2008). The incorporation of cofactors is mediated by the quality control mechanism whereas the assessment of the overall folding state is done by the proofreading mechanism (figure 8).

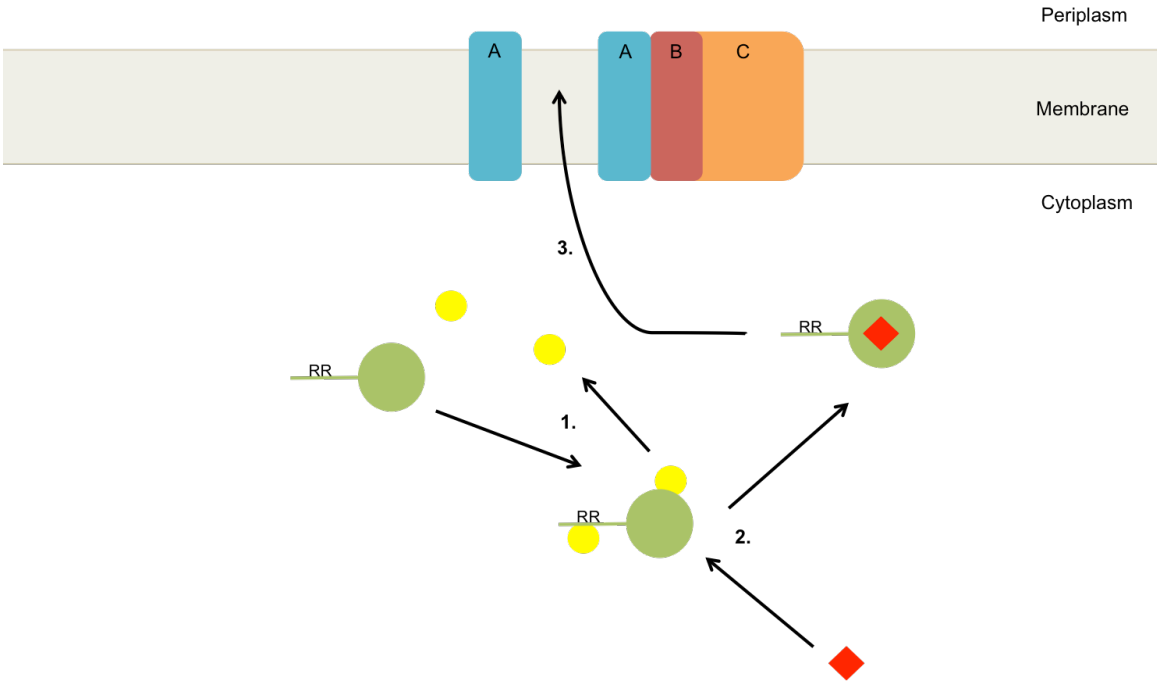


Figure 8. Tat proofreading and quality control

The Tat quality control mechanism is listed. In the membrane the Tat translocase complex is depicted containing: TatA (Blue), TatB (Red) and TatC (orange). (1) A chaperone (yellow) associates with the Twin arginine signal peptide of a precursor substrate (green). In some cases a second chaperone may bind to the mature region of the precursor substrate. (2) The binding of one or more chaperones allows cofactor insertion needed for substrate assembly. Following cofactor insertion the chaperones are released and the fully assembled precursor substrate is targeted by the Tat components for translocation.

To ensure that cofactors are correctly assembled, specific chaperones (redox enzyme maturation protein (REMP)) are involved. The most well-characterised REMPs are TorD, DmsD, HyaE and NapD (Frobel *et al.*, 2012). Trimethylamine N-oxide (TMAO) reductase acquires the redox cofactor, molybdenum guanine dinucleotide (MGD), to be correctly folded prior to export. The third gene of the *torCAD* operon encodes the cytoplasmic 22.5

kDa TorD protein, which interacts specifically with the unfolded Tat substrate, Trimethylamine-N-Oxide reductase (TorA) (Pommier *et al.*, 1998). TorD associates with the TorA protein by recognising two TorD-binding sites, the N-terminal TorA signal peptide which is unique to TorA and the mature region of the protein (Jack *et al.*, 2004). Although a monomeric form of TorD associates with TorA, TorD dimerization has been observed. The binding of a single TorD chaperone to TorA facilitates efficient insertion of the MGD cofactor into the unfolded TorA precursor protein. The structure of precursor TorA is composed of four domains (I-IV), with domain I-III in a folded conformation and the C-terminal domain IV presented to the TorD chaperone. TorD prevents premature closing of domain IV to allow MGD insertion. Moreover, co-factor insertion allows domain IV to fold correctly which triggers TorD release and subsequent translocation of the now correctly folded export competent TorA substrate (Dow *et al.*, 2013). It has been suggested that TorA-TorD interactions are mediated in a GTP-dependent manner (Guymer *et al.*, 2010). Besides facilitating MGD insertion and preventing premature translocation, studies have demonstrated that TorD also has the capability to reduce proteolysis and to facilitate oligomerization of protein subunits. It has been demonstrated that the translocation of the heterologous co-factor less Green Fluorescent Protein (GFP) containing a TorA signal peptide was enhanced possibly by reducing the proteolysis of precursor GFP (Li *et al.*, 2006). Moreover, TorD was required for the assembly of the Tat dependent heterodimer hydrogenase-2 (NiFe) (Jack *et al.*, 2004).

DmsD is related to the TorD family and is the REMP associated with dimethylsulfoxide (DMSO) reductase, which is a heterotrimeric redox enzyme. DMSO comprises the 85.8 kDa MGD cofactor containing DmsA, the 23.1 kDa membrane-associated iron-sulphur protein DmsB, and the 30.8 kDa DmsC which acts as a membrane anchor (Ray *et al.*, 2003). DmsD can interact with the hydrophobic region of the twin arginine signal peptide-

containing precursor DmsA and precursor TorA (Winstone *et al.*, 2013). Studies suggested that DmsD is involved in the biogenesis of DMSO rather than the functioning of the signal peptide (Oresnik *et al.*, 2001). It has been suggested that DmsD guides DmsA through a cascade of chaperone assisted protein biogenesis as it interacts with several chaperones, such as DnaK, DnaJ, GrpE and GroEL (Li *et al.*, 2010). In addition it also interacts with several proteins that play a role in the MGD biosynthesis pathway, such as MoaA and MoeB (Chan and Turner, 2015).

HyaE and HybE are the REMPs associated with the heterodimers hydrogenase-1 and -2 (NiFe) (Dubini and Sargent, 2003). Hydrogenase-1 is a heterotrimeric substrate consisting of a core heterodimer of a Fe-S-binding β-subunit (HyaA) and a NiFe-cofactor binding α-subunit (HyaB). The core dimer associates with a third integral membrane cytochrome b γ-subunit (HyaC). Hydrogenase-2 comprises the HybOC complex containing the α-subunit HybC and the β-subunit HybO. The hydrogenase-1 and -2 β-subunits are equipped with the N-terminal twin arginine signal peptide (Rodrigue *et al.*, 1999). HyaE and HybE prevent premature translocation or interaction with the Tat components before the biogenesis process has been completed. Moreover, they play an important role in the NiFe cofactor insertion of the α-subunits prior to dimerization with the β-subunits. In addition the Fe-S clusters must be assemble in the β-subunits prior to dimerization with the α-subunits (Magalon and Böck, 2000). HybE associates with both the signal peptide containing subunit as well as the second subunit of hydrogenase-2. In contrast, HyaE only interacts with the signal peptide containing β-subunit, HyaA, of the hydrogenase-1 substrate (Dubini and Sargent, 2003).

NapD is the REMP that associates with the nitrate reductase A enzyme (NapA) and they both are encoded in the seven-gene *NapFDAGHBC* operon. NapA is a 90 kDa substrate that binds a Fe-S cluster and accommodates a MGD cofactor (Jepson *et al.*, 2007). NapD is a small 9.3 kDa cytoplasmic chaperone that associates with the n- and h-regions of the N-terminal signal peptide of the NapA precursor substrate. Moreover, significant changes in NapA signal peptide conformation have been demonstrated (Grahl *et al.*, 2012). Studies suggested that NapD plays an important role in the insertion of the MGD cofactor to complete the maturation of NapA (Dow *et al.*, 2014). Beside NapD, NapA is targeted by another cytoplasmic chaperone, NapF. Like NapD, NapF is a quality control related chaperone which plays an important role in protein assembly prior to translocation (Nilavongse *et al.*, 2006).

Beside the ability to carry out chaperone-mediated proofreading, the Tat pathway can also differentiate between folded and unfolded substrates by a proofreading mechanism. This was demonstrated in studies using alkaline phosphatase (PhoA) fused to Tat specific signal peptides. PhoA requires disulfide bonds for proper folding and therefore it was not exported by the Tat system because the proteins were unfolded in the reducing cytoplasm that perturbs the formation of disulfide bonds, which are required for folding. It was demonstrated that PhoA can serve as a Tat substrate when expressed in *E. coli* strains with an oxidising cytoplasm (DeLisa *et al.,* 2003). Similar studies demonstrated efficient export of active, correctly folded disulfide bonded therapeutic proteins to the periplasm such as: PhoA, anti-interleukin 1β (scFv) and human growth hormone (hGH) (Matos *et al.*, 2014). In contrast, other studies have shown that the Tat system tolerates minor structural changes in disulfide-bonded therapeutic proteins (Alanen *et al.*, 2015). Several disulfide bond-containing proteins, such as human interferon α2b (IFN), hGH, scFv and an antibody VH domain, are efficiently exported in the absence of cytoplasmic disulfide bond formation. In

addition, it was demonstrated that hGH can even be exported by the Tat system when disulfide bond formation is perturbed by substitution of the cysteine residues involved. This data suggest that the substrates adopt a near-native structure in the absence of disulfide bond formation that enables them to avoid the Tat proofreading mechanism. A better understanding of the margin for error remains to be elucidated.

A hydrophobicity sensor employed by the Tat apparatus rather than assessment of the folded state of substrates has been suggested (Richter *et al.*, 2007). The hydrophobic interior of a substrate could be revealed in unfolded or poorly folded substrates. There is a possibility that the Tat components sense the hydrophobic interior, which is normally sequestered in correctly folded state. Studies demonstrated that introducing hydrophobic surface patches abolish translocation by the Tat system (Richter *et al.*, 2007). Strikingly, other studies demonstrated that the Tat system tolerates significant changes in substrate surface hydrophobicity (Jones *et al.*, 2016). More recent studies suggested that the Tat systems proofreading mechanism has the ability to sense conformational flexibility rather than changes in surface charge or hydrophobicity (Jones *et al.*, 2016). We have recently shown that the Tat translocase is able to recognize and distinguish the folding state of a *de novo*-designed heme protein, denoted as BT6 maquette (Sutherland *et al.*, 2018). Because of the simple and repetitive protein structure this is a suitable candidate to study the Tat systems proofreading mechanism.

1.4 Recombinant protein production in *Escherichia coli*

The overall market value of the biopharmaceutical industry had an estimated sales value of $140 billion in 2014. Of all biotherapeutics licensed to date, *E. coli* is responsible for the production of more than 30% of the expression of therapeutic recombinant proteins (Overton, 2014). To produce therapeutic proteins in *E. coli*, several strategies have been used including: expression of soluble proteins in the cytoplasm, expression as soluble inclusion bodies, expression in the cytoplasm followed by export to the periplasm. For the latter, extraction of the protein can be done more easily and contaminating proteins, DNA and general debris is minimised. Moreover, the periplasm is essential for efficient disulphide bond formation. This is the only oxidising compartment in wild type cells. Many therapeutic proteins and antibody fragments contain disulphide bonds. Approximately 30% of all currently licensed, recombinantly expressed biotherapeutic products are produced to the periplasm by the general Secretory 'Sec' pathway. However, there are many heterologous proteins that cannot be exported by the Sec pathway. The Sec pathway is unable to export these proteins due to their rapid folding which makes it impossible to transport these folded proteins through a relatively narrow pore in a folded state. In contrast to the Sec pathway, the Tat pathway exports fully folded proteins. The Tat pathway is a powerful alternative by means of export of proteins that cannot be exported by the Sec pathway (Walker *et al.*, 2015). Besides, the Tat system does not only export folded proteins; it consists of a proofreading and a quality control mechanism (Palmer and Berks, 2012). Due to this unique capacity it is likely that the Tat system has a natural tendency to produce high quality, active products. This is particularly relevant for biotechnological exploitation.

1.5 Aims of this project

The overall aim of this study was to gain inside in the *E. coli* Tat proofreading ability based on the folding state and folding kinetics of Tat dependent secretory proteins. As described in this chapter, the Sec and Tat systems are the two general secretory pathways to transport proteins from the cytoplasm to the periplasm. The most described translocase system is the Sec machinery. Unfortunately, the Sec system is unable to transport proteins with rapid folding kinetics within the cytoplasm as the proteins are translocated via a relatively narrow pore. In contrast to the Sec system, the Tat system has the ability to discriminate between folding states. Therefore only fully folded proteins will be exported from the cytoplasm to the periplasm. It is not well understood how the Tat system senses the folding state of a substrate. The project will focus on the Tat system's proofreading ability by altering substrate surfaces to determine what Tat senses as 'correctly folded' in the substrate.

Furthermore, we were also interested to determine whether Tat substrates fold co- or post-translationally and whether they start interacting with the Tat translocase already during or after their synthesis. It is known that Tat dependent proteins use the Tat system because of their rapid folding kinetics. However, it is not known at what stage of the protein synthesis Tat substrates start to interact with the Tat complex. It is also not clear whether Tat dependent proteins prefer to fold post- or co-translationally. Cotranslational nascent chain force measurements provided insight into the folding kinetics of Tat dependent proteins.

Chapter 2:

Materials and methods

Fisher Scientific UK (Thermo Fisher Inc, USA), Sigma (Sigma-Aldrich, USA) or Formedium (UK) supplied all chemicals used in this study unless otherwise stated. L-[35S]-methionine was obtained from PerkinElmer.

2.1 DNA techniques

2.1.1 Preparation of plasmid DNA

Plasmid DNA was extracted and purified using the QIAprep spin miniprep kit (Qiagen, Hilden, Germany). The purifications were performed according to manufacture's instructions. The samples were eluted in 30 µl Elution buffer (10mM Tris-Cl pH 8.5) and the DNA concentration and purity was measured using NanoDrop 2000c (Thermo Fisher Scientific Inc, USA). The Plasmid DNA samples were stored at -20 °C.

2.1.2 Amplification Polymerase Chain Reaction (PCR)

PCR reaction mixes were prepared using Q5® High-Fidelity DNA Polymerase (New England Biolabs, NEB, UK). PCR solutions were made using 1 ng DNA, 5x Q5 reaction buffer, 0.4 µl 10mM dNTPs, 0.8 µl of the forward and reverse primer (10 µM), 0.1 µl Q5 High-Fidelity DNA Polymerase and dH_2O to a final volume of 20 µl. The PCR reaction was performed in a Biometra T3 Thermocycler (Biometra Anachem, UK) and analysed by agarose gel electrophoresis (2.1.4). The following thermocycling conditions were used for DNA amplification:

1= 98.0 °C	1:00 min	
2= 98.0 °C	20 sec	⎫ 29x
3= 60.0 °C	10 sec	⎭
4= 72.0 °C	20-30 sec/kb	
6= 4.0 °C	Pause	

2.1.3 Primers used for PCR

Eurofins MWG Operon Inc supplied all oligonucleotides used in this study. Table 1 lists primers for the polymerase chain reaction, the template DNA used and the resultant constructs. The restriction enzyme sites of the enzymes used in this study are underlined.

Table 1 Primers used for the polymerase chain reaction

Primer	Sequence (5' to 3')
KpnITorAF	CGGCGGGGTACCATGAACAATAACGATCTC
Mut1234XbalR	CGGCGGTCTAGACCTTAGTGGTGATGATGATG
Mut5XbalR	GCCGCCTCTAGACCTTAATGGTGATGATGATG
Mut7XbalR	GCCGCCTCTAGACCTTAATGATGATGATGGTG
Mut8XbalR	GCCGCCTCTAGACCTTAGTGATGATGATGATGGTG
Mut9XbalR	GCCGCCTCTAGACCTTAATGGTGATGATGATG
Mut10XbalR	GCCGCCTCTAGACCTTAATGATGATGATGGTG
BT6WTBamH1R	TTATTAGGATCCGTTAATGGTGGTGATGATGGTGC

2.1.4 Agarose gel electrophoresis

To prepare 1% agarose gels, 1% (v/w) agarose (Bio-Rad Laboratories Ltd, USA) was dissolved in 1x TAE buffer (0.04 M Tris-acetate, 0.001 M EDTA, pH 8.2). The gel was submerged in 1x TAE buffer after it was polymerised. 4 µl of the DNA samples (PCR or restriction digest product) was mixed with 1 µl SYBR Green Nucleic Acid Gel stain (20x in DMSO, Invitrogen, Thermo Fisher Scientific Inc, USA) and 1 µl 6x Gel Loading Buffer (Thermo Fisher Scientific Inc, USA). Agarose gel electrophoresis was carried out at 150V for 30 minutes or until the dye front was at least half way through the gel. The DNA bands were visualised using Bio-Rad Gel doc (Bio-Rad Laboratories Ltd, USA).

2.1.5 Purification of DNA from agarose gels

DNA from agarose gels was excised from the UV transluminator. The DNA of interest was then purified using a QIAprep Gel Extraction kit (Qiagen, Germany according to manufacturer's instructions. The concentration of the extracted DNA was measured using NanoDrop 2000c (Thermo Fisher Scientific Inc, USA) and samples were used immediately or stored at -20 °C

2.1.6 Restriction digest of DNA

Restriction digest reaction mix was prepared to digest DNA following the manufacturer's instructions (New England Biolabs, UK.). The first restriction enzyme was incubated at 37 °C for 30 minutes before further incubation with the second restriction enxyme for another 60 minutes. Table 2 shows restriction enzymes used in this study. The digested DNA products were isolated using agarose gel electrophoresis and DNA purification form agarose gels (2.1.4 and 2.1.5).

Table 2 Restriction enzymes used in this study

Restriction enzyme	Sequence (5' to 3')
NdeI	CATATG
BamHI	GGATCC
KpnI	GGTACC
XbaI	CTCGAG
DpnI	GATC

2.1.7 Ligation of DNA fragments into plasmid vector backbone

In this study, all ligation reactions were performed using T4 DNA ligase (Roche, Sussex, UK). The ligation reaction mix contained 1 µl ligase buffer, 1 µl T4 ligase and the insert and vector were ligated in a 3:1 (9 µl insert to 3 µl vector) ratio. The samples were incubated overnight at 4 °C and 10 µl ligation reaction was transformed into *E. coli* DH5α, XL1 blue or NEB Turbo competent cells (2.2.5).

2.1.8 Site-directed DNA mutagenesis

Site-specific DNA mutagenesis Site-specific DNA mutagenesis was carried out by using Q5® High-Fidelity DNA Polymerase (New England Biolabs, NEB, UK) as described in section 2.1.2. The thermocycling reaction generates circular DNA. To exclude methylated DNA, 0.5 µl DpnI (New England Biolabs, NEB, UK) was added to 20 µl of amplified DNA samples. The samples were incubated overnight at 37 °C. After DpnI digestion, 3 µl of the amplified DNA was transformed into 30 µl competent *E. coli* NEB turbo cells (section 2.2.5).

The mutagenesis occurred due to primers that were designed containing a mismatch in the centre of the forward primers. The forward primers included 12 nucleotides complementary to the plasmid at the 5' end. The desired nucleotide changes were incorporated in the centre of the primer followed by complementary nucleotides on the 3' side of the substitution. The number of nucleotides on the 3' side was based on an annealing temperature of ± 60 °C. The 5' end of the reverse primer was designed next to the 5' end of the forward primer and preceded in the opposite direction on the complementary strand. Table 3 shows the primers used for Site-specific DNA mutagenesis. The desired nucleotide sequence is underlined.

Table 3 Primers used for Site-specific DNA mutagenesis

Primer	**Sequence (5' to 3')**
QCTorARRfor	TTTCAGGCATCACGTCGGCGTTTTCTGGCACAACTCGGC
QCTorAKKfor	TTTCAGGCATCAAAGAAACGTTTTCTGGCACAACTCGG
QCTorAKKRRrev	TGATGCCTGAAAGAGATCGTTATTGT
QCSuflRRfor	ATGTCACTCAGTCGGCGTCAGTTCATTCAGGCATCGGGG
QCSuflRRrev	ACTGAGTGACATGGTATATCTCCTTCTTAAAGTT
Arrest-Cterm-FW	ATCCGTGCTGGCTAGGGGAGCGATTACATCAAGCGC
Arrest-FW	ATCCGTGCTGGCTAGGGGAGCTCCGATAAGCAAGAAG
Arrest-RV	GCCAGCACGGATGCCT
FL-Cterm-FW	ATCCGTGCTGGCGCAGGGAGCGATTACATCAAGCGC
FL-FW	ATCCGTGCTGGCGCAGGGAGCTCCGATAAGCAAGAAG

2.1.9 Primers used for truncation PCR

To delete nucleotides in a specific plasmid region, forward primers were designed that flank the region to be deleted. The forward primers included 12 nucleotides complementary to the plasmid at the 5' end of the to be truncated region. The 12 nucleotides were followed by complementary nucleotides on the 3' side of the to be truncated region. The number of nucleotides on the 3' side was based on an annealing temperature of ± 60 °C. The 5' end of the reverse primer was designed back-to-back to the 5' end of the forward primer and preceded in the opposite direction on the complementary strand. Table 4 shows the primers used for truncation PCR.

Table 4 Primers used for truncation PCR

Primer	Sequence (5' to 3')
Trunc_L50_FW	TCAGGATCGGGCGACAACAGCGCGGACAGC
Trunc_L55_FW	TCAGGATCGGGCATGGGCGACAACCGCGA
Trunc_L40_FW	TCAGGATCGGGCTGCCGGAAGCGAATTGG
Trunc_L30_FW	TCAGGATCGGGCTACCCATACGATGTTCCAGATTACGCT
Trunc_BT6_RV	GCCCGATCCTGACAATTGCTTCA
Trunc_SufI_RV	GCCCGATCCTGACGGTAC
Tunc_hGH_RV	GCCCGATCCTGAGAAGCCAC
BT6_L26_FW	CTGAAGCAATTGTACCCATACGATGTTCCAGATTACGCT
BT6_L26_RV	CAATTGCTTCAGATCTTCAAATTGGTTCAGC
hGH_L26_FW	AGCTGTGGCTTCTACCCATACGATGTTCCAGATTACGCT
hGH_L26_RV	GAAGCCACAGCTGCCCTCC
SufI_L26_FW	AATCCGGTACCGTACCCATACGATGTTCCAGATTACG
SufI_L26_RV	CGGTACCGGATTGACTAACAGT

2.1.10 Primers used for insertion PCR

To insert nucleotides in a specific plasmid region, forward primers were designed that flank the region to be inserted. The forward primers included 12 nucleotides complementary to the plasmid at the 5' end of the to be inserted region. The desired nucleotides to be inserted were incorporated in the centre of the primer followed by complementary nucleotides on the 3' side of the to be inserted region. The number of nucleotides on the 3' side was based on an annealing temperature of ± 60 °C. The 5' end of the reverse primer was designed back-to-back to the 5' end of the forward primer and preceded in the opposite direction on the complementary strand. Table 5 shows the primers used for insertion PCR. The to be inserted region is underlined.

Table 5 Primers used for insertion PCR

Primer	Sequence (5' to 3')
BT6_L35_FW	TCAGGATCGGGCAATTGGATGAGCTCCTACCCATACGATGTTCC AGATTACGCT
BT6-L40_FW	TCAGGATCGGGCTTTGTGCCGGAAGCGAATTGGATGAGCTCCT ACCCATACGATGTTCCAGATTACGCT
BT6_L45_FW	TCAGGATCGGGCAGCCGTTACTGGGGCTTTGTGCCGGAAGCGA ATTGGATGAGCTCCTACCCATACGATGTTCCAGATTACGCT
pING_BT6_RV	GCCCGATCCTGACAATTGCTTCA
pING_hGH_RV	GCCCGATCCTGAGAAGCCAC
pING_SufI_RV	GCCCGATCCTGACGGTACC

2.1.11 Gibson assembly

Constructs (insert) were transferred into pING, pET or pEXT22 plasmids (vector) by using Gibson assembly. By using Q5® High-Fidelity DNA Polymerase (New England Biolabs, NEB, UK) insert and vector fragments were amplified and linearized as described in section 2.1.2. Methylated DNA was digested by adding 0.5 µl DpnI (New England Biolabs, NEB, UK) to 20 µl of amplified DNA samples. The samples were incubated overnight at 37 °C before both fragments were then joined using Gibson assembly® (Gibson *et al.*, 2009). 240 aliquots were prepared from 1.2 ml Gibson reaction mix and stored at -20 °C. The 1.2 ml Gibson reaction mix contained, 320 µl 5x isothermal (ISO) reaction buffer (25% PEG-8000, 500 mM Tris-HCl pH 7.5, 50 mM $MgCl_2$, 50 mM DTT, 1mM of the 4 different dNTPs and 5 mM NAD), 0.64 µl of 10 U/µl T5 exonulease, 20 µl of 2 U/µl Phusion polymerase, 160 µl of 40 U/µl Taq ligase and 700 µl dH_2O. To assemble the overlapping DNA fragments, 1.25 µl of the vector and 1.25 µl of the insert were added to 5 µl Gibson reaction mix and incubated for 1 h at 50 °C. 3 µl of the ligated DNA samples was transformed into 30 µl competent *E. coli* NEB turbo cells (section 2.2.5).

2.1.12 Primers used for Gibson assembly

To amplify the insert fragment, overhang oligonucleotides were used with 25 nucleotides to either the 5' and 3' ends of the sequence. This overhang was designed to be complementary with the respective 5' and 3' ends of the vector. To linearize the vector, standard forward and reverse primers were designed that flank the region of the insert. These primers were complementary to the overhang region of the primers to amplify the insert fragments. The preferred annealing temperature of the primers was 60 °C. In preparation of the co-translational folding studies, the primers used to amplify the insert

fragments for co-translational folding studies are shown in table 6. The primers to linearize and amplify the vector fragments are listed in table 7. To engineer constructs into pEXT22 for fractionation assays, the primers used to amplify the insert are shown in table 8. The primers to linearize and amplify the vector fragments are listed in table 9.

Table 6 Primers to amplify the insert fragments for co-translational folding studies

Primer	**Sequence (5' to 3')**
L60_BT6_FW	TCTGAAGCAATTGTCAGGATCGGGCGGACAATACTTCATG ATGGGCGACA
L60_hGH_FW	CAGCTGTGGCTTCTCAGGATCGGGCGGACAATACTTCATGA TGGGCGACA
L60_SufI_FW	CAATCCGGTACCGTCAGGATCGGGCGGACAATACTTCATG ATGGGCGACA
L60_RV	CGTAATCTGGAACATCGTATGGGTAGGAGCTCATCCAATTC GCTTCCG
Insert_SufI_KK_FW	TTTAACTTTAAGAAGGAGATATACCATGTCACTCAGTAAGA AGCAGTTCATTCAG
Insert_SufI_RR_FW	TTTAACTTTAAGAAGGAGATATACCATGTCACTCAGTCGGC GTCAG
Insert_TorA_FW	TTTAACTTTAAGAAGGAGATATACCATGAACAATAACGAT CTCTTTCAGGCATCAAAG
Insert_lep_RV	GCGCAAACGTGAACGAAGATGGCTATTAATGGATGCCGCC AATGCGA
SufI_L40_FW	GCAACTGTTAGTCAATCCGGTACCGTCAGGATCGGGCTTTG TGCCG
SufI_L40_RV	CCGCGCGCTTGATGTAATCGCTCCCAGGGCCAGCACGGAT GC

Table 7 Primers to amplify the vector fragments for co-translational folding studies

Primer	Sequence (5' to 3')
pING_HA_FW	TACCCATACGATGTTCCAGATTACGCT
BT6_L26_RV	CAATTGCTTCAGATCTTCAAATTGGTTCAGC
hGH_L26_RV	GAAGCCACAGCTGCCCTCC
SufI_L26_RV	CGGTACCGGATTGACTAACAGT
pING_FW	GAAGCAGAAAGAAGGTAAGAAACG
pING_RV	CCTTTCGTCTTCAAGAATTCC
pET_FW	TAGCCATCTTCGTTCACGTTTGCG
pET_RV	GGTATATCTCCTTCTTAAACAAAATTATTTCTAGAGGGG
SufI_L40_FW	CGGTACCGGATTGACTAACAGTTGC
SufI_L40_RV	GGGAGCGATTACATCAAGCGCG

Table 8 Primers to amplify the insert fragments for fractionation assays

Primer	Sequence (5' to 3')
TorApEXT22_FW	GGTACCCTACCACAGAGGAACATGTATGAATAACAACGACCT GTTTCAGGC
ADR1a_RV	GACGGAGCTCGAATTCGGATCCTTATTAGTGATGGTGATGATG ATGGTTACCG
R16_RV	GACGGAGCTCGAATTCGGATCCTTATTAATGATGATGGTGATG ATGTTCCAGGC
SOD_RV	GACGGAGCTCGAATTCGGATCCTTATTAGTGATGGTGGTGATG ATGCTGT
ProteinG_RV	CGACGGAGCTCGAATTCGGATCCTTATTAGTGATGGTGATGAT GATGTTCGGT

Table 9 Primers to amplify the vector fragments for fractionation assays

Primer	Sequence (5' to 3')
pEXT22_FW	TAAGGATCCGAATTCGAGCTCCG
pEXT22_RV	ACATGTTCCTCTGTGGTAGGGT

2.1.13 Sequencing of plasmid DNA

Subsequent sequencing was done to confirm successful site-directed mutagenesis and Gibson assembly. The sequencing was carried out by GATC sequencing service. The primers used for sequencing are shown in table 9.

Table 10 Primers used for sequencing

Primer	Sequence (5' to 3')
T7-forward	TAATACGACTCACTATAGGG
T7-reverse	CTAGTTATTGCTCAGCGGT
Ptac-forward	GAGCGGATAACAATTTCACACAGG
Ptac-reverse	AAAAGGCCATCCGTCAGGATG
pING-forward	GAAGCAGAAAGAAGGTAAGAAACG
pING-reverse	CCTTTCGTCTTCAAGAATTCC

2.1.14 Plasmids used in this study

The plasmids used and engineered in this study are shown in the appendices. The plasmids used in the BT6 variants studies are shown in appendix I. The BT6 constructs used in this are all in a pEXT22 plasmid containing a pTac promoter, which is IPTG inducible (appendix III). The pEXT22 plasmids contain a kanamycin resistance cassette. The

constructs used in the co-translational folding studies can be found in appendix II. These constructs are in pING (appendix V) or pET plasmids (appendix IV) that both contain an ampicillin resistance cassette and are inducible with arabinose.

2.2 Growth and maintenance of *E. coli* cultures

2.2.1 *E. coli* strains used in this study

The *E. coli* strains presented in table 11 were used in this study.

Table 11 Strains used in this study

Strain	Description	Reference
W3110	*lambda-IN(rrnD-rrE)1 rph-1*	(Hayashi *et al.*, 2006)
W3110 TE2	W3110 carrying a ptac promoter upstream of *tatABCD E.*	(Browning *et al.*, 2017)
MC1061	*araD139 Del(araA-leu)7697 Del(lac)X74 galK16 galE15(GalS) lambda- e14- mcrA0 relA1 rpsL150(strR) spoT1 mcrB1 hsdR2*	(Casadaban and Cohen, 1980)
DH5α	*supE44 ΔlacU169(ϕ80lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	ThermoFisher Scientific, UK
XL1 blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac*	Stratagene, USA
NEB Turbo	*proA$^+$B$^+$ lacI^q ΔlacZM15 / fhuA2 Δ(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)* Tet^S *endA1 thi-1 Δ(hsdS-mcrB)5*	NEB, UK

2.2.2 Media

For growth of *E. coli* in liquid media, Luria Bertani (LB) medium (10 g/L sodium chloride, 10 g/L tryptone and 5 g/L yeast extract) was frequently used. For growth on agar plates, Luria Bertani agar (LBA) (10 g/L sodium chloride, 10 g/L tryptone, 5 g/L yeast extract and 10 g/L bacto-agar) was used. In section 2.8.1, the *E. coli* cells were grown in M9 medium (1x M9 salts, 0.1 mM $CaCl_2$, 0.2% casamino acids, 0.2% thiamine, 2% fructose and 2 mM $MgSO_4$). Suitable antibiotics to the vector were used at the following concentrations: Ampicillin (100 μg/mL) and Kanamycin (50 μg/mL).

2.2.3 Glycerol stocks

The constructs used in this study were transformed in NEB Turbo cells and stored as glycerol stocks at – 80 °C. Glycerol stocks were prepared by adding stationary phase cells to LB media containing 17% glycerol.

2.2.4 Preparation of competent cells

To prepare the cells to be made competent, 5 mL of LB was inoculated and incubated overnight at 37 °C shaking at 220 rpm. 10 mL LB was inoculated with 100 μl of the overnight culture and incubated at 37 °C shaking at 220 rpm until OD_{600} = 0.3 – 0.4. Once the cells reached the desired OD_{600} they were centrifuged for 10 min at 3000g at 4 °C (Thermo Scientific Heraeus Megafuge 16R, UK). The pellet was resuspended in 100 mM ice cold $MgCl_2$, and incubated on ice for 5 min. The cells were repelleted as in the previous step and resuspended in 1 mL ice cold 100 mM $CaCl_2$ and incubated for 2 – 24 hours at 4 °C. The cells were used fresh or stored at -80 °C.

2.2.5 Transformation of competent cells

In a pre-cooled Eppendorf tube, 0.3 µl of plasmid DNA was gently mixed with 30 µl of competent cells (section 2.2.4 and 2.2.1) and incubated on ice for 20 minutes. Following incubation, the cells were transferred to a 42 °C water bath for 45 sec, after which they were shortly incubated on ice. After the heat shock, 0.3 mL of LB was added to the cells followed by incubation at 37 °C for 20 min. Following incubation, the cells were harvested by centrifugation at 7000 rpm for 2 minutes (Eppendorf centrifuge, 5417R, Hamburg), the supernatant was removed and the resuspended pellet was plated on LBA plates containing the appropriate selective antibiotics (section 2.2.2). The LBA plates were incubated overnight at 37 °.

2.3 Protein production

2.3.1 Culture of *E. coli* and plasmid induction

Prior to fractionation, *E. coli* strains carrying the respective plasmids were grown in 5 ml LB medium containing the appropriate antibiotics (section 2.2.2) by inoculating with a single colony and grown overnight at 37 °C, shaking at 220 rpm. 50 mL of fresh LB medium containing the appropriate antibiotics was then inoculated with the overnight culture to OD_{600} = 0.05. Cultures were then grown at 37 °C, 200 rpm to OD_{600} = 0.5 followed by induction with IPTG. Typically 3 hours post induction, an amount equivalent to OD_{600} = 10 was collected and fractionated into Cytoplasm (C), Membrane (M), and Periplasm (P) fractions (section 2.3.2). The periplasmic fraction was collected using the EDTA/lysozyme/cold osmotic shock and centrifugation method (Matos *et al.*, 2014). Cells

used in the BT6 variant studies were induced with 0.5 mM IPTG and cells used for co-translational folding studies were induced with 100 µM IPTG.

2.3.2 Fractionation of the *E. coli* cells

After the induction period, the cells equal to OD10 were harvested by centrifugation at 3000 rpm for 10 min at 4 °C (Eppendorf centrifuge, 5417R, Hamburg). The pellets were resuspended in ice cold 500 Buffer 1 (100 mM Tris-acetate pH 8.2, 500 mM sucrose, 5 mM EDTA). In addition, 500 µl ice cold MilliQ H_2O and 40 µl lysozyme (from egg-white, 1 mg/mL stock (Sigma Aldrich, UK)) was added and the suspension was incubated on ice for 5 min. To prevent contamination of the periplasmic fraction (P) from the spheroplast, 20 µl 1 M $MgSO_4$ was added to stabilise the inner membrane. The samples were centrifuged at 14.000 rpm for 2 min at 4 °C. The spheroplast pellet was then washed by resuspension in 1 mL ice cold buffer 2 containing 50 mM Tris-Acetate (pH 8.2), 250 mM sucrose and 10 mM $MgSO_4$ and centrifuged at 14,000 rpm for 5 min at 4 °C. The supernatant was discarded and the resulting pellet was then resuspended in ice cold buffer 3 containing 50mM Tris-Acetate, 2.5 mM EDTA (pH 8.2) and sonicated for 6 x 10 s, 8 µm amplitude (Soniprep 150plus, Sanyo Gallenkamp, Loughborough, UK). The sonicated samples were then centrifuged at 70,000 rpm, 4 °C, 30 min (Beckmann TL100, TLA100.3 rotor). The supernatant was taken as the cytoplasmic fraction (C) and the pellet was resuspended in buffer 3 containing 50 mM Tris-Acetate, 2.5 mM EDTA (pH 8.2) to obtain the Membrane fraction (M). All cell fractions were stored frozen at – 20 °C.

2.3.4 Time course assay

Samples were cultured in LB containing the appropriate antibiotics and induced with IPTG as described in section 2.3.1. 1, 3 and 5 hours post induction, an amount equivalent to OD_{600} = 10 was collected and fractionated into Cytoplasm (C), Membrane (M), and Periplasm (P) fractions (section 2.3.2).

2.3.4 Export assay for purification of BT6 variants

W3110 'Tat Express' *E. coli* strains carrying the respective plasmids were grown in 50 ml LB medium overnight at 37 °C, 200 rpm (Browning *et al.*, 2017). 400 mL fresh LB medium containing the appropriate antibiotics was inoculated with the overnight culture to OD_{600} = 0.05. Cultures were then grown at 30 °C, 200 rpm to OD_{600} = 0.5 before induction with 0.5 mM IPTG and cultures were incubated for 24 h. Following incubation, the cells were harvested by centrifuging at 4000 rpm for 20 min at 4 °C (Beckman Avanti J-25, JA-10 rotor). To obtain the periplasmic fraction, cells were resuspended in 10 mL buffer containing 100 mM Tris-acetate pH 8.2, 500 mM sucrose, 5 mM EDTA pH 8.0 and 10 mL ice-cold dH_2O before the addition of 800 µl hen egg white lysozyme (1 mg/mL). The samples were then incubated on ice for 10 min. To stabilise the inner membrane 800 µl 1 M $MgSO_4$ was added to the solution before centrifuging at 4000 rpm for 20 min at 4 °C (Beckman Avanti J-25, JA-25.5 rotor). The supernatant was taken to collect the periplasmic fraction. The spheroplast pellet was resuspended in 10 mL 0.5 M Sodiumphosphate buffer (pH 7.3) and a protease inhibitor (cOmplete Mini Protease Inhibitor, Roche) was added. The cells were sonicated for 6 x 30 s, 8 µm amplitude (Soniprep 150plus, Sanyo Gallenkamp, Loughborough, UK) and centrifuged (14,000 rpm, 20 min, 4 °C). The supernatant was collected as the spheroplast fraction.

2.4 Protein purification

2.4.1 Immobilised Metal Affinity Chromatography (IMAC), Nickle

An IMAC Sepharose Fast Flow Column (GE Healthcare) was used to purify the BT6 variants. After equilibration of the column with 20 mM Sodium Phosphate, periplasmic or spheroplast fractions were applied to the column. The column was washed with 20 mL of buffer containing 25 mM imidazole, 0.5M NaCl, 20mM Sodiumphosphate, pH 7.3. BT6 variants were eluted with 9 mL buffer containing a gradient of 100 – 200 mM imidazole, 50 mM EDTA, 20 mM Sodiumphosphate pH 7.3. The elution fractions were collected and concentrated using Vivaspin centrifugal concentrators (Sartorius).

2.4.2 Protein quantification assay

Proteins of interest were purified as described in section 2.3.4 and 2.4.1. The protein concentration was quantified by a Bradford assay (Bio-Rad DC™ Protein Assay, Herts, UK) according to the manufacturer's instructions. In several cases the protein concentration was calculated by using the molar extinction co-efficient and reading A_{280} from a quartz cuvette (Starna Scientific Ltd, Essex, UK).

2.4.3 De-salting IMAC peak fractions

In order to de-salt the IMAC peak fractions, by dialysis tubing, samples were dialysed overnight at 4 °C into 50 mM sodium phosphate buffer overnight at 4 °C (SnakeSkin® dialysis tubing, Thermo scientific). To de-salt smaller volumes (0.5 mL to 3 mL), dialysis cassettes were used (Slide-A-Lyzer® Dialysis Cassette, Thermo Scientific, UK).

2.4.4 Concentrating IMAC peak fractions

Dialysed IMAC peak fractions were concentrated to an end volume between 0.5 and 1.0 mL by centrifuging the samples in VIVASPIN 20 10,000MWCO columns (Santorius, Gloucestershire, UK) at 3000 rpm.

2.5 Protein resolution

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

Protein separation was carried out using the PROTEAN® Tetra System (Bio-Rad, UK) according to the manufacturer's instructions. The 15% 1.5 mm thick separation gel contained: 15% acrylamide, 0.3% bis-acrylamide (37.5:1, Bio-Rad, UK), 375 mM Tris-HCl pH 8.85, 0.1% SDS, 0.1% APS and 0.06% TEMED. The 5% stacking was composed of: 5% acrylamide, 0.0375% bis-acrylamide, 125 mM Tris-HCl pH 6.8, 0.1% SDS, 0.6% APS and 0.06% TEMED. Proteins samples were prepared by diluting aliquots 6:1 with 6x loading dye (125 mM Tris-HCl pH 6.8, 20% glycerol, 4% SDS, 0.02% bromophenol blue, 5% β-mercaptoethanol) and heated at 95 °C for 10 min. The gels were run at 60 mA until the dye front migrated off the gel, typically 50 min. The protein gel running buffer was composed of 25 mM Tris, 192 mM glycine and 0.1% SDS.

2.5.2 Detection of proteins with Coomassie

The separated protein bands were visualised with Coomassie stain (0.05% w/v Coomassie Brilliant Blue R-250 (Bio-Rad, UK), 50% Ethanol, 1.74 M Acetic acid). The gels were stained with Coomassie stain for 1 hour at room temperature on a shaker. Following

staining, the gels were destained by incubating at room temperature on a shaker with D-stain (5% Ethanol, 7.5% Acetic acid) until the background was clear.

2.6 Western blotting

Protein samples separated by SDS-PAGE as described in section 2.5.1, were transferred to a PVDF membrane (GE healthcare, UK) using a Mini Trans-Blot® Cell system (Bio-Rad, UK) with transfer buffer composed of 192 mM glycine, 25 mM Tris, 10% ethanol). The PVDF membrane was prepared by soaking in methanol prior to contact with the gel. The transfer was carried out for 1 hour at 80V and after transfer the membranes were submerged in blocking milk (2.5% skimmed mild powder in 0.1% PBS-Tween20, 3% BSA) and incubated for at least 1 hour at 4 °C.

The blocked membranes were washed with 1x PBS-tween20 (0.1%) before incubation with 3.5 µl 6x-His C-terminal Tag HRP antibody (New England Biolabs, UK) in 20 mL 1x PBS tween20 (0.1%) for 1 hour. Membranes were washed and proteins were visualised using an ECL (enhanced chemiluminescence) kit (Biorad, Herts, UK) according to the manufacturer's instructions. A Bio-Rad chemiluminescence imager and corresponding software was used to develop the membranes.

2.7 1D-1H-NMR spectra

1D-1H NMR spectra of BT6, BT6M1, TorA-BTM0 and BT6-2Asp were acquired on a Bruker Advance III 600MHz spectrometer with a cryogenically cooled probe at 298K. Water supressed spectra were acquired using a Double Pulse Field Gradient Selected Excitation Experiment (DPFGSE) (Hwang and Shaka, 1995). All experiments were acquired with the same parameters. Spectra were acquired using 2048 complex points with a spectral width of 9615Hz, an inter-scan delay of 1s and 2048 repetitions for improved signal to noise. Data were processed in topspin 3.5pl7. All spectra were zero filled once, apodised with an exponential function (LB=3) and the residual water signal was supressed with a convolution based solvent filter (quadrature Gaussian window of width 0.2ppm) (Marion *et al.*, 1989).

2.8 Pulse labelling analysis

2.8.1 *In vivo* pulse labelling analysis

Competent *E. coli* MC1061 cells were transformed with the respective pING1 plasmid (Dalbeys and Wickner, 1986) and grown overnight at 37°C in M9 minimal medium (section 2.2.2). Cells were diluted into fresh M9 medium to an OD_{600} of 0.1 and grown until an OD_{600} of 0.3. Expression of the constructs was induced with 0.2% (w/v) arabinose and continued for 5 min at 37°C. Proteins were then radiolabeled with ^{35}S-methionine for 2 min at 37°C before the reaction was stopped by adding ice-cold trichloroacetic acid (TCA) to a final concentration of 10%. Samples were put on ice for 30 min and precipitates were centrifuged for 10 min at 14.000 rpm at 4°C (Eppendorf centrifuge, 5417R). After one wash with ice-cold acetone, centrifugation was repeated and pellets were subsequently

solubilized in Tris-SDS buffer (10 mM Tris-HCl pH 7.5, 2% (w/v) SDS) for 5 min while shaking at 1300 rpm at 37°C (Thermomixer comfort, Eppendorf). Another centrifugation for 5 min at 14.000 rpm was performed to remove insoluble material. The supernatant was then added to a buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1 mM EDTA-KOH, 2% (v/v) triton X-100, and supplemented with GammaBind™ G Sepharose™ (GE, Healthcare, UK). After 15 min incubation on ice, proteins that bind un-specifically to G Sepharose were removed by centrifugation as before. The supernatant was used for immunoprecipitation using G Sepharose an anti-HA. 11 Epitope Tag antibody (Nordic BioSite, Sweden). The incubation was carried out at 4°C whilst rolling. After centrifugation for 1 min immunoprecipitates were washed with 10 mM Tris-Cl pH 7.5, 150 mM NaCl, 2 mM EDTA and 0.2% (v/v) triton X-100 and subsequently with 10 mM Tris-Cl pH7.5. Samples were spun down again and pellets were solubilized in SDS sample buffer (67 mM Tris, 33% SDS, 0.012% bromophenol blue, 10 mM EDTA-KOH pH 8.0, 6.75% glycerol, 100 mM DTT) for 10 min while shaking at 1300 rpm. Additionally, solubilized proteins were incubated with 0.25 mg/ml RNase for 30 min at 37°C and subsequently separated by using SDS-PAGE. Gels were fixed in 30% (v/v) methanol and 10% (v/v) acetic acid) and dried by using the Bio-Rad gel dryer model 583. Radiolabelled proteins were detected by exposing dried gels to film plates, which were scanned in a Fuji FLA-3000 scanner. Band intensities were generated and quantified with ImageJ and EasyQuant, respectively. Data was collected from two independent biological replicates, and standard errors were calculated.

2.8.2 *In vitro* pulse labelling analysis

In vitro translation was performed using the PURE system (NEB, UK) following the manufacturer's instructions. PCR products were used as the template DNA by using the primers listed in table 12. Amplification of the template DNA fragments was carried out as

described in section 2.1.2. ^{35}S-methionine labelled proteins were separated by SDS-PAGE and analysed by phosphor imaging as described in section 2.8.1.

Table 12 Primers used for *in vitro* pulse labelling analysis

Primer	**Sequence (5' to 3')**
T7 promoter primer	CCCGCGAAATTAATACGACTCACTATAGGG
T7 Terminator	GCTAGTTATTGCTCAGCGG

Chapter 3:

Probing the quality control mechanism of the twin-arginine translocase with folding variants of a *de novo*-designed heme protein.

3.1 Introduction

Escherichia coli is used for the production of more than 30% of recombinant therapeutic proteins. To produce therapeutic proteins in *E. coli*, several strategies have been used. Export to the periplasm is preferred because extraction of the protein can be done more easily and contaminating proteins, DNA and general debris is minimised. Approximately 30% of all currently licensed, recombinantly expressed biotherapeutic proteins are translocated to the *E. coli* periplasm by the general Secretory 'Sec' pathway. However, there are many heterologous proteins that cannot be exported by the Sec pathway. The Sec pathway can only transport proteins in their unfolded state since more bulky conformations cannot fit through a relatively narrow channel of the Sec translocon. In contrast to the Sec pathway, the Twin-arginine translocase (Tat) exports fully folded proteins. Therefore, the Tat pathway is known for exporting periplasmic proteins, which require co-factor insertion in the cytoplasm, such as FeS and molybdopterin. (Berks, 1996).

In *E. coli* the Tat translocase consists of three integral membrane proteins TatA, TatB and TatC. Tat dependent substrates are characterised by an N-terminal signal peptide containing a twin arginine (RR) motif. After translocation of the precursor substrate to the periplasm, the signal peptide is cleaved off by signal peptidase. Alterations of the RR motif to a lysine pair perturbs Tat-dependent export to the periplasm (Alami *et al.*, 2003, Stanley *et al.*, 2000).

The Tat system consists of a proofreading and a quality control mechanism, which only allows export of folded proteins. Various studies have demonstrated the Tat system's ability to reject unfolded proteins (Robinson *et al.*, 2011, Maurer *et al.*, 2009). However, emerging evidence suggests that quality control involves more than the proofreading of the

substrates folding state as misfolded substrates rapidly degrade after interaction with the Tat components (Matos *et al.*, 2008). Other studies have shown that the Tat system tolerates minor structural changes (Alanen *et al.*, 2015). Furthermore, studies seem to suggest a hydrophobicity sensor employed by the Tat apparatus rather than assessment of the folded state of substrates (Richter *et al.*, 2007). However, other studies demonstrated that the Tat system tolerates significant changes in substrate surface hydrophobicity (Jones *et al.*, 2016). Evidently, the Tat proofreading mechanism is not well understood.

The aim of this study was to test whether the *E. coli* Tat system could recognize and export a *de novo* designed di-heme protein. *De novo* designed proteins can be designed containing specific sequence information for a given function but are much simpler than natural substrates as natural substrates have accumulated structural complexity by undergoing natural selection (Degrado *et al.*, 1989, Edelman and Gally, 2001). Altering natural substrate residues can have unpredictable effects on protein function and conformation; whereas the effect of altering residues of a *de novo* designed substrate is more predictable. The first *de novo* designed maquette family was designed in order to study proton coupling to heme oxidation/reduction. To create a model substrate with oxidoreductase capabilities, a family of simplified model proteins was designed based on a heptad repeat four-helix bundle with inserted cofactor ligation sites (Shifman *et al.*, 1998). These four-helix bundle substrates are composed of two identical di-helical subunits accommodating one heme per subunit by bis-histidine ligation. Two di-helical subunits linked by a disulfide bond formation form the four-helix bundle (Grosset *et al.*, 2001). More recently, four α-helix maquette substrates were designed where the disulfide bond was removed and the N-terminus of one subunit was fused to the C-terminus of the second subunit (Lichtenstein *et al.*, 2013).

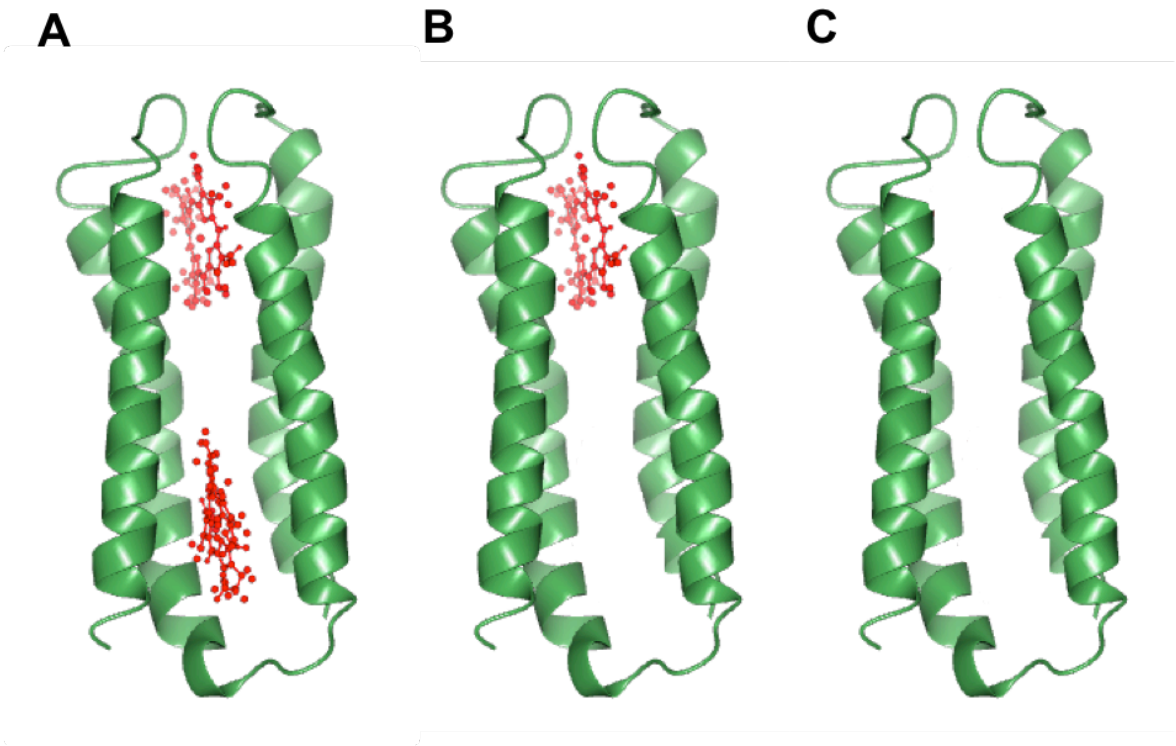


Figure 9 Structural models of BT6 proteins used in this study

(A) BT6 coordinates two heme *b* (red) cofactors. **(B)** BT6M1 coordinates a single heme *b* due to removal of one coordinating histidine residue. **(C)** BT6M0 coordinates no heme *b* due to removal of two coordinating histidine residue.

The so-called BT6 substrate is entirely synthetic and is not adapted from nature. The structure of the maquette BT6 substrate we use is shown in figure 9A. BT6 shares no sequence identity with natural proteins and the complexity was minimized to increase engineering freedom (Farid *et al.*, 2014). The protein is water-soluble and consists of four α-helix loops that form a hydrophobic cavity that accommodates two cofactors. The BT6 interior maintains *bis*-histidine ligation sites for binding heme *b*. BT6 needs to fold and acquire its heme *b* cofactors in the cytoplasm prior to interaction with the Tat translocase in order to be accepted by the Tat proofreading mechanism. The sequences of the BT6 maquette variants used in this study are listed in table 13. By substituting the heme coordinating histidine residues, the four-α-helix bundle BT6 maquette proteins were engineered to bind two, one or no heme *b* cofactors, to perturb folding (Figure 9). This was

assessed with spectroscopic analysis and one-dimensional 1H-NMR spectroscopy of purified maquette heme binding variant proteins. An N-terminal TorA signal peptide was fused to the BT6 variants for the Tat system to recognise them as Tat dependent substrates. The data suggest that the Tat system can sense the folding conformation and conformational flexibility of the BT6 substrates due to the observed correlation between the export efficiency and the folded state of the proteins. We proposed that the *de novo*-designed BT6 substrate is an ideal candidate to study the Tat systems proofreading mechanism.

Table 13 Amino acid sequences of maquettes

Name	**Sequence**	**Details**
BT6	GGDGENLYFQG	Di-heme
	EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG	
	EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG	
	EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG	
	EIWKQ**H**EDALQKFEEALNQFEDLKQLHHHHHH	
BT6M1	GGDGENLYFQG	Single heme
	EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG	
	EIWKQAEDALQKFEEALNQFEDLKQLGGSGSGSGG	
	EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG	
	EIWKQ**H**EDALQKFEEALNQFEDLKQLHHHHHH	
BT6M0	GGDGENLYFQG	No heme
	EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG	
	EIWKQAEDALQKFEEALNQFEDLKQLGGSGSGSGG	
	EIWKQAEDALQKFEEALNQFEDLKQLGGSGSGSGG	
	EIWKQ**H**EDALQKFEEALNQFEDLKQLHHHHHH	

3.2 Results

The Tat system can recognize and export a de novo designed di-heme protein

The first aim of this study was to test whether the *E. coli* Tat system could recognize and export a *de novo* designed BT6 maquette protein. Furthermore we wanted to test if the Tat proofreading mechanism can sense conformational differences. Two BT6 variants were designed to test whether the Tat system can recognize and export a *de novo* designed BT6 maquette protein, coordinating 2 (BT6) or no (BT6M0) heme *b* co-factors (Table 13). Figure 9A and C show the predicted secondary structure of the BT6 variants. BT6M0 was created in order to perturb folding by substituting two histidine residues to alanine (H53A/H88A). In table 13 the histidine residues that ligate heme are highlighted in bold. In order to be recognized and translocated by the Tat system, an N-terminal TorA-signal peptide was fused to the BT6 maquette protein followed by a C-terminal 6xHis tag (Table 13). The BT6 maquette variants were cloned into a low copy pEXT22 vector and expressed in *E. coli* W3110 cells. Sutherland et al., (2018) showed that the expression of BT6 in a high copy plasmid massively increased the production of recombinant BT6 protein resulting in contamination of the periplasm with BT6 protein.

The expression of the BT6 variants was induced with 0.5 mM IPTG. 3 hours post induction the cells were harvested and fractionated into cytoplasmic, membrane and periplasmic fractions (C/M/P). Figure 10 shows that the di-heme binding BT6 was efficiently exported to the periplasm whereas the no heme binding BT6 variant was rejected by the Tat system. There is a mature sized band of BT6 (17 kDa) observed in the periplasmic fraction (P). A faint unprocessed precursor BT6M0 band (21 kDa) is observed

in the cytoplasmic (C) fraction. The overall weak signal intensity of the precursor band is presumably due to degradation in the cytoplasm.

However, in order to determine whether the Tat system rejected BT6M0 system, additional experiments are required. There is a possibility that the BT6M0 substrate aggregated and is present in inclusion bodies. The fractionation procedure used in this study does not visualise the insoluble fraction separately. Possible insoluble proteins are now present in the membrane fraction. By using a different fractionation procedure a fourth cell fraction can be separated from the other three fractions, the insoluble fraction.

Previous studies have shown that the Tat system tolerates Tat dependent signal peptides with a RR or KR motif. However, it does not recognize the signal peptides when the twin arginine motif is substituted by, a twin lysine motif (Buchanan *et al.*, 2002). In order to confirm that the export of BT6 was mediated by the Tat system, a signal peptide with a KR (R12K) and a KK motif (R12K/R13K) were made. These TorA signal peptide variants were fused to the N-terminus of the di-heme binding BT6 variant. These constructs are denoted as TorA-KR-BT6 and TorA-KK-BT6. Table 14 lists the amino sequences of the TorA signal peptides used in this study. The different RR, KR and KK motifs are in bold and the mature regions are underlined.

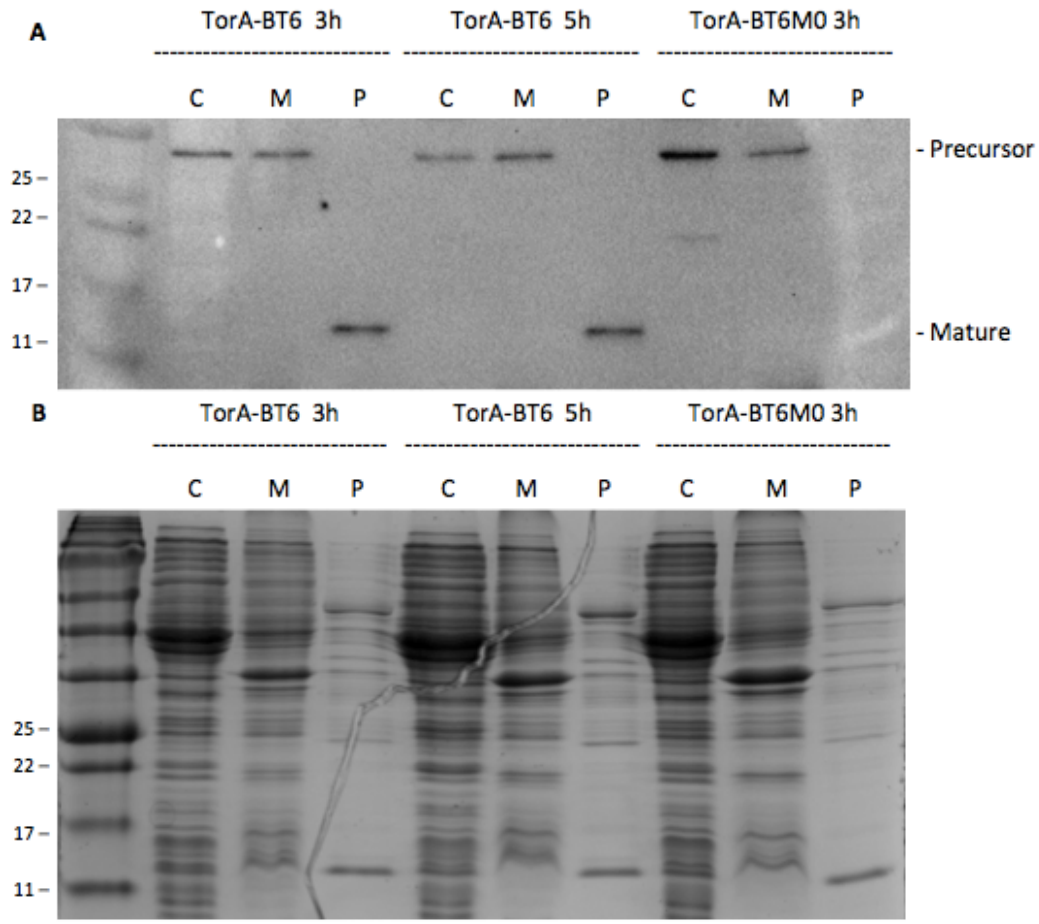


Figure 10. The export of BT6 and BTM0 determined by western blotting

The two samples on the left display the expression of the di-heme binding BT6 3 and 5 hours post induction, whereas the third sample demonstrates the expression of the no heme binding BT6M0 3 hours post induction. The cells were fractioned to cytoplasmic (C), membrane (M) and periplasmic (P) fractions. **(A)** The fractions were immunoblotted using a C-terminal His antibody. The molecular mass markers (kDa) are shown on the left and the expected size of precursor and mature bands on the right side of the blot. BT6 shows a mature sized band (17kDa) in the periplasmic fraction (P). A faint precursor-sized band (21 kDa) can be seen in the cytoplasmic fraction, which is, BT6M0 maquette precursor protein. **(B)** Coomassie-stained gel of C/M/P fractions with the molecular mass markers shown on the left.

Table 14 Amino acid sequence of TorA signal peptides used in this study

Name	Sequence
TorA	MNNNDLFQASR**RR**FLAQLGGLTVAGMLGPSLLTPRRATAAQA
TorA (R12K)	MNNNDLFQASR**KR**FLAQLGGLTVAGMLGPSLLTPRRATAAQA
TorA (R12K//R13K)	MNNNDLFQASR**KK**FLAQLGGLTVAGMLGPSLLTPRRATAAQA

The TorA-KR and TorA-KK variants were expressed as described previously and the cells were fractionated into cytoplasmic, membrane and periplasmic fractions (C/M/P). Figure 11 illustrates a very faint unprocessed precursor TorA-KK-BT6 band (21 kDa) in the cytoplasmic (C) fraction. The TorA-KR-BT6 variant as a mature sized band of BT6 (17 kDa) was observed in the periplasmic fraction (P). This data shows that the Tat system can recognize and export a fully folded *de novo* designed BT6 maquette but rejects an incompletely folded variant. Thus, the Tat system has the ability to recognize and export a *de novo* designed BT6 maquette protein and the Tat proofreading mechanism can discriminate between a folded and misfolded state. Furthermore, Tat dependent translocation by the tat system was completely abolished when BT6 variants contain an N-terminal TorA-KK signal peptide. All the data together indicates that BT6 can acquire the heme *b* cofactors *in vivo* to complete folding and that this folding event and the presence of a twin arginine TorA signal peptide can be detected by the Tat system.

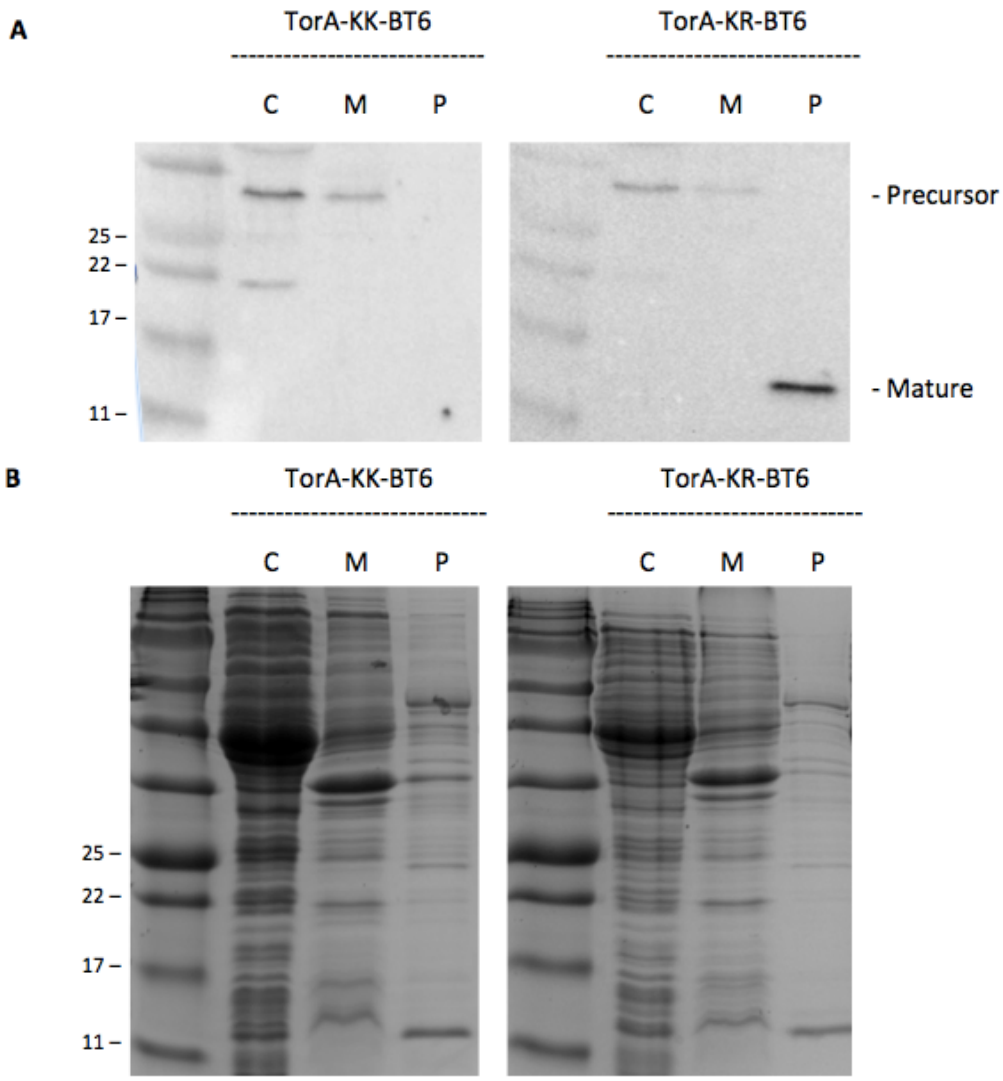


Figure 11. The export of TorA-KK-BT6 and TorA-KR-BT6 determined by immunoblotting. The left side displays the expression of TorA-KK-BT6 whereas the right side demonstrates the expression of TorA-KR-BT6. The cells were fractioned to cytoplasmic (C), membrane (M) and periplasmic (P) fractions. **(A)** The fractions were immunoblotted using a C-terminal 6xHis antibody. The molecular mass markers (kDa) and the expected size of precursor and mature bands are indicated alongside of the blot. A faint precursor-sized band (21 kDa) can be seen in the cytoplasmic fraction, which represents TorA-KK-BT6 maquette precursor protein. A 17 kDa-sized band, which represents mature TorA-KR-BT6 can be observed in the periplasmic fraction (P). **(B)** Coomassie-stained gel of C/M/P fractions with the molecular mass markers shown on the left.

Export efficiency decreased with increasing conformational flexibility

In order to test whether the Tat proofreading system can sense differences in conformational flexibility, a partially folded BT6 heme binding variant was designed (BT6M1). For BT6M1 (H53A) one heme was substituted to alanine in order perturb the ligation of one heme *b* cofactor (figure 9B). Expressions assays with the BT6 and BT6M1 substrates were performed to test whether the Tat proofreading mechanism has the ability to sense differences in substrate conformation. Similar as described previously, BT6 and BT6M1 were constructed with a TorA signal peptide and they were expressed from the same plasmid and cell line. The cells were induced with 0.5 mM IPTG and 3 hours post induction the cells were harvested and fractionated into cytoplasmic, membrane and periplasmic fractions (C/M/P).

The presence of the substrates in the cell fractions was determined by immunoblotting (figure 12). As demonstrated before, the di-heme binding BT6 was efficiently exported to the periplasm as a 17 kDa mature sized band was observed in the periplasmic fraction (P). For the one heme binding partially folded BT6M1 variant a faint 17 kDa mature sized band was observed in periplasmic fraction (P). Compared to BT6, considerably less BT6M1 protein was translocated to the periplasm by the Tat system indicating weaker export efficiency. Because the export efficiency decreased with increasing conformational flexibility we suggest that the Tat system is not only able to recognise whether BT6 is folded or largely unfolded but also has the ability to discriminate between partial folding states of BT6 substrates.

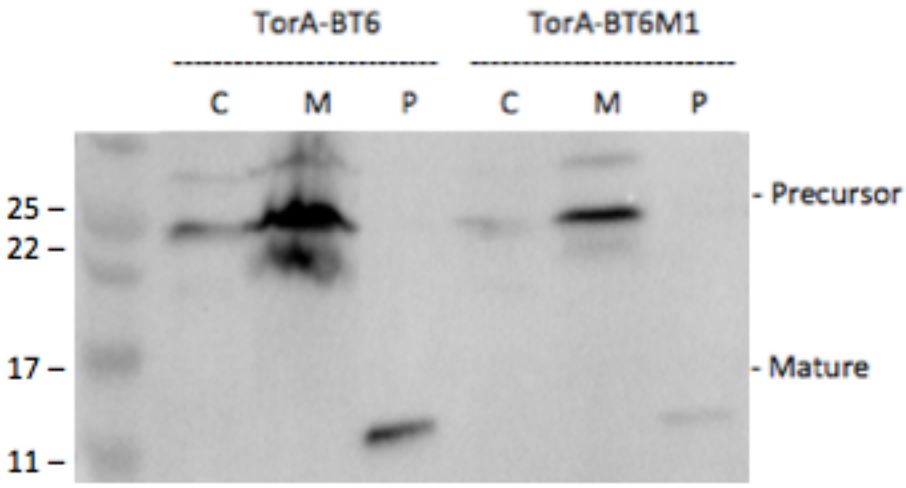


Figure 12 Export of BT6 heme binding variants by the Tat system determined by immunoblotting

The left side displays the expression of the di-heme binding BT6 whereas the right side demonstrates the expression of the one heme binding BT6M1 (H53A). The cells were fractioned to cytoplasmic (C), membrane (M) and periplasmic (P) fractions and the fractions were immunoblotted using a C-terminal 6xHis antibody. The molecular mass markers (kDa) and the expected size of precursor and mature bands are indicated alongside of the blot. A mature-sized band (17 kDa) can be seen in the cytoplasmic fraction, which represents BT6 maquette precursor protein. A faint 17 kDa-sized band, which represents mature BT6M1, can be observed in the periplasmic fraction (P). Precursor bands (21 kDa) representing both BT6 and BT6M1 are present in the cytoplasmic and membrane fractions (Data provided by Alex Jones, University of Kent).

The heme content has an effect on the conformational flexibility of BT6 maquette

Following the export assays using three BT6 maquette variants, accommodating 2 (BT6), 1 (BT6M1) or 0 (BT6M0) heme *b* cofactors, the conformational flexibility of the substrates was assessed. The BT6 variants were expressed in *E. coli* BL21 (DE3) cells and the proteins were purified from the cytoplasmic fraction by immobilized metal affinity chromatography (IMAC). Conformational changes in tertiary structure upon heme binding to the maquette variants were determined by using one-dimensional proton (^{1}H) NMR spectroscopy (work carried out by George Sutherland at University of Sheffield).

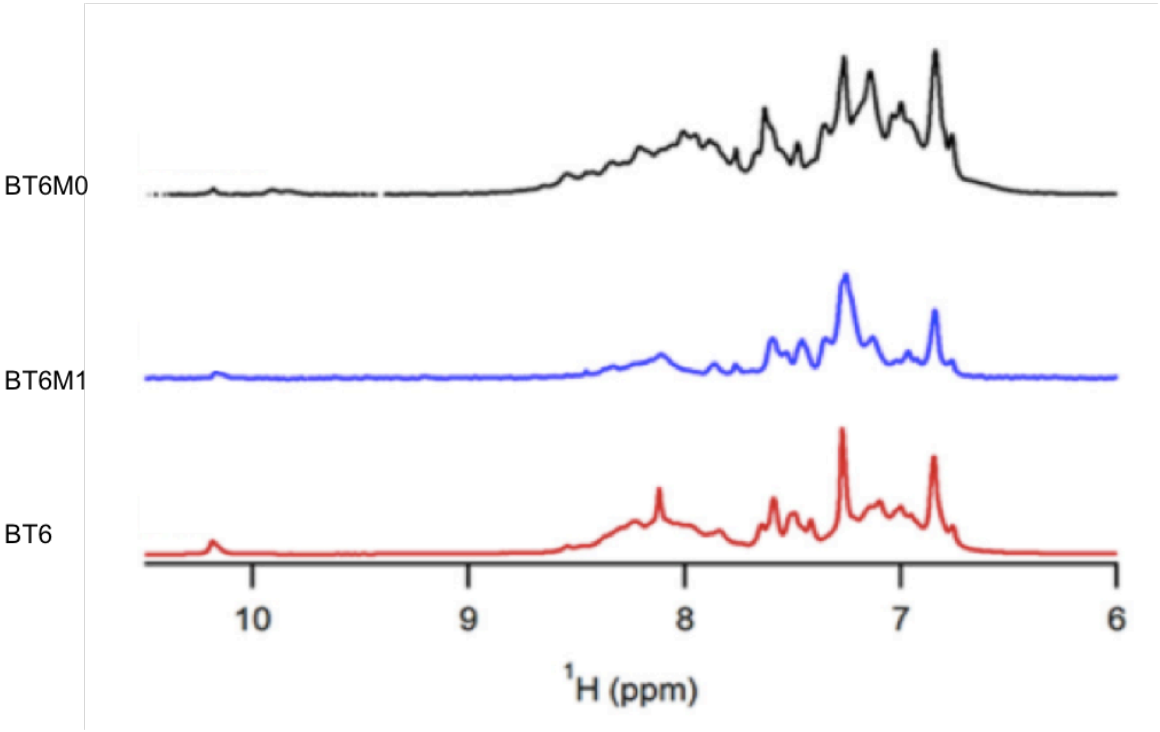


Figure 13 Proton NMR resonances of apo-BT6, apo-BT6M1 and apo-BT6M0

The figure shows a blow-up of the amide proton region between 9.5 and 12.5ppm for each of the 1D spectra. The NMR resonances of apo-BT6M0 (black), apo-BT6M1 (blue) and apo-BT6 (red) are displayed. The spectra of all maquette scaffolds indicate limited tertiary folding because the dispersion of resonances is limited (work carried out by George Sutherland at University of Sheffield).

Figure 13 displays proton NMR resonances of apo-BT6 (red), apo-BT6M1 (blue) and apo-BT6M0 (black). In the absence of heme, the three ^{1}H NMR spectra are distinctive for a protein with limited tertiary structure. Broadening of the peaks is indicative of tertiary folding.

Figure 14 shows resonances of BT6, BT6M1 and BT6M0 with (dashed line) and without (solid line) bound heme. The addition of heme *b* to BT6 resulted in a stable tertiary structure due to heme *b* cofactor binding to the BT6 scaffold as significantly broader peaks are observed (Figure 14A). Furthermore, heme binding to BT6 results in a larger dispersion in the amide proton region and at around 10ppm more peaks can be observed. For BT6M1, following the addition of heme, the same resonance dispersion was observed albeit to a reduced level (figure 14B). In contrast to BT6 and BT6M1, the NMR resonance for BT6M0 showed no significant changes in amide proton dispersion (figure 14C). In addition, Sutherland et al., demonstrated an increase in the methyl proton region for BT6 and BT6M1. In addition to the NMR experiments, Sutherland *et al* (2018) tested whether the number of bound hemes has an effect on the thermostability of the proteins using temperature-dependent CD spectroscopy. This data indicated an increase stability of the four-helix bundle structure of BT6 maquette when the number of bis-histidine ligated hemes increased.

Together, the ^{1}H NMR spectroscopy and CD spectroscopy data confirm the ligation of heme into the BT6 and BT6M1 variants and suggest an increase in protein folding upon ligation. Thus, the ligation of heme results in changes in the extent of substrate conformation of the maquette variants. The ^{1}H NMR spectra suggest a folded structure for the di-heme binding BT6, an intermediate structure for one heme binding BT6M1, and a limited tertiary structure for the no heme binding BT6M0.

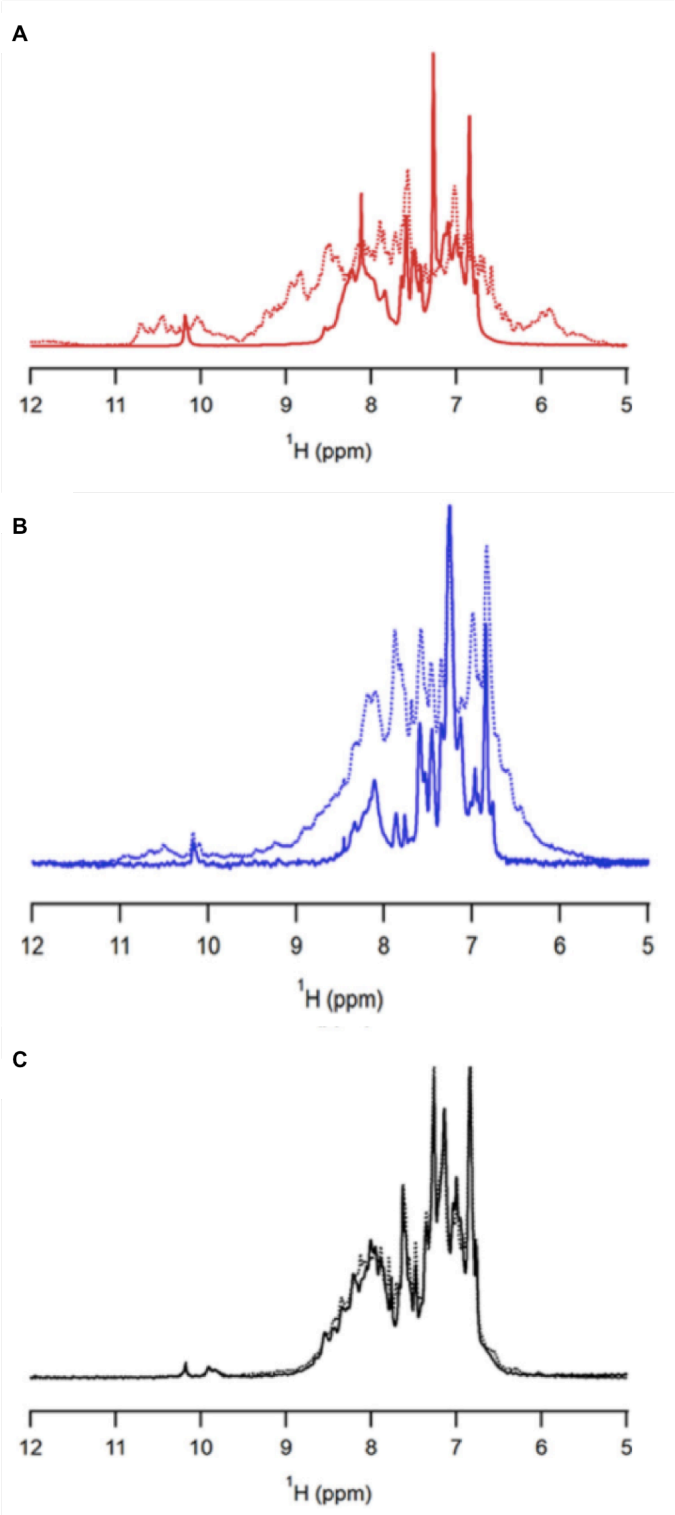


Figure 14 NMR resonances in the absence and presence of heme *b* cofactor. Comparisons of the NMR spectra for BT6 **(A)**, BT6M1 **(B)**, BT6M0 **(C)** with (dashed line) and without (solid line) are shown. In the presence of heme, changes in the amide proton resonance dispersion are demonstrated for BT6 and BT6M1. For BT6M0 no changes were observed. The changes indicate a heme binding event in correlation with a change in protein conformation (work carried out by George Sutherland at University of Sheffield).

Tat-dependent localization in the periplasm depends on heme ligation into BT6 maquette

Previously we demonstrated that heme ligation into BT6 was necessary for tertiary structure formation resulting in translocation to the periplasm by the Tat system. The number of bound hemes to the BT6 maquette variants was verified by absorption spectroscopy. The BT6 variants had a C-terminal 6xHIS tag and were expressed on a larger scale in *E. coli* W3110 Tat Express cells (Browning *et al.*, 2017). Following expression, the cells were fractionated and BT6 and BT6M1 proteins were purified from the periplasmic fraction by immobilized metal affinity chromatography (IMAC). BT6M0 was purified from the cytoplasmic fraction by IMAC. The purified maquette protein eluates were concentrated after which extensive colour differences were observed (figure 15). The fully folded (2 heme binder) purified BT6 sample has an intense red color, the partially folded (1 heme binder) BT6 sample has an orange color and the unfolded (no heme binder) is colorless. The intense red color is indicative for bound heme, as this color is caused by porphyrin complexes in heme.

Figure 15 shows the absorption spectra normalized to absorbance at 280 nm. The spectra demonstrate that both BT6 (black) and BT6M1 (blue) bind heme. For these BT6 variant a peak around 412 nm was observed which is indicative for bound heme as heme absorbs light at 412 nm. No peak at 412 nm was observed for BT6M0 (red), suggesting that no heme was bound to BT6M0. Sutherland et al., tested the absorption spectra of an *in vitro* heme reconstituted BT6 maquette to confirm that the two heme binding sites of BT6 were fully saturated. No difference was demonstrated between the spectra of the *in vivo* exported heme-loaded BT6 maquette and the *in vitro* reconstituted BT6 maquette. This data confirms that the two heme binding sites of the by Tat-transported BT6 purified from the periplasmic fraction are fully saturated.

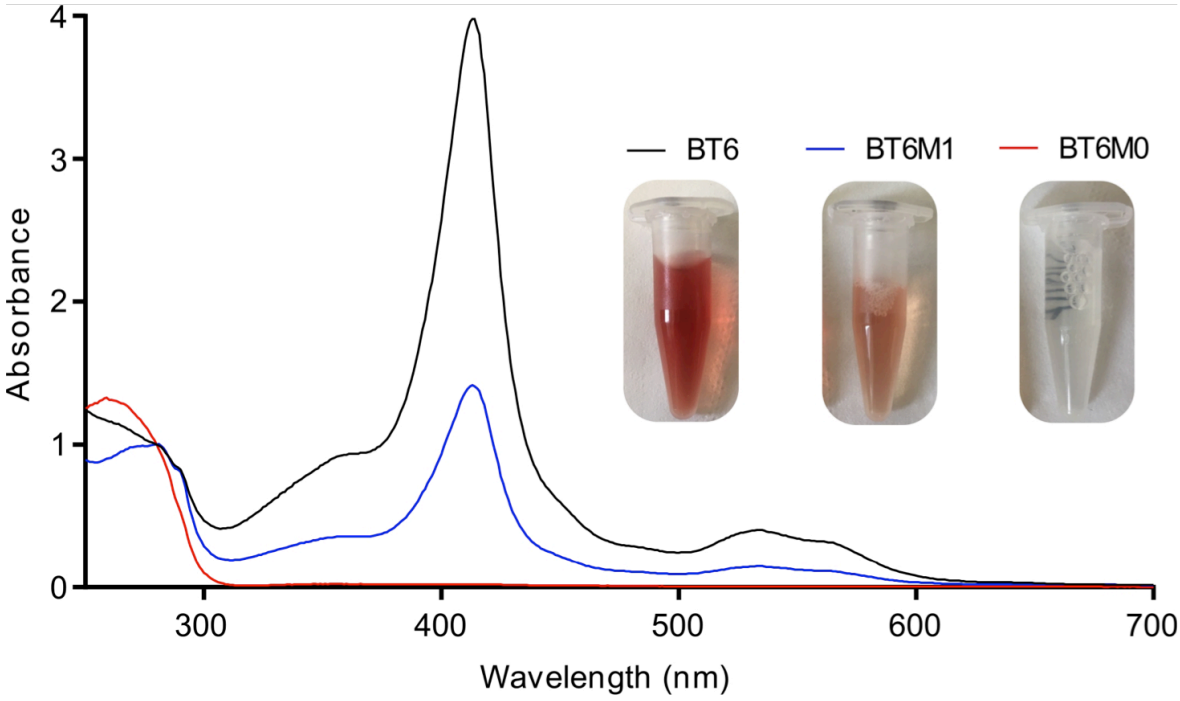


Figure 15 Spectroscopic analysis of purified maquette proteins

UV-visible absorption spectra of BT6 (black), BT6M1 (blue) and BT6M0 (red) normalized at 280 nm. A peak at 412 nm indicates bound heme. Compared the one heme binding BT6M1, the measured absorbance at 412 nm for the di-heme binding BT6 variant was more abundant. In addition, the figure shows the colour difference of the BT6 protein eluates. The purified di-heme binding BT6 sample has an intense red colour, whereas the one heme binding BT6M1 variant has a less intense red colour. The non-heme binding purified BT6M0 elution sample has no visible colour.

Although mature BT6 maquette proteins exhibit a molecular weight of ~17 kDa, they have the tendency to run slightly differently on a SDS-PAGE gel. Immediately after purification, we analysed the BT6 maquette protein after purification from the periplasm, using mass spectrometry to define the protein more precisely. The mass spectrometry data shows a mass that corresponds to the predicted molecular weight of BT6 (16,981.29 Da) (figure 16).

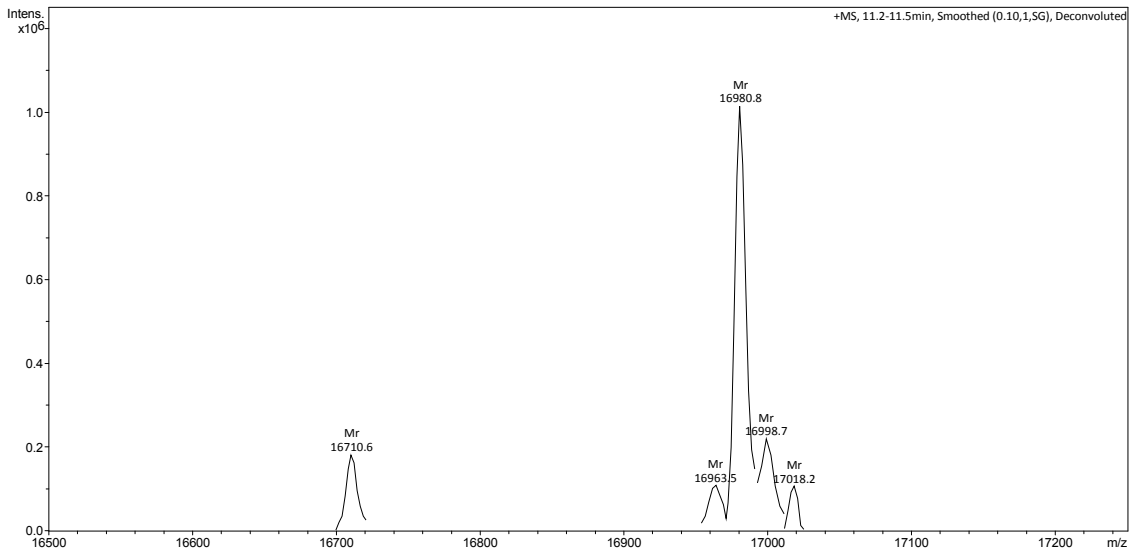


Figure 16 Mass spec data verifying purified di-heme binding BT6

The spectra of the electrospray mass spectrometry analysis of di-heme binding BT6 purified from the periplasmic fraction. The 16980.8 Da peak corresponds to the molecular mass of BT6.

3.2 Discussion

Protein trafficking is essential as more than one third of the proteome of Gram-negative bacteria are secreted across the inner and outer membrane (Economou, 2010a). The twin arginine translocase (Tat) is one of the major pathways to transport proteins from the cytoplasm to the periplasm. Many substrates that are targeted by the Tat system fold around a co-factor in the cytoplasm which bypasses an extra mechanism to export the separate cofactor to the periplasm (Berks, 1996). Thus, the Tat system only transports folded proteins across membranes. Therefore, Tat has the ability to discriminate between folded and unfolded proteins because it consists of a proofreading mechanism. It is not well understood how the Tat proofreading mechanism is able to determine that a protein is folded and export competent.

In this study we investigated whether the Tat proofreading mechanism is capable of discriminating between different conformational states of a *de novo* designed substrate, denoted BT6 maquette. BT6 maquette is an artificial four α-helix water-soluble protein encompassing a hydrophobic interior that accommodates two heme *b* cofactors. A TorA signal peptide was fused to the N-terminus for periplasmic localization by the Tat system and for immunodetection a 6xHis tag was fused to the C-terminus. To perturb the folding state of BT6 maquette the number of heme cofactors was reduced. Therefore, two BT6 folding variants, BT6M1 and BT6M0, were generated by point mutations affecting the heme binding sites to counteract the ligation of 1 or two heme cofactors. UV-visible absorption spectra confirmed the reduced numbers of heme in BT6M1 and BT6M0 (figure 15).

Exported assays revealed that the Tat system has the ability to recognize and export an artificial cofactor binding substrate (figure 10). Moreover, there is a correlation between the conformational flexibility and the export efficiency of BT6, BT6M1 and BT6M0. When targeted for export by the Tat system, the export efficiency decreased with increasing conformational flexibility (figure 12). The Tat proofreading mechanism rejected the unfolded BT6M0 (figure 10). Compared to the fully folded di-heme binding BT6, the partially folded one heme binding BT6M1 substrate was exported in a less efficient manner (figure 11).

^{1}H NMR spectroscopy analysis showed that the heme content has an effect on the conformational flexibility of BT6 maquette proteins (figure 14). This data showed a correlation between the number of bound heme cofactors and the folding state of the BT6 proteins. In addition, Sutherland et al., tested whether the number of bound hemes has an effect on the thermostability of the proteins using temperature-dependent CD spectroscopy. This data demonstrated that the four-helix bundle structure was significantly stabilized by the bis-histidine ligation of hemes within the BT6 maquette cavity. Together, the ^{1}H NMR spectroscopy and CD spectroscopy data show consequent effects of heme content on substrate conformation.

This data shows that the Tat system is able to recognize the artificial designed BT6 maquette proteins. Moreover, it does not only recognise the substrate but the proofreading mechanism can discriminate between folded states. The intermediate folded BT6M1 was translocated differently compared to a fully folded BT6 and unfolded BT6M0. This suggests that the Tat proofreading pathway involves interactions between Tat components and the substrate at the membrane surface to sense the structural flexibility and overall

export competence of a substrate. This cascade of interactions can sense flexible conformations, resulting in the rejection of an unfolded substrate.

This study shows that the *de novo*-designed BT6 substrate, which the native *E. coli* Tat system will never encounter, is an ideal candidate to study the Tat systems proofreading mechanism. Due to this ability, the Tat system has a natural tendency to produce high quality, active proteins. Improving our understanding of this remarkable proofreading mechanism is relevant for biotechnological exploitation. Expanding this proof-of-concept study will contribute to the current knowledge about the proofreading mechanism. By using the same principle we can probe the tolerance of the Tat systems proofreading mechanism.

Chapter 4:

The twin-arginine proofreading ability of *Escherichia coli* can sense localized unfolded regions of a man-made maquette protein variant.

4.1 Introduction

In *E. coli,* the Tat system consists of a proofreading and quality control mechanism, which only allows export of folded proteins. Due to this unique capacity it is likely that the Tat system has a natural tendency to produce high quality, active proteins. Various studies have demonstrated the Tat system's ability to reject unfolded proteins (Robinson *et al.*, 2011, Maurer *et al.*, 2009). Other studies have shown that scFv and hGH substrates lacking disulphides were translocated by Tat which indicates that the Tat system tolerates minor structural changes (Alanen *et al.*, 2015). This was also observed in chapter 3, where it was demonstrated that a partially folded one heme binding BT6 variant was translocated by Tat albeit less efficiently (Sutherland *et al.*, 2018). Furthermore, studies seem to suggest a hydrophobicity sensor employed by the Tat apparatus rather than assessment of the folded state of substrates (Richter *et al.*, 2007). Strikingly, other studies demonstrated that the Tat system tolerates significant changes in substrate surface hydrophobicity (Jones *et al.*, 2016). This study suggested that the Tat proofreading mechanism senses structural flexibility of a substrate rather than the surface characteristics. Evidently, the Tat proofreading mechanism is not well understood.

The aim of the study was to increase the understanding of the Tat export system's proofreading ability, using a simple model substrate to create various mutations in order to change (i) the surface characteristics (ii) or the addition of unfolded domains. Export assays were used to see which variants are exported and which are rejected. To analyse the proofreading capacity we chose a *de novo*-designed BT6 maquette protein substrate (Farid *et al.*, 2014). The structure of the so-called maquette BT6 we used is shown in Figure 17. BT6 shares no sequence identity with natural substrates and the complexity was minimized to increase engineering freedom (Farid *et al.*, 2014). The protein consists of four α-helix

loops that form a hydrophobic cavity that accommodates two cofactors. The BT6 interior maintains *bis*-histidine ligation sites to coordinate two heme *b* co-factors.

We have previously shown that the Tat translocase is able to recognize and distinguish between folded and unfolded BT6 maquette substrates (Sutherland *et al.*, 2018). In chapter 3, we tested whether there is a difference in translocation of fully folded (BT6), partially folded (BT6M1), and unfolded BT6 (TorA-BT6M0) variants. The folding state of the BT6 maquettes was affected by reducing the number of bound hemes. The resulting hypothesis is that the Tat translocase must sense differences in structural flexibility of BT6 maquette proteins. This data suggest that the *de novo*-designed BT6 substrate is an ideal candidate to study the Tat systems proofreading mechanism. In addition, altering natural substrate residues can have unpredictable effects on protein function and conformation; whereas the effect of altering residues of a *de novo* designed substrate is more predictable.

In this study, we examined the Tat system's proofreading ability by using BT6 maquette. BT6 substrates were mutated in order to change the surface characteristics to probe the tolerance of the Tat system's proofreading mechanism. Figure 17 shows the secondary and tertiary structure of the di-heme binding BT6 maquette protein. The results suggest that the Tat proofreading system does not sense a global unfolded state of the substrate but has the ability to sense localised unfolded regions.

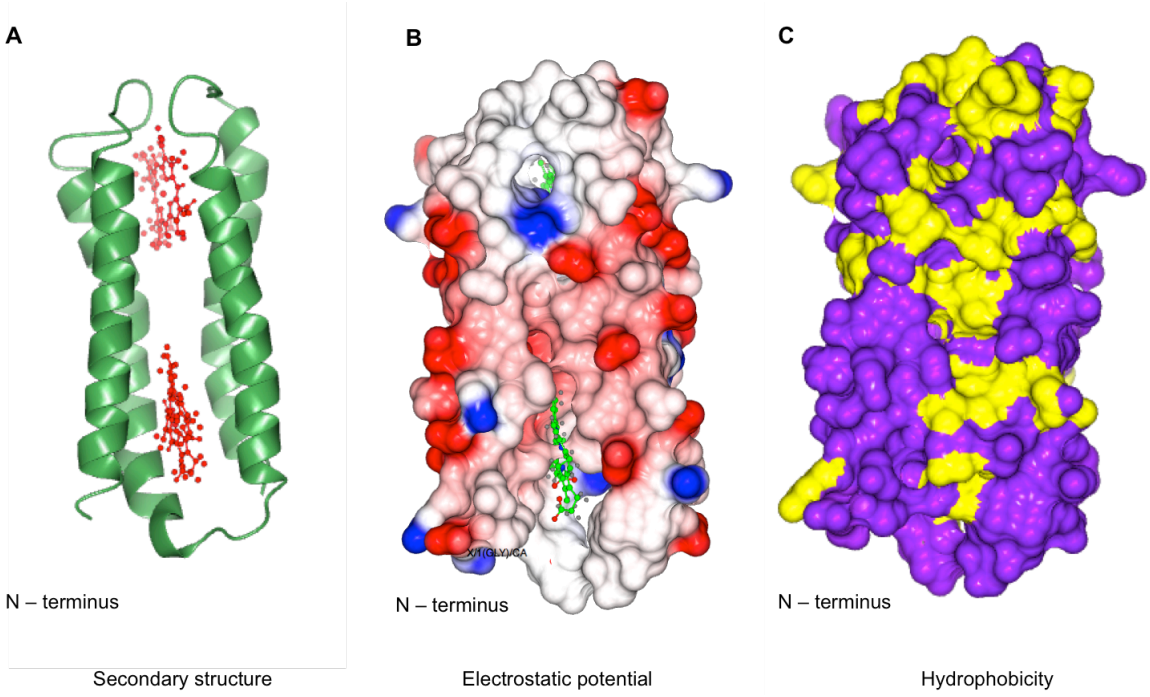


Figure 17 Secondary and tertiary structure of the di-heme binding BT6 protein.

(A) The ribbon diagram of the overall structure of the two-heme binding maquette BT6 protein shows four identical α-helix loops (green). The loops accommodate four histidine residues to ligate two heme *b* molecules (red). **(B)** The electrostatic potential map of the BT6 outer surface shows the positively charged (red), negatively charged (blue) and uncharged (grey) residues. **(C)** The hydrophilic (purple) and hydrophobic (yellow) outer surface residues are shown in the hydrophobicity map.

4.2 Results

The Tat system does not tolerate changes in substrate surface hydrophobicity of BT6 maquette

There is a possibility that the Tat translocase rejects proteins containing hydrophobic surface patches (Richter *et al.*, 2007). It has been speculated that the presence of hydrophobic surface patches prevents transport. The hydrophobic interior of a substrate could be revealed in unfolded or partially folded substrates. There is a possibility that the Tat components sense the hydrophobic interior, which is normally sequestered in a fully folded state. Richter et al. (2007) demonstrated that Tat substrates were rejected upon insertion of a short hydrophobic sequence. We therefore first tested whether introducing hydrophobic patches on the surface of a fully folded BT6 maquette (BT6-3NLeu) affected the export by the Tat translocase. An N-terminal Tat signal peptide TorA and a C-terminal 6x His-tag were fused to the maquette protein and Table 15 lists the amino acid sequence of BT6-3NLeu with the substitution shown in bold.

Figure 17C shows the predicted surface hydrophobicity map of the fully folded two-heme binding BT6 maquette (WT). The outer surface is predicted to be relatively hydrophilic with the hydrophilic surface residues displayed in purple and the hydrophobic surface residues in yellow. We substituted 3 hydrophilic charged surface residues (Asp, Lys, GLu) with leucine in the N-terminal domain (D9L, K13L, E16L). Figure 18A shows the altered hydrophilic outer surface residues (purple) on the predicted ribbon diagram.

The hydrophobicity mutant was cloned into pEXT22 and expression was induced with 0.5 mM IPTG. 3 hours post induction the cells were harvested and fractionated into cytoplasmic, membrane and periplasmic fractions (C/M/P). Figure 18B shows that the

BT6-3NLeu BT6 variant is rejected by the Tat translocase. There is no mature sized band of BT6-3NLeu (17 kDa) observed in the periplasmic fraction (P). Unprocessed precursor with a higher molecular weight (21 kDa) is observed in the cytoplasmic (C) and membrane (M) fraction. It is more likely that the overall weak signal intensity of the bands is caused by a technical problem rather than a lack of protein expression as the data also shows weak presence of the positive control 'BT6'. Thus, increasing the outer surface hydrophobicity of BT6 maquette is not tolerated by the Tat translocase. However, the outer surface alterations might perturb the folding of the substrate, which might result in the Tat proofreading mechanism to reject the substrate because it is not folded properly. Moreover, by fractionating the cells into a cytoplasmic, membrane and periplasmic fraction we cannot assess whether the protein was not translocated due to possible inclusion body formation. This can be determined by harvesting a fourth cell fraction, the insoluble fraction.

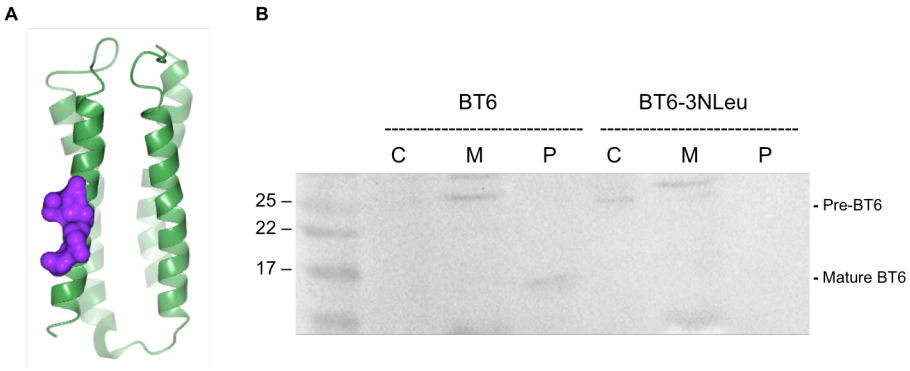


Figure 18 Changes in substrate surface hydrophobicity of BT6 is not tolerated

(A) Structure of the BT6 maquette hydrophobicity variant (BT6-3NLue). The altered hydrophilic outer surface residues (purple) are shown on the ribbon diagram. 3 hydrophilic surface residues were substituted with leucine in the N-terminal domain (D9L, K13L, E16L). **(B)** The cells were fractioned to cytoplasmic (C), membrane (M) and periplasmic (P) fractions. The fractions were immunoblotted using a C-terminal His antibody. The molecular mass markers (kDa) are shown on the left and the expected size of precursor and mature bands on the right side of the blot. A precursor-sized band (21 kDa) can be seen in the cytoplasmic and membrane fraction, which is, unprocessed BT6-3NLeu BT6 maquette.

Addition of an unfolded C-terminal domain is not tolerated by the Tat system

Previous studies suggested that the addition of an unfolded element to a folded substrate aborts export by the Tat system (Jones *et al.*, 2016). A 26-residue unfolded domain, chosen at random, was added to the C-terminus of a folded scFv protein. Tat detected the addition of the unfolded domain and the protein was rejected. Jones *et al* (2016) demonstrated that the additional domain did not affect the confirmation of the core scFvM whereas the additional domain forms an unfolded region. This was indicated by secondary structure modelling using the Phyre^2 program (Kelley *et al.,* 2015).

In this study, we tested whether the addition of different sized domains to a fully folded BT6 maquette affects export. Therefore, we designed three BT6 variants by adding 26 residues (SNVIIIIITNKDPNSSSVDLAAALE), 13 residues (SNVIIIIITNKDP) or 6 residues (SNVIII) to the C-terminus of BT6 maquette (WT). As described previously, an N-terminal TorA signal peptide and a C-terminal 6x His-tag were fused to the BT6 variants.

Figure 19 shows that there is no mature sized band of (17 kDa) observed for any of the three BT6 variants in the periplasmic fraction (P). The Tat substrates are degraded in the cytoplasm when export to the periplasm does not take place. However, the blots show a decrease in signal intensity when the length of the unfolded domains increases. This indicates that there might be a correlation between how unfolded the Tat substrate is and the efficiency of degradation. In summary, Tat can detect the presence of an additional 6-, 13- and 26-residue C-terminal unfolded domain and aborts translocation of these BT6 maquette variants.

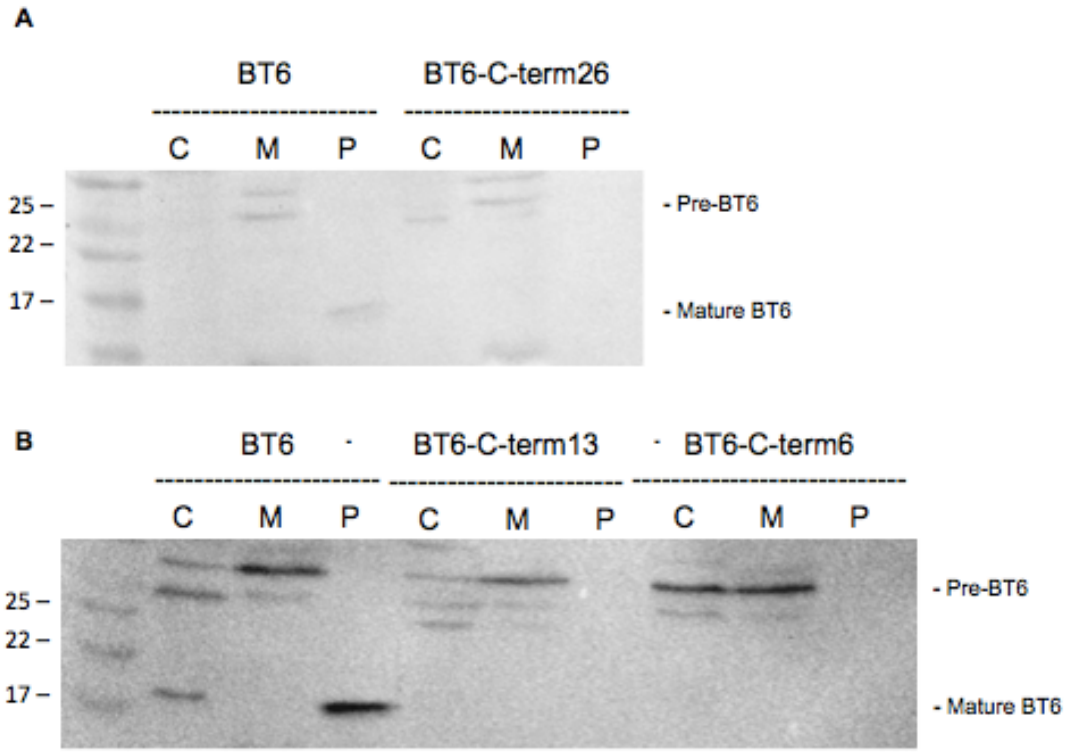


Figure 19 An addition of an unfolded domain to a folded BT6 is rejected by the Tat system. The immunoblots show BT6 maquette variants containing an additional unfolded domain. The cells were fractionated to cytoplasmic (C), membrane (M) and periplasmic (P) fractions. The fractions were immunoblotted using a C-terminal His antibody. The molecular mass markers (kDa) are shown on the left side of the blot. No mature sized (17 kDa) bands were observed in the periplasmic fraction (P). The different immunoblots show the export of the di-heme binding BT6, **(A)** BT6-C-term26, **(B)** BT6-C-term13 and BT6-C-term6.

The Tat system is highly tolerant of changes in substrate surface charge

We tested whether changes on the substrate's outer surface electrostatic potential (figure 17B) affects efficient export to the periplasm by the Tat translocase. It has been suggested that the Twin-arginine translocase is highly tolerant towards surface charge changes (Jones *et al.*, 2016). However, the Tat translocase might sense certain outer surface characteristics as unfolded.

To test whether alterations in the surface charge of maquette BT6 affects export, four BT6 mutants with altered outer surface characteristic were designed (figure 20A). Two mutants were designed with 2 or 4 positively charged Lysine residues substituted to 2 or 4 negatively charged Aspartic acid residues (denoted BT6-2Asp and BT6-4Asp). In contrast, two mutants were designed with 2 or 4 negatively charged Aspartic acid residues substituted to 2 or 4 positively charged Lysine residues (denoted BT6-2Lys and BT6-4Lys). An N-terminal Tat signal peptide TorA and a C-terminal 6x His-tag were fused to the BT6 electrostatic potential mutants.

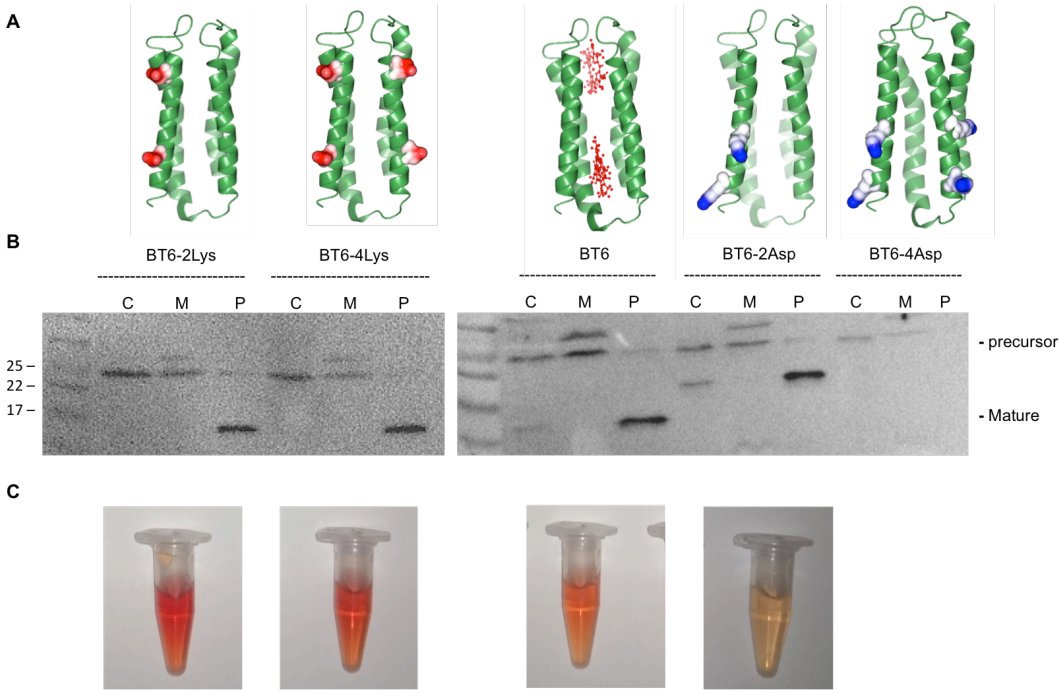


Figure 20 Changes in substrate surface electrostatic potential of BT6 is highly tolerated **(A)** Representation of the altered outer surface residues of the BT6 maquette variants: BT6-2Lys, BT6-4Lys, BT6-2Asp and BT6-4Asp. The substituted residues (negative = red and blue = positive) are shown on the simulated ribbon diagram. The two left mutants had 2 and 4 negatively charged residues substituted to 2 and 4 positively charged residues (denoted BT6-2Lys and BT6-4Lys). In contrast, the 2 and 4 positively charged residues of the two right mutants were substituted to 2 or 4 negatively charged residues (denoted BT6-2Asp and BT6-4Asp). **(B)** The immunoblots of the BT6 variants shows mature sized bands of (17 kDa) for BT6-2Lys, BT6-4Lys and BT6-2Asp in the periplasmic fraction (P). There is no mature sized band in the periplasmic fraction observed for BT6-4Asp. **(C)** The color of the purified BT6 variant (BT6-2Lys, BT6-4Lys and BT6-2Asp) from the periplasmic fraction by immobilized metal affinity chromatography (IMAC). The intense red color indicates that BT6-2Lys and BT6-4Lys accommodate two hemes and therefore are fully folded. BT6-2Asp had an orange color, which suggest this BT6 variant only accommodates one heme.

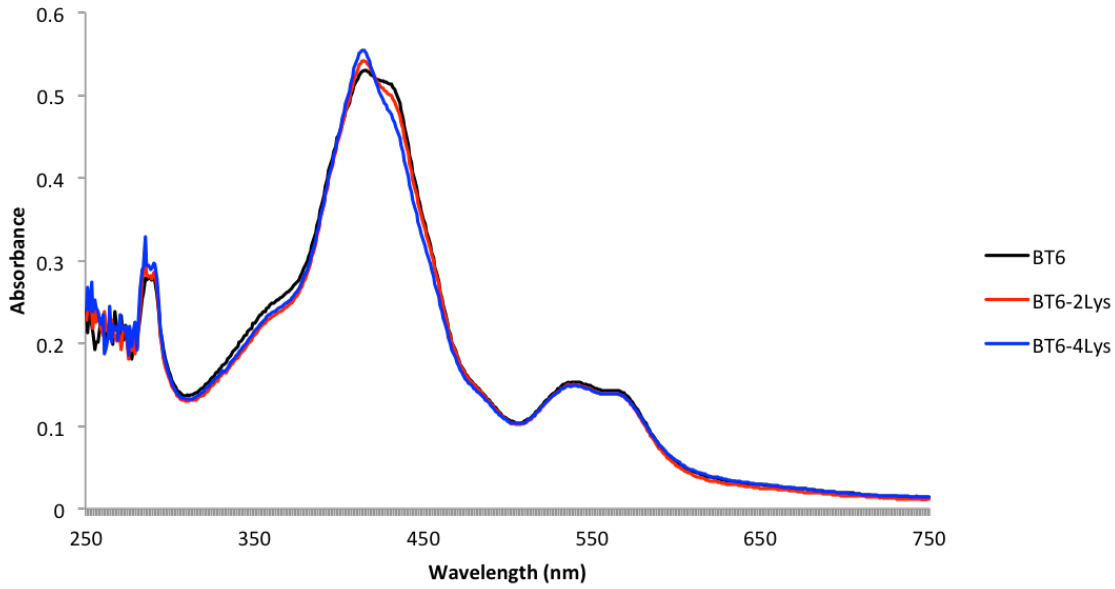


Figure 21 UV-visible absorbance spectra of BT6-2Lys and BT6-4Lys

The absorbance spectra of BT6-2Lys (red) and BT6-4Lys (blue) compared to the previously published absorbance spectra of BT6 (black). BT6-2Lys and BT6-4Lys show high similarity with the spectrum of the di-heme binding BT6, suggesting that the two BT6 mutants are accommodating two heme cofactors.

Figure 20B shows that there is a mature sized band of (17 kDa) observed for BT6-2Lys, BT6-4Lys and BT6-2Asp in the periplasmic fraction (P). The 3 mutants are all exported efficiently. There is no mature sized band in the periplasmic fraction observed for BT6-4Asp. It is unknown whether the export of BT6-4Asp by Tat was blocked due to the altered outer surface. The outer surface changes might have perturbed folding of the BT6 variant.

In order to test whether the surface changes of the exported BT6 variants introduced structural changes we purified BT6-2Lys, BT6-4Lys and BT6-2Asp from the periplasmic fraction by immobilized metal affinity chromatography (IMAC). Figure 20C shows the color of the purified BT6 variants. Previously, it has been demonstrated that a fully folded (2 heme binder) purified BT6 sample has an intense red color, a partially folded (1 heme

binder) BT6 sample has an orange color and an unfolded (no heme binder) is colorless (Sutherland *et al.*, 2018). After purification the intense red color indicated that BT6-2Lys and BT6-4Lys accommodate two hemes and therefore are fully folded (figure 21). BT6-2Asp had an orange color, which suggests this BT6 variant only accommodates one heme.

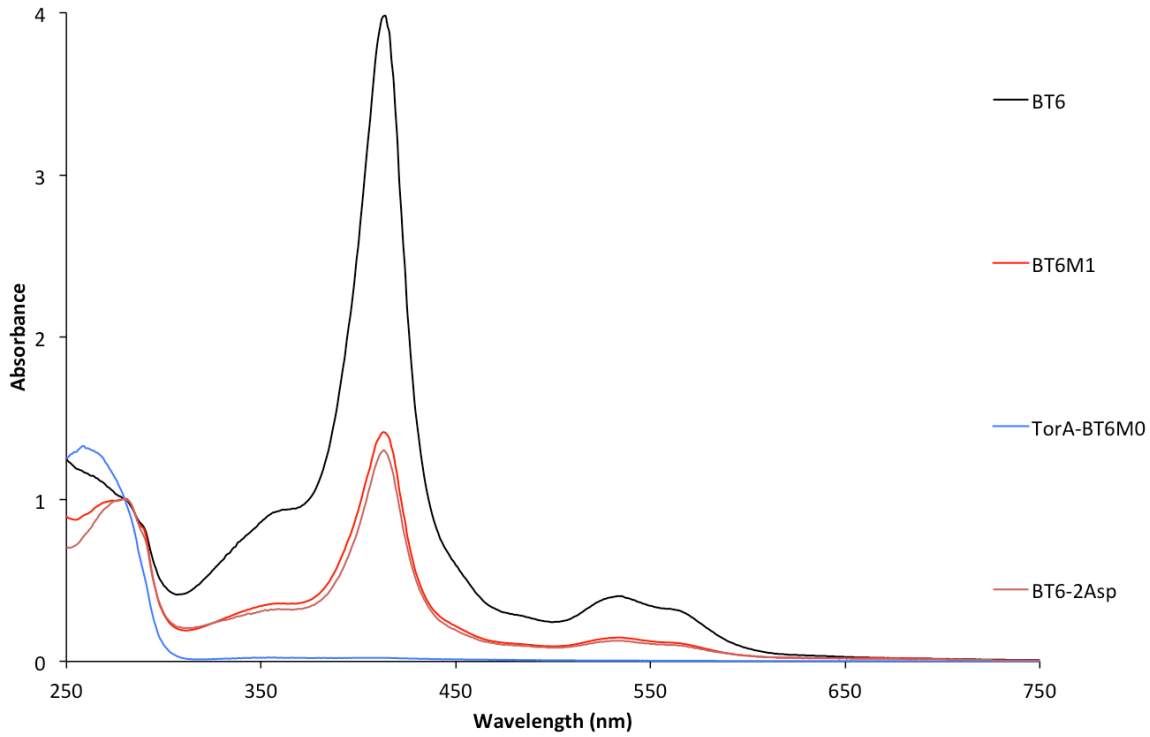


Figure 22 UV-visible absorbance spectra of purified BT6-2Asp

The absorbance spectra of BT6-2Asp (orange) compared to the previously published absorbance spectra of BT6 (black), BT6M1 (red), TorA-BT6M0 (blue). There is high similarity between the spectrum of BT6M1 (one heme binder) and BT6-2Asp, suggesting that Bt6-2Asp only accommodates one heme.

Figure 22 shows the absorbance spectra of BT6, BT6M1, TorA-BT6M0 and BT6-2Asp normalized at 280nm to demonstrate that BT6-2Asp only accommodates one heme. There is high similarity between the spectrum of BT6M1 (one heme binder) and BT6-2Asp. Efficient export of fully folded BT6 was demonstrated, whereas BT6M1 is exported with moderate efficiency.

Figure 23 shows a time course expression assay of BT6, BT6-2Asp and BTM1. The cells where induced with IPTG and the periplasmic fraction was harvested after different time points (60, 120, 300 min). The figure shows a mature 17 kDa BT6 and a faint mature BT6M1 band in the periplasm. However, the mature one heme binding BT6-2Asp bands are more intense which indicates that it is exported even more efficiently than the non-mutated fully folded BT6. Note that the BT6-2Asp protein runs higher on SDS gels, presumably due to the presence of additional negative charges.

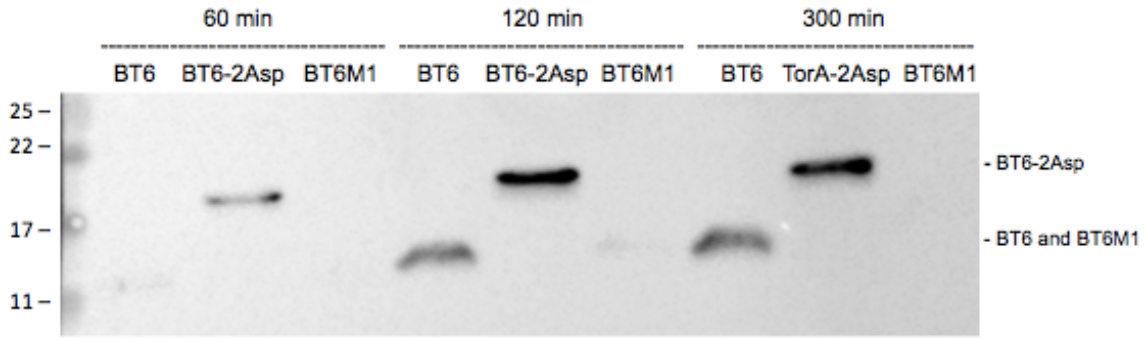


Figure 23 The alteration of two positively charged residues into two negatively charged residues results in an increase of export. An Immunoblot of BT6, BT6-2Asp and BT6M1 is shown. W3110 'TatExpress' cells where induced with (o.5mM) IPTG and the periplasmic fraction was harvested after different time points (60, 120, 300 min). Mature sized 17kDa bands of BT6, BT6-Asp and a faint mature BT6M1 band were observed in the periplasm. However, the mature one heme binding BT6-2Asp bands are more intense which indicates that it is exported more efficiently than the fully folded BT6.

1D NMR indicates that BT6-2Asp is generally less tightly folded than BT6, but has a specific region that is more tightly folded

We purified the maquette variants and subjected them to NMR analysis because the nature of the technique is very sensitive to the structure and dynamics of molecules at the atomic scale on timescales of ms-ns. However, the spectra of the maquettes are complicated

because of the presence of heme groups with a high spin Fe^{III} present (low spin heme groups only occur when O_2 is bound to the heme's Fe^{III} ion). This leads to two processes occurring as hemes are bound to the maquettes. The first is that the presence of heme group in the maquette may increase the 'foldedness' or order of the protein leading to an increase in chemical shift dispersion and a broadening of the proteins peaks to the line width expected for a protein of this size. In the case of the maquettes (mass ~ 21 kDa) we may predict a linewidth of somewhere around 18Hz for an amide or 0.03 ppm at 600 MHz (Cavenagh *et al,* 2006) assuming only dipole-dipole relaxation is present and making the approximation that the protein is approximately spherical.

The second effect is that the lone pair from the paramagnetic centre in the heme can lead to extreme shifts for nuclei both within the porphyrin ring that encloses it and also in nearby residues in the protein. On top of this the paramagnetic centre present in the heme can also cause broadening due to increased T2 relaxation. In interpretation of the results it is assumed based on the observed data and previous analysis that TorA-BT6M0 contains no heme and so is not broadened, but that BT6, BT6M1 and BT6-2Asp are all bound to heme to some extent and so show broadening and extreme shifts. Examples of these extreme shifts can be seen in figure 24 for the low field region of their spectra, with broad dispersed peaks being observed between 9.5 and 10.5 ppm.

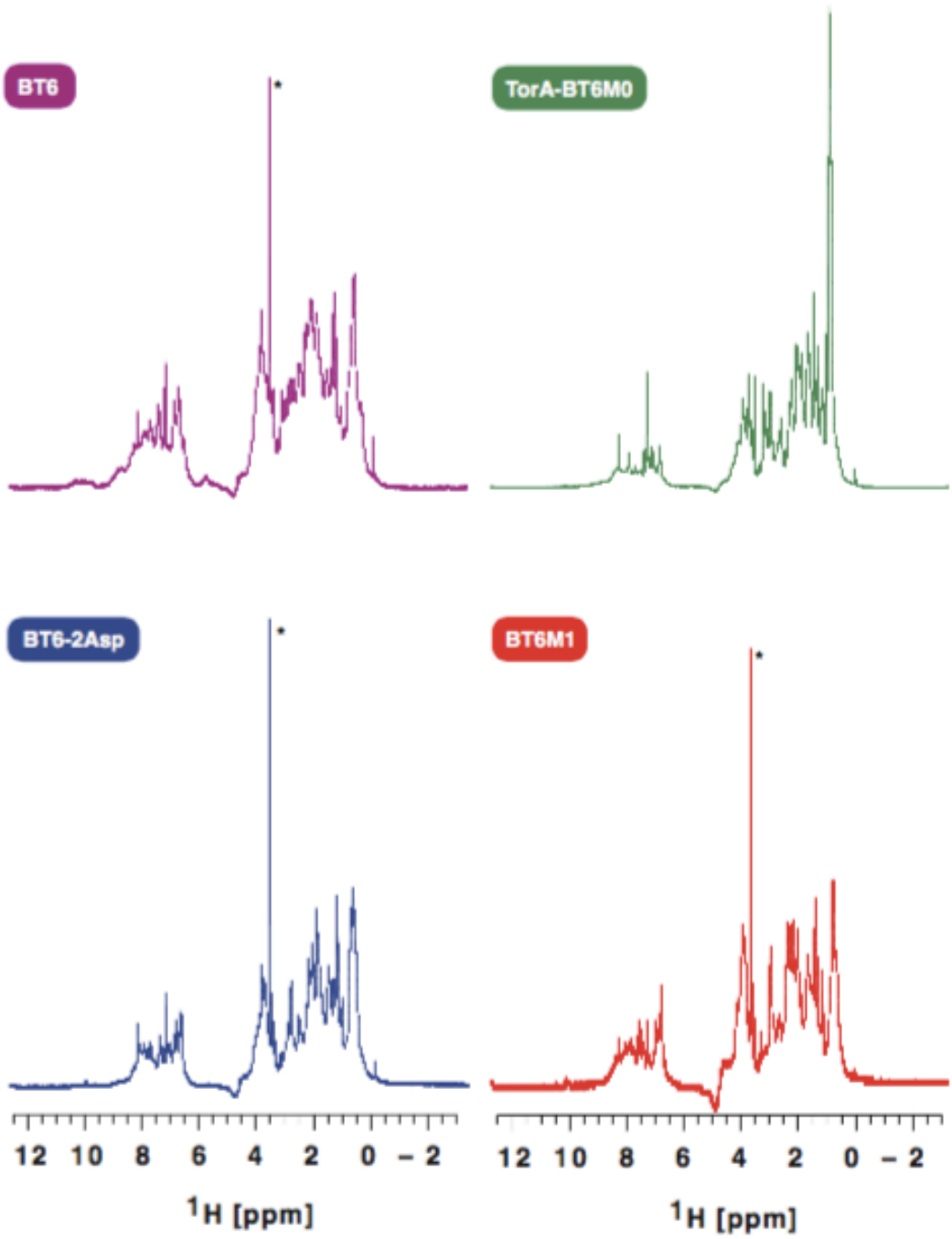


Figure 24 1D-1H-NMR spectra suggest that BT6-2Asp is more folded than BT6M1 and BT6 1D 1H NMR spectra of BT6, BT6M1, TorA-BT6M0, and BT6-2Asp measured at 298K and 600MHz in 90%H_2O/10%D_2O using a DPFGSE water supressed pulse sequence. The peak marked with a * is presumed to be from a small molecule contaminant.

The spectrum of TorA-BT6M0 is relatively sharp and has a narrower overall envelope for the amide and aliphatic sidechain protons indicating that it is at least partially folded (it is still too broad and dispersed to be a fully unfolded spectrum). On the other hand BT6 and BT6M1 both show broadening of the envelope for the amide and aliphatic sidechain protons compared with TorA-BT6M0 showing that Fe^{III} heme groups are bound and also possibly suggesting an increase in the structuring of the protein (BT6M1 is somewhat sharper but still well dispersed, most probably because there is less relaxation from the single bound heme-Fe^{III} group). The mutant protein BT6-2Asp also shows similar broadening of the aliphatic and amide envelopes similar to BT6M1 suggesting that the structure present is similar to that in BT6M1.

However, analysis of the extremely shifted peaks in the proteins above 9.5 ppm (Figure 25) may suggest that a more subtle and interesting analysis can be made. For both BT6M1 and especially BT6 a number of broad peaks are observed in this region which could be attributed to highly shifted peaks from the heme and the protein due to aromatic ring currents and hyperfine shifts from the heme-Fe^{III} moiety in a similar fashion to those observed in previous studies (Du *et al.*, 2003). However, the observed peaks are considerably broader than those observed in Du et al leading to the conclusion that there is some movement of the heme within the structure of the maquette leading to the heme taking up multiple orientations which leads to further broadening due to exchange effects. It can also be seen that these peaks cannot be attributed to just the tryptophan side chains which are also present in this region as area of the observed peaks appears to be too large to be consistent with the 4 tryptophan peaks present (the area of the region 9.5-10.5 ppm is $1/32^{nd}$ of the amide and aromatic ring region; 6.1-9.5 ppm whereas it should be ~1/70th). However, analysis of the same region in the protein BT6-2Asp shows two additional peaks at 10.05 ppm, which are entirely consistent with two of the tryptophan residues present in

the protein. As these peaks are not observed in the spectra of any other proteins studied this suggests that there is an ordering around two of the tryptophans (if it was an ordering of the heme group it would be expected that many more signals would be observed). Furthermore, the linewidths observed (10 Hz) are not inconsistent with the peaks expected from the tryptophans of a protein of mass. It could be suggested, as an alternative that the peaks could be caused by a decrease of order as this can also sharpen peaks, however, as there is no other sign of sharpening in the overall envelope of the protein this seems unlikely.

Thus, the BT6-2Asp variant yielded unexpected information. With BT6-2Asp and BT6M1, the NMR data raise the possibility that a localised domain differs in terms of conformational flexibility, and that this domain may include unspecified Trp residues in the BT6 protein. The inference is that this domain is particularly rigid in the BT6-2Asp variant, and that the Tat system accepts this variant as folded because the domain is one of the primary interaction sites during the binding/ proofreading process. Therefore, overall it can be suggested that BT6-2Asp is more ordered and structured than BT6M1 and possibly even BT6. In order to confirm this suggested hypothesis additional data is necessary. For example, the acquired NMR data can by validated by measuring the difference in structural stability of BT6-2Asp and BT6M1 by preforming circular dichroism (CD) spectroscopy assays. Exposing the BT6 substrates to a temperature gradient can assess the thermostability. BT6 substrates with a more rigid structural conformation will be more tolerant to higher temperatures compared to the less tightly folded substrates. Herewith the thermal stability of the two BT6 scaffolds can be compared. Moreover, ^{15}N NMR or coupled ^{1}H -^{15}N NMR spectroscopy analysis could be performed in order to gain more sensitive structural data on the structural confirmation and the location of the *specific region that is possibly more tightly folded*

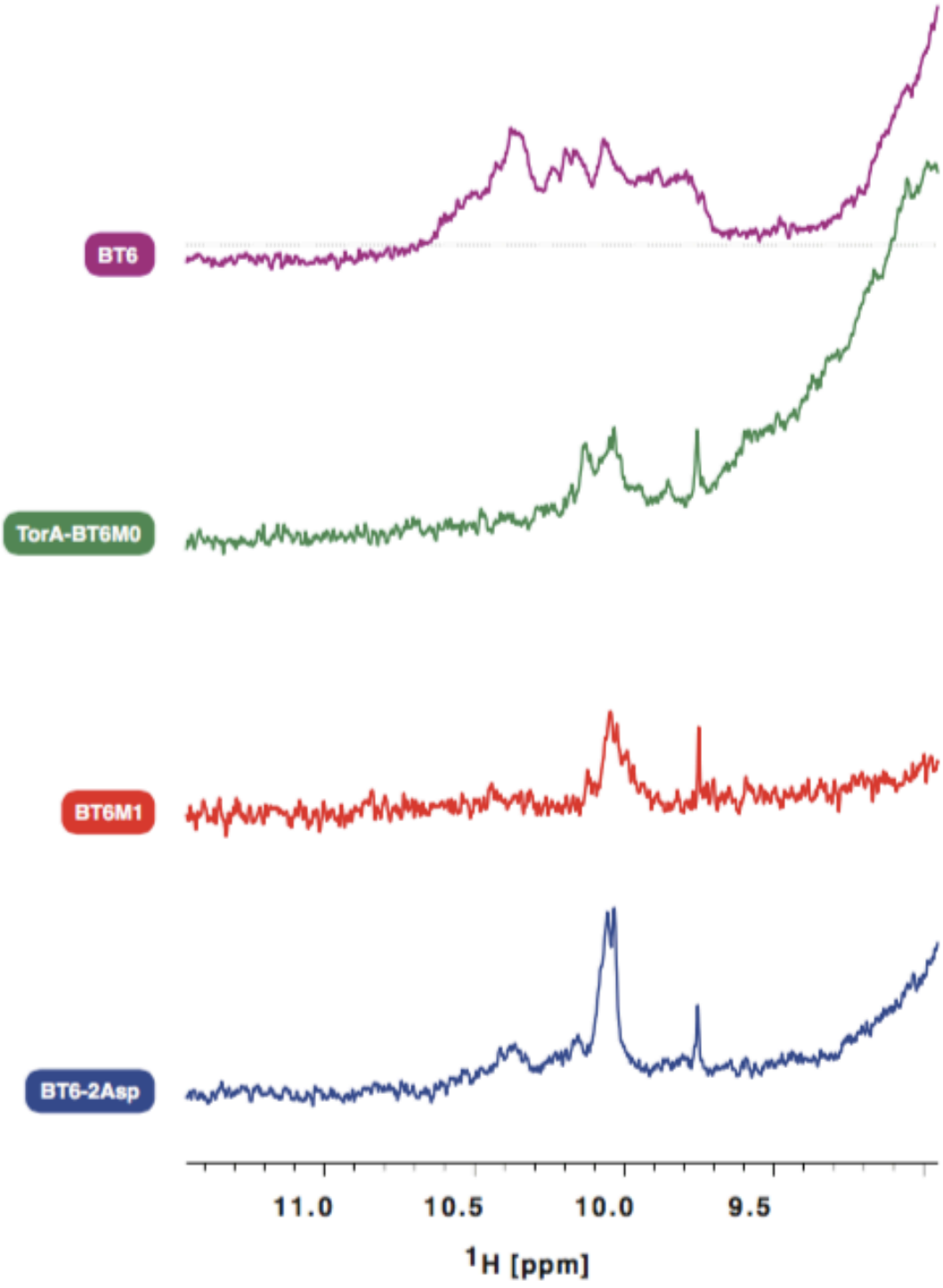


Figure 25 1D-1H-NMR spectra

BT6M1, TorA-BT6M0, and BT6-2Asp measured at 298K and 600MHz in 90%H_2O/10%D_2O using a DPFGSE water supressed pulse sequence. A blow-up of the region between 9.5 and 12.5ppm for each of the 1D spectra is displayed.

4.3 Discussion

The Tat translocase exports folded and cofactor containing proteins from the cytoplasm to the periplasm. It is not well understood how the Tat translocase senses the folding state of a substrate and how it prevents premature export to the periplasm. In this study, we investigated whether Tat's remarkable proofreading system senses outer surface residues as unfolded by using a *de-novo* co-factor binding 'maquette' protein. We therefore designed a variety of BT6 maquette mutants with various alterations on the substrate's outer surface. Export assays reveal whether the mutants were accepted or rejected by the Tat components.

We designed a mutant where we introduced hydrophobic patches on the surface (BT6-3NLeu). The export assay showed that changes in hydrophobicity were not tolerated by the Tat translocase. No export to the periplasm was observed. Unfortunately, no analysis of the folding state of this mutant has been possible. Moreover, it is thus uncertain whether the protein was rejected because of the hydrophobic outer surface patches or because the alterations perturb folding. Additional analysis on the structural confirmation of BT6-3NLeu could not be performed, as there was not enough protein to analyse. Most likely the limited presence of the protein was due to degradation of the BT6-3NLeu precursor as Tat recruits proteases after rejection of a Tat substrate (Matos *et al.,* 2008*)*.

We tested whether the Tat components will reject a folded substrate that contains an unfolded domain. Three BT6 mutants were designed with an additional 6-, 13- or 26-residue C-terminal unfolded domain (BT6-C-term26, BT6-C-term13 and BT6-C-term6). Tat sensed all three different sized unfolded domains, suggesting that in this case, the Tat

system senses conformational flexibility and has the ability to sense relatively short unfolded domains attached to a small *de-novo* designed α-helical protein.

To measure Tat's tolerance towards alterations of the outer surface charges we designed four different mutants. We designed two mutants with two or four negatively charged residues substituted to two and four positively charged residues (denoted BT6-2Lys and BT6-4Lys). Moreover, we designed two more mutants where two or four positively charged residues were substituted to two or four negatively charged residues (denoted BT6-2Asp and BT6-4Asp). BT6-4Asp was rejected by the Tat system, however the folding state of this variant is unclear. In contrast, BT6-2Lys, BT6-4Lys and BT6-2Asp were efficiently exported to the periplasm. These results suggest that Tat proofreading is tolerant of a range of both positively- and negatively-charged residues on the surface of substrate proteins.

The BT6-2Asp variant yielded unexpected information. As with BT6-2Lys and BT6-4Lys, this variant was efficiently exported, but whereas the former were shown to bind two heme molecules it is clear that BT6-2Asp binds only one, and this protein should therefore be partially unfolded, as was clearly shown for the other 1-heme mutant, BT6M1 (Sutherland et al., 2018). 1D-NMR data confirm that this is the case: BT6-2Asp and BT6M1 have nearly identical spectra which shows that at a general level their folding states are very similar indeed. We conclude that the Tat proofreading system does not assess the overall folding status of the substrate, but rather must respond to the folding state of localised domains in the substrate.

With BT6-2Asp and BT6M1, we believe that one of these important domains is one of the few regions that differ in terms of conformational flexibility, and the NMR data

furthermore raise the possibility that this domain may include unspecified Trp residues in the BT6 protein. The inference is that this domain is particularly rigid in the BT6-2Asp variant, and that the Tat system accepts this variant as folded because the domain is one of the primary interaction sites during the binding/ proofreading process.

Chapter 5:

Co-translational folding of Tat dependent proteins monitored *in vivo* and *in vitro*.

5.1 Introduction

As previously described, the twin arginine translocase (Tat) transports folded proteins across membranes. Tat is located in the inner membrane of bacteria, archaea, thylakoid membranes of plant plastids and mitochondria of plants (Bogsch *et al.*, 1998). Tat substrates play an important role in energy metabolism, formation of the cell envelope, iron and phosphate acquisition, cell division and cell motility (Hinsley *et al.*, 2001). The rationale for using Tat is not well understood. However, various reasons have been described for substrates to be translocated by the Tat system.

In order to bypass an extra mechanism to export the cofactor to the periplasm, some Tat dependent substrates fold around a co-factor in the cytoplasm prior to translocation (Berks, 1996). In addition, by insertion of metal ions in the cytoplasm, Tat dependent substrates avoid competition between metal ions that compete for insertion into the active site (Tottey *et al.*, 2008). The third rationale for the use of Tat is a so-called hitchhiker co-translocation mechanism. Several Tat dependent proteins form hetero-oligomeric complexes in the cytoplasm when a subunit binds a twin arginine containing subunit to form a dimeric protein (Rodrigue *et al.*, 1999). Moreover, there is data suggesting that, in archaea, Tat exports monomers and cofactor-less proteins because of their fast folding kinetics (Rose *et al.*, 2002).

In this study we tested whether Tat substrates fold in a co- or post-translational manner in order to determine how rapidly Tat substrates fold prior to translocation. To test whether Tat substrates fold co-translationally and if so at what point during translation they fold we performed arrest-peptide (AP)-mediated force measurements.

In recent years, several amino acid sequences have been found to interact with the ribosomal exit tunnel and therefore stall protein synthesis (Ito *et al.*, 2010). So-called 'arrest sequences' are involved in cellular processes like the regulation of SecA secretion carried out by secretion motor SecM in *E. coli* (Nakatogawa *et al*, 2004). Synthesized SecM stalls its own translation leading to an increased translation of SecA (Nakatogawa *et al*, 2001). The 17-residue long arrest peptide (AP) FSTPVWISQAQGIRAGP is essential for the stalling, which happens when the last proline codon is positioned at the A site of the ribosome (Nakatogawa et al., 2002, Muto et al., 2006, Garza-Sánchez et al., 2006). If however, there is dynamic or interaction of the nascent chain outside of the ribosomal exit tunnel, e.g. membrane protein insertion, protein translocation or protein folding, ribosomal stalling can be overcome and translation can continue beyond the AP (Ismail *et al.*, 2013). The mechanism of arrest release was later termed as 'pulling' model, which was originally suggested for SecM (Butkus et al., 2003).

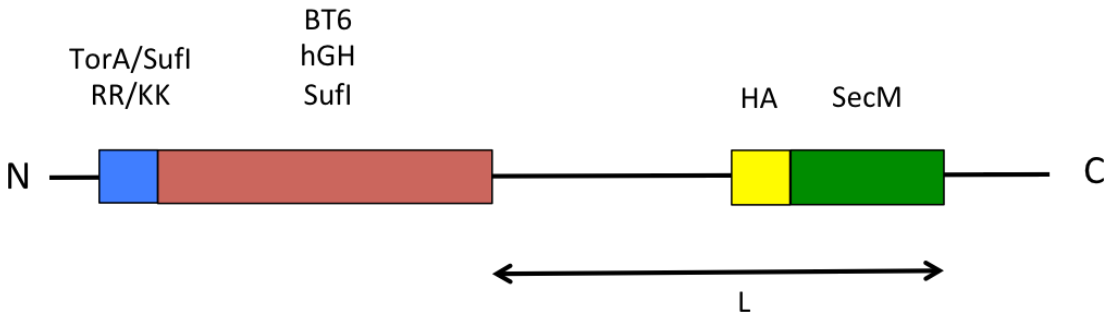


Figure 26 Design of constructs used in this study.

Substrates containing signal peptide with both a RR or a KK motif were designed. The substrate is placed linker (L) residues upstream of the C-terminal *E. coli* SecM AP. An additional 23-residues segment is added to the C-terminus in order to obtain a clean separation between arrested and full-length protein on a SDS PAGE gel. An HA tag was added for immunoprecipitation.

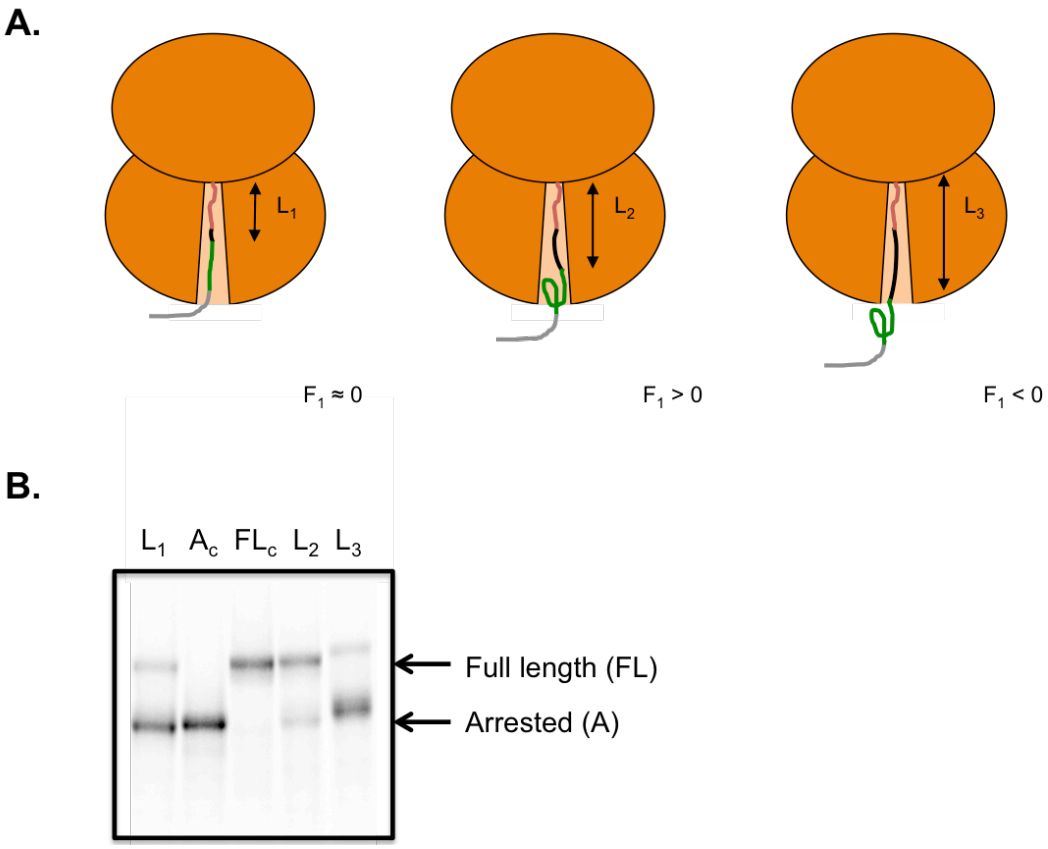


Figure 27 Arrest-peptide (AP)-mediated force measurement assay of the substrates

(A) The principle of arrest peptide (AP)-mediated measurement assay. The protein domain (green) is placed L residues (black) upstream of the C-terminal *E. coli* SecM AP (red). An additional 23-residues C-terminal domain (grey) was introduced to separate the arrested and the full-length version of the protein on a SDS-PAGE gel. Three constructs with different L values are displayed (L_1, L_2, and L_3). No or low pulling force (F) is generated at L_1 and L_3 as the protein did not fold yet at L_1, and the protein is already folded at L_3 when the ribosome reaches the AP. At L_2 there is a folding event that generates a pulling force (F) on the nascent chain. **(B)** Image of an expression provided by Felix Nicolaus, Stockholm University. Full-length (FL_c) and arrested-length (A_c) controls are indicated. Substitution of the crucial proline in SecM to a stop codon or alanine generates A_c or FL_c respectively. The full-length protein can only be visualised if pulling forces prevents AP-mediated stalling, leading to a continued translation beyond the SecM AP. At L_2, mainly full-length protein is visible whereas at L_1 and L_3 most of the protein is arrested. The fraction full-length (f_{FL}) can be plotted as a function of L to create a pulling force profile.

Translational APs, especially SecM have already been proven suitable to study the integration propensity of an artificial hydrophobic segment (8L/11A) during transition through the lateral gate of the SecY translocon in *E. coli*. This technique measures pulling forces acting on transmembrane segments during integration into the membrane. In addition, these force measurements have further been used to investigate the insertion of single transmembrane helices of multi-spanning membrane proteins into the *E. coli* inner membrane (Cymer *et al.*, 2013). More recently, the pulling force measurements of stalled ribosome-nascent chain complexes (RNCs) have been used to study co-translational folding of proteins (Nilsson *et al.*, 2015).

In this study, the AP from the *E. coli* SecM was used as a force sensor to study the folding kinetics of three Tat dependent substrates, human growth hormone (hGH), maquette BT6 and a natural substrate SufI. A variable linker (L) between the substrates and SecM AP was added, which was also previously used for *in vivo* studies (Ismail *et al.*, 2013, Cymer and von Heijne, 2013).

The design of the constructs is shown in Figure 26. The force profiles of the series of constructs generated by varying the number of residues (L) between the AP and the substrate provide a direct insight into the folding transition that the substrate undergoes as it is translated by the ribosome (figure 27). Substrates with a signal peptide containing a KK-motif in place of the normal RR-motif were also constructed as interaction of the substrate with the Tat translocon might also trigger a 'pulling' force, in which case the absence of the RR-motif would remove this force due to failure to interact with the Tat translocon. Previous cryo-EM (electron microscopy) studies showed that the substrate is located inside the ribosome tunnel at $L_{max} < 27$ residues (Nilsson *et al.*, 2015).

Both *in vivo* and *in vitro* pulse-labelling analyses were performed. The same constructs were used for both the *in vivo* and *in vitro* analyses (figure 26). For *in vitro* assays the constructs were cloned into a pET plasmid expressed using the PURExpress® *in vitro* protein synthesis kit. During a coupled transcription/translation reaction for 15 minutes, proteins were labelled with [^{35}S]-methionine. For *in vivo* assays the constructs were cloned into a pING plasmid comprising an arabinose-inducible promoter and expressed in *E. coli* MC1061 cells. Following the induction of the cells with arabinose for 5 minutes the cells were pulse-labelled with [^{35}S]-methionine for 2 minutes. For immunoprecipitation, a HA-tag (amino acid sequence: YPYDVPDYA) was introduced upstream of the SecM AP. The proteins were resolved on a SDS-PAGE to visualise the protein bands using a Fuji FLA-3000 phosphor imager and the ImageJ software. The proteins were quantified by using the EasyQuant software. All experiments were repeated in duplicate.

Translation stalling at the AP is overcome by a folding event due to an increased tension on the nascent chain upstream of the arrest peptide. This event will increase the fraction of the full-length protein, $f_{FL,}$ determined on the SDS-PAGE gel (figure 27B). The full-length protein (FL) can only be visualised if pulling forces prevents AP-mediated stalling, leading to a continued translation beyond the SecM AP. Then, the f_{FL} can be used to measure the force initiated by the folding event at a certain linker length (*L*). In the absence of a pulling force acting on the nascent chain an arrested (A) band will be visible. The f_{FL} is calculated by using the following calculation: f_{FL} = FL / (FL + A). Plotting the f_{FL} versus L can generate a so-called force profile, which shows how the folding force varies with the distance of the folding protein to the ribosome exit tunnel.

In addition, we investigated whether the Tat system can recognize and translocate substrates that fold co-translationally. ADR1a, Spectrin R16, superoxide dismutase

(SOD1), and protein G binding domain fold in a co-translational manner. ADR1a is a small zinc-finger domain protein that can fold inside the ribosome exit tunnel (Hedman *et al.*, 2015). Spectrin R16 serves for structural purposes (e.g. cytoskeletal interactions) (Nilsson *et al.*, 2017). SOD1 is an important antioxidant in nearly all living cells exposed to oxygen and the Protein G binding domain is the B1 binding domain of *Streptococcal* protein G for binding of the human Fc fragment of IgG (SOD1 and Protein G is unpublished data). A TorA signal peptide was fused to the N-terminus for periplasmic localization by the Tat system and for immunodetection a 6xHis tag was fused to the C-terminus. The constructs were cloned into a low copy pEXT22 plasmid and expressed in *E. coli* BL21 (DE3) and W3110 Tat Express cells. Export assays revealed whether the Tat system has the ability to export substrates that fold co-translationally.

5.2 Results

Force profile of BT6 in vitro *suggests a co-translational folding event*

To investigate whether Tat dependent substrates fold co- or post-translationally, we tested the previously described four α-helix di-heme binding BT6 maquette substrate. As described in chapter 3, BT6 maquette is a cofactor containing substrate, which folds inside the cytoplasm and is efficiently translocated to the periplasm in a Tat system dependent manner. A series of constructs was designed in which BT6 was inserted variable linker (L) residues upstream of the *E. coli* SecM AP comprising a critical C-terminal proline residue, where translation stalls. An additional 23 residues followed by a stop codon were inserted downstream of the AP (figure 26). In addition to the series of BT6 constructs, full-length (FL_c) and arrested (A_c) controls were tested. If no pulling force is acting on the nascent chain when the ribosome reaches the AP, the translation will stall resulting in a short arrested protein. In contrast, only if a pulling force acts on the nascent chain at the precise moment when the ribosome reaches the proline codon, stalling will be overcome and mostly the full-length protein will be produced (figure 27).

The fraction of full-length (f_{FL}) was plotted as a function of L to create a pulling force profile (figure 28). To differentiate between pulling forces caused by a folding event or a possible interaction with the Tat components two BT6 maquette variants were tested. One BT6 construct contained a native twin arginine TorA signal peptide, whereas the second BT6 construct contained a twin lysine TorA signal peptide. *In vivo* translation of these two BT6 maquette variants in *E. coli* MC1061 cells was performed in duplicate and showed a similar force profile. The force profile shows that the f_{FL} starts to increase at L ≈ 45 residues. There is a peak at L ≈ 50 residues after which it returns at the baseline at L ≈ 60

residues. It has to be mentioned that the increased f_{FL} values is only minimal. However, this data indicates a co-translational folding event of BT6 maquette. Possible interaction with the Tat components during translation was not visible as both BT6 variants show a similar force profile. The reason for a limited f_{FL} increase can be due the presence of chaperones that might inhibit co-translational folding, such as GroEL (Nilsson *et al.*, 2016).

To generate a stronger force profile, *in vitro* translation of BT6 was performed in duplicate using the PURE *in vitro* translation system. Therefore the constructs were cloned into a high copy pET vector. In this study only the twin arginine TorA BT6 maquette was used, as there are no Tat components present in the *in vitro* translation reaction mix. Figure 28 shows the pulling force profile of BT6 *in vitro,* presenting that the f_{FL} starts to increase at L ≈ 35 residues to reach a sharp peak at L ≈ 40 residues. An increase of pulling force from L ≈ 30 residues is reasonable as 30 aa is the approximate length of a nascent that the ribosome exit tunnel can accommodate. Therefore, at L ≈ 30 residues, BT6 is supposed to be completely outside the ribosome, which gives the protein enough conformational space to fold. For the *in vitro* force profile, a more significant increase in f_{FL} is detectable, compared to the f_{FL} values of the *in vivo* expression. The lack of proteases and chaperones results in a stronger pulling force acting on the nascent chain. It has to be mentioned that BT6 has to fold around two heme *b* cofactors to fold completely, which is absent in the *in vitro* translation system. Therefore the force profile does not represent complete folding of BT6 resulting in weaker f_{FL} values. Hence, the *in vitro* expression should be repeated in the presence of heme in order to see the complete folding event. Another noticeable difference is that the *in vitro* f_{FL} peak is shifted compared to the *in vivo* f_{FL} peak. Previous studies have demonstrated that this occurs in the absence of chaperones, such as trigger factor (Nilsson *et al.*, 2016).

We suggest that BT6 substrates might fold co-translationally as force acting on the nascent chain can be determined *in vivo* and more specifically *in vitro* when the protein emerges from the ribosome exit tunnel. *In vivo,* the force profiles of the two TorA-BT6 variants showed a similar curve, indicating that there was no detectable interaction with the Tat components during translation and folding by the pulling force measurements. It is possible that there is interaction with the Tat components, which however is not significant enough to be captured by the AP-mediated force measurements assay.

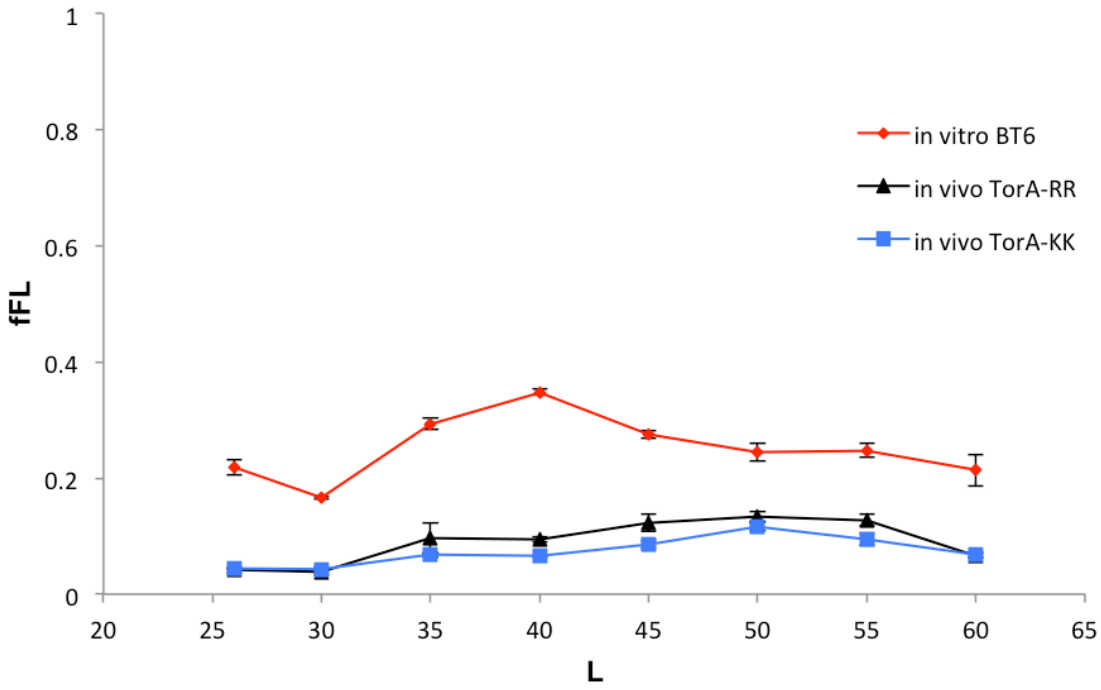


Figure 28 Pulling force profile of BT6 *in vitro* and *in vivo*

The fraction full-length (f_{FL}) was plotted as a function of L to create the force profile. The force profile of TorA-RR-BT6 (black) and TorA-KK-BT6 (blue) constructs translated *in vivo* in *E. coli* MC1061 cells is shown. These force profiles show that the f_{FL} starts to increase at L ≈ 35 residues. There is a peak at L ≈ 50 residues after which it returns at the baseline at L ≈ 60 residues. The force profile of TorA-RR-BT6 translated in the PURExpress® *in vitro* system is indicated in red. The f_{FL} starts to increase at L ≈ 30 residues to reach a sharp peak at L ≈ 40 residues.

Pulse labelling assay of hGH in vitro *and* in vivo *suggests a co-translational folding event*

In addition to the previously discussed force profile of BT6 maquette we tested the heterologous human growth hormone (hGH). hGH is a cofactor less protein that can fold inside the cytoplasm and is efficiently exported by the Tat system (Alanen *et al.*, 2015). A similar series of constructs, with varying amounts of L residues, was designed as described for BT6. In addition, two hGH variants, TorA-RR-hGH and TorA-KK-hGH, were tested to identify possible forces caused by interaction with the Tat components.

In vivo translation of these two hGH variants in *E. coli* MC1061 cells was performed in duplicate and fairly identical force profiles were seen. The fraction full-length (f_{FL}) was plotted as a function of L to create a pulling force profile (figure 29). The f_{FL} starts to increase when the protein emerges from the ribosome exit tunnel at L ≈ 35 residues going through a maximum at L_{max} ≈ 40 - 45 residues after which it returns back to the baseline at L ≈ 60 residues. This rather significant peak suggests a co-translational folding of hGH outside of the ribosome exit tunnel *in vivo.* As observed for BT6, a possible interaction with the Tat components during translation could not be detected by using this technique as both hGH variants show a similar force profile.

To investigate co-translational folding in the absence of inhibiting chaperones the TorA-RR-hGH constructs were cloned into a pET vector for translation in the PURExpress® *in vitro* system. The constructs were translated in the absence and presence of a disulfide bond enhancer, as hGH is a protein that contains disulfide bonds in the native state. Surprisingly, the *in vitro* pulling force profile of hGH shows that the f_{FL} already starts to increase at L ≈ 26. At this stage of translation the protein is still inside the relatively narrow ribosome exit tunnel. The force profile goes through a peak at L ≈ 30 - 35 residues

when the protein has emerged from the ribosome exit tunnel. Only the PURExpress® reaction mix, containing the disulfide bond enhancer, enables disulfide bond formation for complete folding resulting in an overall higher pulling force. Because the chaperones, trigger factor and GroEL, reduce the force on the nascent chain *in vivo*, the *in vitro* f_{FL} peak has higher values and is shifted to shorter linker lengths compared to the *in vivo* f_{FL} peak, as it was seen for BT6 as well.

The fact that the pulling force in the *in vitro* assay never reaches zero values could be explained by interaction of the emerging proteins with the surface of the ribosome. The tested proteins have an overall negatively charged surface in the folded state, which results in repelling dynamics at the negatively charged ribosome surface. These dynamics could in turn lead to pulling force on the nascent chain even after folding has taken place. However, this cannot be seen *in vivo* since chaperones might shield the nascent chain from interaction with the ribosome.

The data indicates that hGH might fold co-translationally as force acting on the nascent chain can be determined *in vivo* and *in vitro*. Moreover, the data implicates that preliminary folding can already take place inside the ribosome exit tunnel. However, it is theoretically unlikely that the rather large hGH can already fold at this stage. We therefore propose that either first folding intermediates start to form before the protein has completely left the tunnel or the high number of negatively charged residues already interact with the surface of the ribosome. It remains to be uncertain whether interaction with the Tat components occurs during translation, as it was not detected using the force measurements assay.

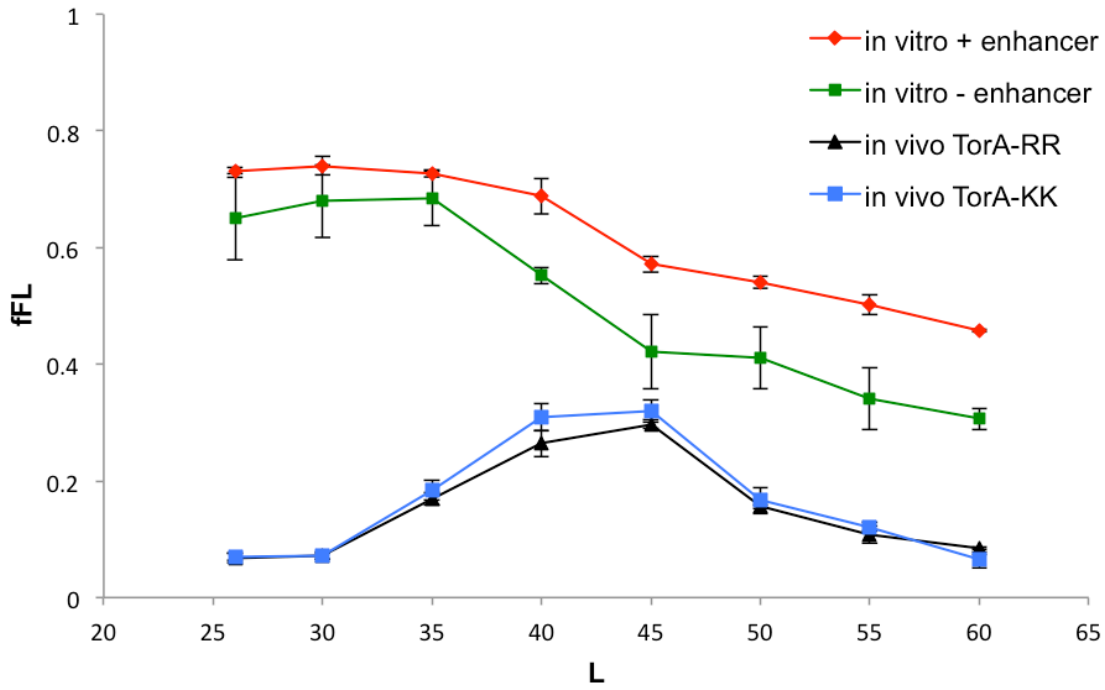


Figure 29 Pulling force profile of hGH *in vitro* and *in vivo*

The *in vivo* force profiles TorA-RR-hGH (black) and TorA-KK-hGH (blue) show that the f_{FL} starts to increase at L_{max} ≈ 30 – 35 residues. There is a peak at L_{max} ≈ 40 - 45 residues after which it returns at the baseline at L ≈ 60 residues. TorA-RR-hGH was translated in the PURE *in vitro* system in the presence (red) and absence (green) of a disulfide bond enhancer. The f_{FL} starts to increase at L ≈ 26 residues to reach a peak at L ≈ 35 residues. *In vitro,* higher f_{FL} values were measured in the presence of a disulfide bond enhancer.

Pulse labelling assay of SufI suggests a possible co-translational interaction with Tat

In addition to the force profiles of BT6 maquette and hGH we tested the natural *E. coli* model Tat substrate SufI (FtsP). The SufI protein is frequently used as a model substrate for studies of the Tat system due to the favourable properties, as it is water soluble, monomeric and it does not require cofactor insertion during folding. As before a series number of constructs, with varying amounts of L residues and both RR and KK variants were tested.

In vivo translation of these two SufI variants in *E. coli* MC1061 cells was performed in duplicate. The fraction full-length (f_{FL}) was plotted as a function of L to create a pulling force profile (figure 30). Unfortunately, no significant pulling force could be detected for SufI KK. This observation is reasonable as SufI is a significantly larger protein than BT6 and hGH and therefore needs more space outside of the ribosomal exit tunnel. However, it has been seen that multi-domain proteins can fold co-translationally (Zhang and Ignatova, 2011). Another interesting observation is the fact that even though the f_{FL} values are weak, the force profile of the SufI-RR variant shows higher f_{FL} values compared to the force profile of the SufI-KK variant. The different force profiles might be caused by a possible interaction with the Tat components during translation, as only the SufI-RR variant can be recognized and translocated by the Tat system.

Another indication for force initiated by interaction with the Tat components is the *in vitro* force profile of the two SufI variants (figure 31). The *in vitro* force profile shows a very weak increase in f_{FL} as seen for the *in vivo* expression assays. However, the difference is that the force profile of SufI-RR and SufI-KK *in vitro* are similar. The fact that Tat

components are absent in the *in vitro* expression assay might be the reason why we see similar force profiles if the force is initiated by interaction with Tat.

The data we obtained does not give indication for a co-translational folding event, however the assay provides first insight in the interaction of SufI with the Tat components. One reason why there is no detectable folding of SufI is that it is too large in order to be folded in the set-up that we choose. Introducing longer linker length downstream of SufI to give it more conformational freedom around the ribosomal exit tunnel could solve this issue. Another reason *in vivo* could be that the folding is inhibited by chaperones or initial interactions with Tat.

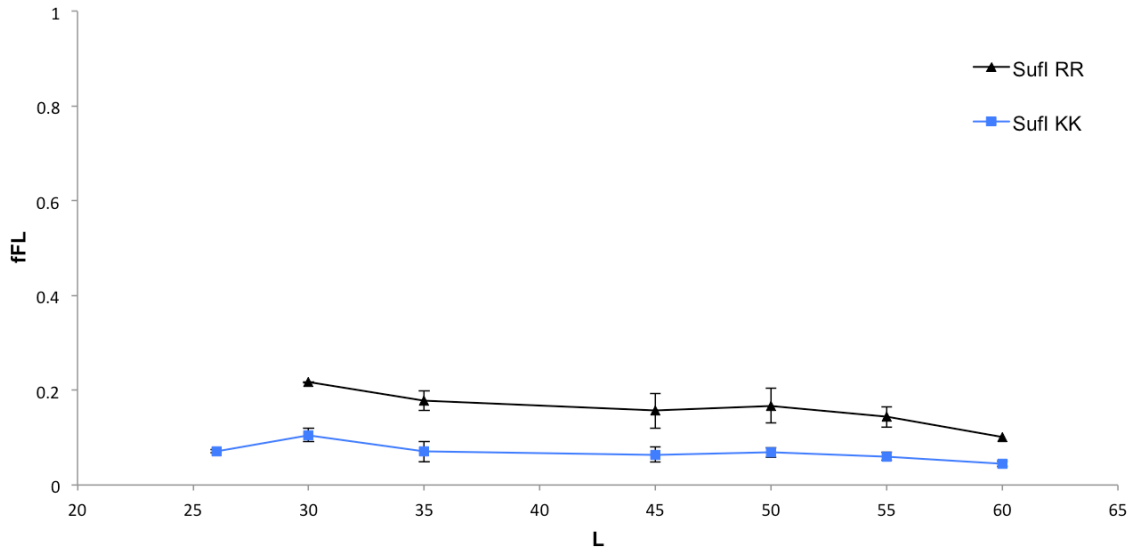


Figure 30 Pulling force profile of SufI *in vivo*

The SufI constructs were translated *in vivo* in *E. coli* MC1061 cell. The force profiles of both SufI-RR (black) and SufI-KK (blue) show overall weak pulling force. The force profile shows that there is stronger force acting on the nascent chain of the SufI-RR variant compared to SufI-KK.

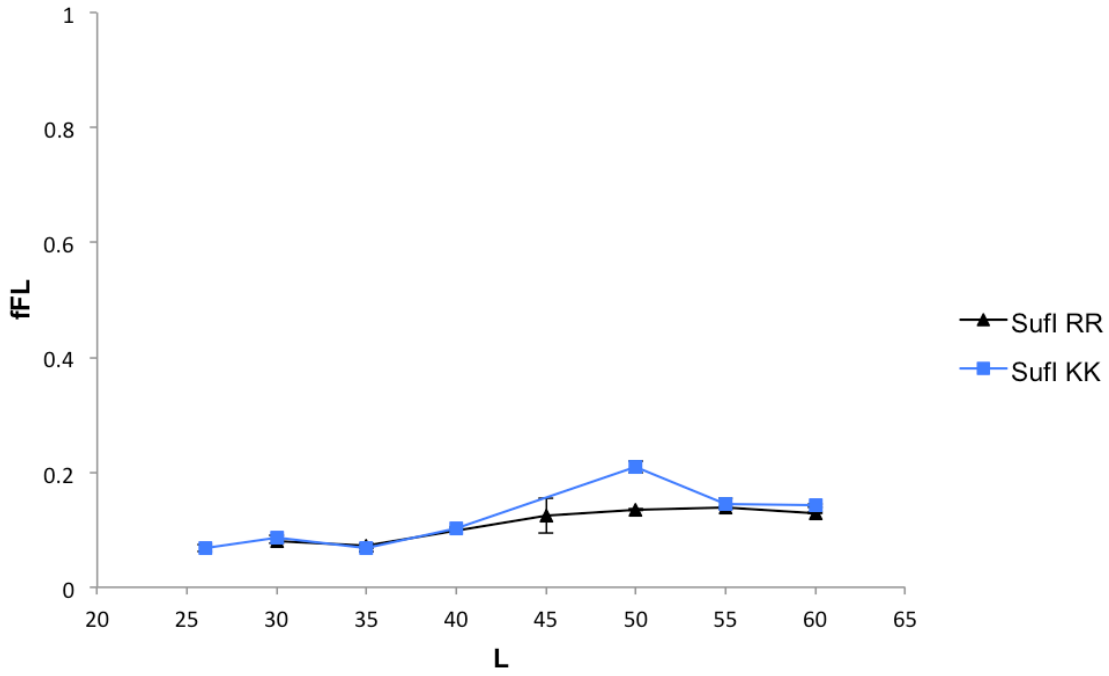


Figure 31 Pulling force profile of SufI *in vitro*

The *in vitro* force profiles of SufI-RR (black) and SufI-KK (blue) are shown. The fraction full-length (f_{FL}) was plotted as a function of L to create the force profile. No pulling force acting on the nascent chain was detected using the AP-mediated force measurements assay. The force profiles both SufI variants show a similar curve.

The Tat translocase can translocate co-translationally folded proteins

To determine whether the Tat translocase exports co-translationally folded proteins we cloned four different proteins into a low copy pEXT22 plasmid. The proteins of interest, ADR1a, R16, SOD and protein G B1 binding domain, fold in a co-translational manner. For immunodetection a 6xHis tag was fused to the C-terminus and a TorA signal peptide was fused to the N-terminus for periplasmic localization by the Tat system. The constructs were expressed in *E. coli* BL21 (DE3) and W3110 Tat Express cells (Browning *et al.*, 2017). Export assays revealed whether the Tat system has the ability to export co-translationally folded substrates. Expression assay with ADR1a, R16, SOD and protein G B1 binding domain were performed to test whether the Tat system can export co-translationally folded substrates. The cells were induced with 0.5 mM IPTG and 3 hours post induction the cells were harvested and fractionated into cytoplasmic, membrane and periplasmic fractions (C/M/P). The presence of the substrates in the cell fractions was determined by immunoblotting.

Figure 31 shows the results of the export assay of ADR1a in *E. coli* BL21 (DE3) and W3110 Tat Express cells. It is to mention that ADR1a is a very small protein, which is difficult to detect on a SDS PAGE gel. Therefore, the low molecular weight proteins were separated using tricine protein gels. Despite using tricine gels, the image does not show a 3.5-kDa mature sized band in the periplasmic fraction (P). In addition, no precursor proteins were detected in the cytoplasmic (C) or membrane (M) fractions. ADR1a is a cofactor containing protein and folds around a Zn^{2+} ion. It remains unclear whether ADR1a was rejected by the Tat system due to the lack of Zn^{2+} ions in the cytoplasm. However, the small size of ADR1a might be the limiting factor in visualizing exported protein to the periplasm as it might run off the SDS gel too quickly to analyse.

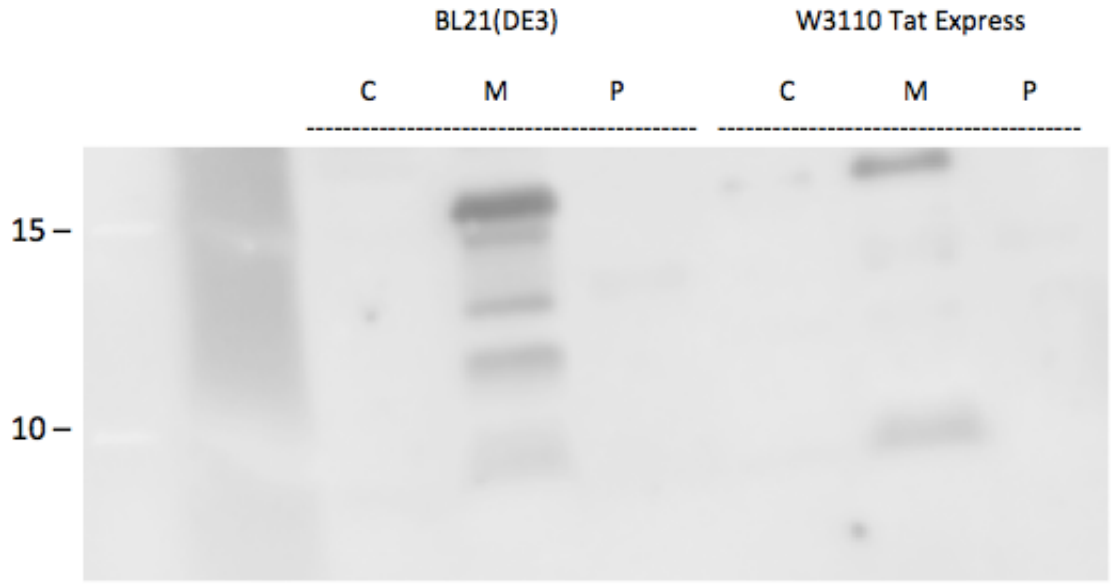


Figure 32 The Tat system does not export the zinc cofactor binding ADR1a.

E. coli BL21 (DE3) and W3110 Tat Express cells were fractioned to cytoplasmic (C), membrane (M) and periplasmic (P) fractions. The fractions were immunoblotted using a C-terminal His antibody. The molecular mass markers (kDa) are shown on the left. The export assay of the ADR1a protein shows no mature sized bands of (3.5 kDa) in the periplasmic fractions (P). There are no precursor proteins visible in the cytoplasmic (C) and membrane (M) fraction observed for BT6-4Asp.

As mentioned previously we tested whether three more cotranslational folders can be exported by the Tat translocase system. The export assay was performed in the manner as described for ADR1a and fractionated into cytoplasmic, membrane and periplasmic fractions (C/M/P). The substrates present in the three different cell fractions were visualised by immunoblotting using a C-terminal his antibody.

The results of the export assay of, R16, SOD and protein G B1 domain are shown in figure 32. The results of the expression in *E. coli* W3110 Tat Express cells are displayed in figure 32A, whereas figure 32B shows the results of the export assay in *E. coli* BL21 (DE3). The expected molecular weight of the substrates is 13.4 kDa for R16, 11.9 kDa for SOD and 7 kDa for Protein G B1 domain. The proteins were separated using SDS-PAGE gels, as the molecular weight of the proteins is not as low as for ADR1a. Figure 32A shows mature sized bands for R16, SOD and Protein G B1 domain after expression in *E. coli* W3110 Tat Express cells. Unprocessed precursor bands are present in the cytoplasmic (C) and membrane (M) fraction. In addition, also after expression in *E. coli* BL21 (DE3) cells, mature sized bands are visible the periplasmic fractions of all three different substrates.

This data shows that substrates that undergo co-translational folding can be recognised and translocated by the Tat system. This can be an indication that some substrates might be exported by the Tat system because of their fast folding kinetics. Due to the rapid folding within the cytoplasm the Sec translocase system is incompetent to export these substrates. This might result in the targeting of the folded substrates by the Tat system.

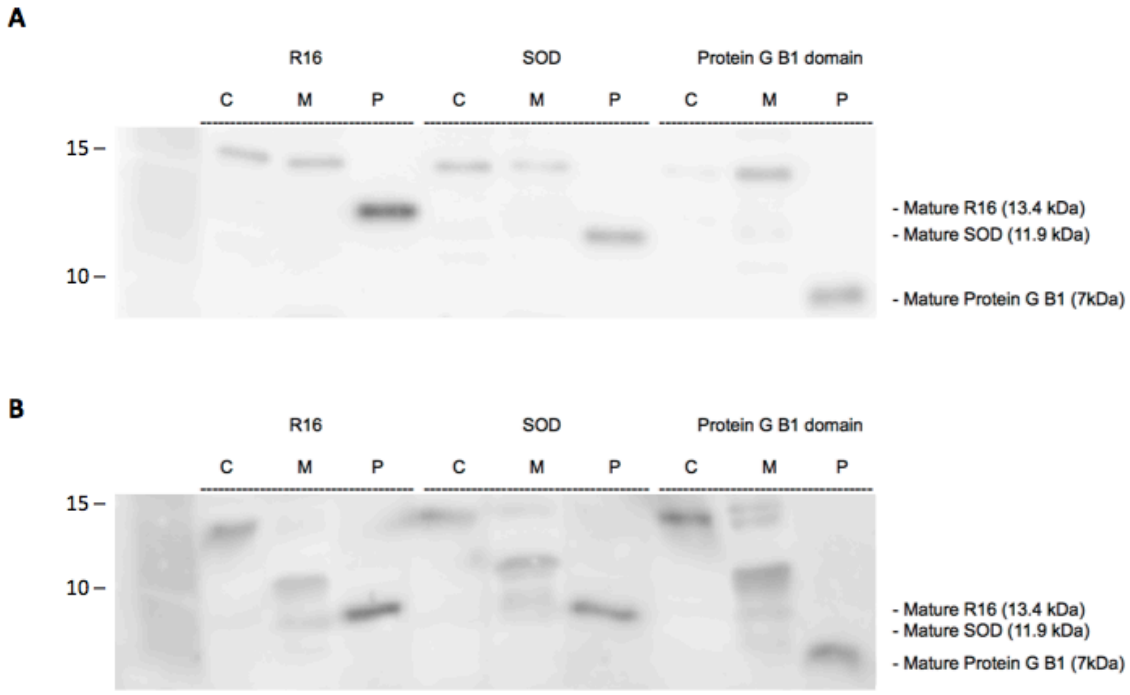


Figure 33 Export of co-translationally folded substrates by the Tat system

(A) Representation of the expression of R16, SOD and Protein G B1 domain. The cells were fractioned to cytoplasmic (C), membrane (M) and periplasmic (P) fractions and these fractions were immunoblotted using a C-terminal His antibody. The molecular mass markers (kDa) and the expected size of mature bands are indicated alongside of the blot. The expected mature sizes of the substrates are 13.4 kDa for R16, 11.9 kDa for SOD and 7 kDa for Protein G B1 domain. (A) The immunoblot of the cell fractions after expression in *E. coli* W3110 Tat Express cells. Mature sized bands are visible in the periplasmic fractions of R16, SOD and Protein G B1 domain. For all three substrates precursor bands are present in the cytoplasmic (C) and membrane (M) fraction. (B) The immunoblot of the cell fractions after expression in *E. coli* BL21 (DE3) cells. As observed after expression in W3110 Tat Express cells, mature sized bands are visible in the periplasmic fractions of R16, SOD and Protein G B1 domain. Despite sporadic unspecific binding of the antibody, precursor bands are present in the cytoplasmic (C) and membrane (M) fraction for all three substrates.

5.3 Discussion

Most of the secretory substrates are translocated via the Sec pathway. Unfortunately, the Sec system cannot translocate those substrates that fold rapidly within the cytoplasm. Therefore, the Tat system exports proteins in a folded state, however the rationale for why these proteins cannot be exported in an unfolded state by the Sec pathway is not well understood. Only a limited amount of reasons for the use of the Tat system are described, such as co-factor insertion, avoidance of metal ion competition and hetero-oligomeric complex formation (Palmer and Berks, 2012). Some proteins fold in the cytoplasm because they require reducing conditions for folding and therefore are unable to fold in the oxidising environment of the periplasm. For archeae, it has been suggested that substrates are translocated to the periplasm by the Tat system because of their rapid folding kinetics (Rose *et al.*, 2002). Nevertheless, the rationale for using the Tat system is not well understood. In this study, we tested whether the Tat system exports substrates because of their possible rapid folding kinetics in *E. coli*. Therefore, we investigated whether Tat dependent substrates fold co- or posttranslationally by using AP-mediated force measurements. The three Tat substrates used in this study were BT6 maquette, hGH, SufI.

In vivo expression assays with BT6, showed increased pulling forces at L = 45 indicating a possible co-translational folding event after the protein has emerged from the ribosome exit tunnel. We tested BT6 substrates with a twin-arginine signal peptide as well as substrates with a twin-lysine signal peptide variant to investigate whether folding or an interaction with the Tat components initiated the force acting on the nascent chain. However, possible interaction with the Tat components during translation was not visible, as the force profile of these two TorA signal peptide variants showed a similar curve. To avoid chaperones that might inhibit co-translational folding, we investigated the force

profile of BT6 *in vitro*. The pulling force profile of BT6 *in vitro* showed a stronger force profile where the f_{FL} starts to increase at L ≈ 30 residues. Even though, the *in vitro* expression was done in the absence of the BT6 binding cofactor heme *b*. We suggest that this data indicates that BT6 substrates might fold co-translationally, as force acting on the nascent chain was seen *in vivo* and more specifically *in vitro* when the protein emerges from the ribosome exit tunnel. The force profiles of the two TorA-BT6 variants showed a similar curve, indicating that there was no detectable interaction with the Tat components during translation and folding by the pulling force measurements. It is possible that there is interaction with the Tat components, which however is not significant enough to be captured by the AP-mediated force measurements assay.

Similar *in vivo* and *in vitro* force profiles were observed for the expression assays with hGH, albeit with a stronger pulling force acting on the nascent chain. This might be due to different chaperones needed for the folding of hGH and BT6 which might interfere with the folding process. In addition, there is a noticeable difference between the *in vitro* and *in vivo data* of BT6 and hGH. The *in vitro* f_{FL} peaks are shifted compared to the *in vivo* f_{FL} peaks. Previous studies have demonstrated that this occurs in the presence of chaperones, such as Trigger factor and GroEL (Nilsson *et al.*, 2016).

The *in vivo* force profile data of the natural Tat substrate SufI did not show a folding event. Surprisingly, the force profile of the SufI-RR variant shows higher f_{FL} values compared to the force profile of the SufI-KK variant *in vivo*. The different *in vivo* force profiles and the *in vitro* data might suggest a possible interaction of SufI with the Tat components during translation.

In addition to the force profiles, export assays revealed that the co-translational folding substrates, R16, SOD1 and Protein G B1 binding domain, could be exported via the Tat pathway. Only for ADR1a, efficient export to the periplasm was not detected. It might be possible that ADR1a did exported by the Tat system but was not detected because of the very low molecular weight of this protein (3.5 kDa). A second reason for no visible export to the periplasm could be a lack of zinc within the cytoplasm, which is required for folding of ADR1a. The force profile of ADR1a was done *in vitro* in the presence of an access of zinc (Nilsson *et al.*, 2015). This could suggest that ADR1a was not exported by the Tat system due to insufficient folding because of the lack of zinc cofactors. Optimization of the export assays of ADR1a is necessary to confirm whether the Tat system has the ability to export a small zinc finger binding domain protein.

Chapter 6:

Final discussion

In Gram-negative bacteria, more than one third of the proteome are secreted across the inner and outer membrane (Economou, 2010a). In *E. coli*, the twin arginine translocase (Tat) is one of the major pathways to transport proteins from the cytoplasm to the periplasm. Despite the periplasm being the site of function for a lot of secretory proteins, some of these proteins require folding in the cytoplasm. The Tat system only allows export of folded proteins for which it consists of a proofreading and quality control mechanism. Various studies have demonstrated the Tat system's ability to reject unfolded proteins (Robinson *et al.*, 2011, Maurer *et al.*, 2009). However, it is not well understood how the Tat proofreading mechanism is able to determine the folding state of a substrate. The aim of this study was to increase our understanding on what the Tat proofreading mechanism can sense in order to 'proofread' the substrate's structural conformation.

Only a limited amount of reasons for the use of the Tat system are described, such as co-factor insertion, avoidance of metal ion competition and hetero-oligomeric complex formation (Palmer and Berks, 2012). Some proteins fold in the cytoplasm because they require reducing conditions for folding and therefore are unable to fold in the oxidising environment of the periplasm. For archeae, it has been suggested that substrates are translocated to the periplasm by the Tat system because of their rapid folding kinetics (Rose *et al.*, 2002). Nevertheless, the rationale for using the Tat system is not well understood. In this study, we tested whether possible rapid folding kinetics of Tat substrates could be another reason for using Tat in addition to the previously described reasons. Therefore, we investigated whether Tat dependent substrates fold co- or posttranslationally by using AP-mediated force measurements. In this chapter the major findings in this study will be summarised.

Probing the quality control mechanism of the twin-arginine translocase with folding variants of a *de novo*-designed heme protein.

The first chapter in this study probed the quality control mechanism of the Tat system with folding variants of a *de novo* designed substrate, denoted BT6 maquette. This was accomplished by reducing the number of heme cofactors in order to perturb the folding state of BT6 maquette. Therefore, two BT6 folding variants were generated by point mutations affecting the heme binding sites to prevent the ligation of 1 or two heme cofactors (figure 9).

Export assays revealed that the Tat system has the ability to recognize and export BT6 maquette substrates (figure 10). Moreover, there seems to be a correlation between the conformational flexibility and the export efficiency of the folding variants (figure 12). The Tat proofreading mechanism rejected the unfolded heme deficient BT6M0 (figure 10) and the partially folded one heme binding BT6M1 substrate was exported in a less efficient manner (figure 11). In addition, the heme content has an effect on the conformational flexibility of BT6 maquette proteins as confirmed by ^{1}H NMR spectroscopy and CD spectroscopy analysis (figure 14). However, in order to determine whether the Tat system rejected BT6M0, additional experiments are required. There is a possibility that the BT6M0 substrate was aggregated and is present in inclusion bodies. This can be determined by harvesting a fourth cell fraction, the insoluble fraction.

In order to confirm the presence or absence of an interaction of BT6 substrates with the Tat components, FRET or cross-linking studies could be applied. This would give information on whether the substrates and Tat are in close vicinity. However, such approaches are

rather time-consuming and would add alterations to the structure, which might affect the proofreading mechanism of Tat.

In summary, the results of this study suggest that the Tat system can recognize the BT6 maquette proteins and that the proofreading mechanism can discriminate between folded states. This suggests that the Tat proofreading pathway involves interactions between Tat components and the substrate at the membrane surface to sense the structural flexibility and overall export competence of a substrate. This cascade of interactions can sense flexible conformations, resulting in the rejecting of an unfolded substrate. This study shows that BT6 maquette is an ideal candidate to study the Tat systems proofreading mechanism. By using BT6 in future studies, our understanding of the proofreading mechanism will increase. This proof-of-principle technique we will allow us to probe the tolerance of the Tat systems proofreading mechanism.

The twin-arginine proofreading ability of *Escherichia coli* can sense localized unfolded regions of a man-made maquette protein variant.

After determining that BT6 maquette is an ideal candidate to study the Tat proofreading mechanism, we tested whether Tat can sense outer surface characteristics as unfolded by BT6. Therefore a variety of BT6 maquette mutants were designed with altered outer surface residues. In this study, BT6 substrates with an altered surface electrostatic potential, an increased surface hydrophobicity, and with an additional unfolded domain were tested. Export assays revealed whether the mutants were accepted or rejected by the Tat proofreading mechanism.

The export of the BT6 variants with an increased surface hydrophobicity and an additional unfolded domain was perturbed. However, it is uncertain whether Tat rejected the protein due to hydrophobic surface patches or the folded state as the alterations in surface hydrophobicity might have counteracted the folding of BT6. In order to determine whether the introduction of the hydrophobic surface patch perturbed the folding of this BT6 variant, additional experiments need to be performed. In the future this protein could be expressed *in vitro* as initial attempts to purify the protein after an *in vivo* expression failed. We suspect that the protein degrades too quickly to analyse due to proteases recruited by Tat after rejection of the protein. *In vitro,* there are no proteases present to degrade the protein resulting in a higher protein concentration after expression and purification. The overall conformational state of the purified BT6 variant can be investigated by using 1D NMR and CD spectroscopy experiments. This will give us more information on why the Tat system did not export this variant.

The majority of the variants with alterations in the surface electrostatic potential were exported efficiently by the Tat system. This data shows that the Tat proofreading is tolerant to an increase in both positively- and negatively-charged residues on the surface of substrate proteins. These findings correspond with other studies, which have suggested that the Tat proofreading mechanism is tolerant to surface charge (Jones *et al.*, 2016).

Surprisingly, spectroscopic measurements revealed that one BT6 variant with an increased negatively charged surface only binds one heme. Previously we have shown that the heme content has an effect on the conformational flexibility of BT6 maquette proteins. However, 1D NMR data suggests that this one heme binding BT6 variant is possibly more tightly folded than the di-heme binding BT6. Therefore we suggest that the proofreading mechanism of the Tat translocase has the ability to sense structural flexibility rather than

outer surface characteristics. Moreover, the Tat system might sense localised flexible regions rather than a global unfolded state of the substrate.

In this study, we have shown that the Tat proofreading mechanism has the ability to sense the localised unfolded regions to assess the conformational state of the substrate. In order to increase our insight of these unfolded regions, ^{15}N NMR or coupled ^{1}H -^{15}N NMR spectroscopy analysis could be performed as a more sensitive alternative to the ^{1}H NMR spectroscopy analysis performed in this study. The data of these additional experiments in addition to the data presented in this study will increase our understanding of the Tat proofreading mechanism. Moreover, the previously discussed methods such as FRET and cross-linking studies can give more insight into the dynamics of the proofreading mechanism of the Tat system. In addition, the mechanism of the Tat proofreading ability remains unclear as we are lacking structural data of the Tat(A)BC complex interacting with the Tat substrates. Further high-resolution structural analysis of the Tat components is needed in order to test the proofreading ability of the different Tat components and to determine which domains are responsible for proofreading the structural conformation. Hopefully, the relatively new electron cryomicroscopy (Cryo-EM) technique will broaden our structural knowledge on the different Tat components in the near future. Mutagenesis studies on the Tat substrates and Tat components can lead to an increased insight in the Tat proofreading mechanism. Finally, this study and previous studies primarily focused on heterologous and artificial proteins. Therefore, mutagenesis studies on native Tat substrates as described in this study will be a valuable addition to our knowledge.

Co-translational folding of Tat dependent proteins monitored *in vivo* and *in vitro*.

In the third study, we tested whether *E. coli* Tat substrates fold co- or posttranslationally by using AP-mediated force measurements. The obtained force profile of the BT6 substrates might indicate co-translational folding, as force acting on the nascent chain was seen *in vitro* when the protein emerges from the ribosome exit tunnel. Moreover, a possible interaction with the Tat components during translation and folding was not detected. However this possibility is not excluded, as the force generated by the interaction with Tat might not be significant enough to be captured by the AP-mediated force measurements assay.

Compared to the force profile of BT6, stronger pulling force acting on the nascent chain was observed for hGH. Possibly, because hGH and BT6 require different chaperones, which might interfere differently with the folding process. Interestingly, the pulling forces probably generated by folding were reduced and detected later *in vivo* during translation. This was previously described in studies that demonstrated that this occurs in the presence of chaperones, such as Trigger factor and GroEL (Nilsson *et al.*, 2016). Furthermore, a noticeable increase of the folding force was observed when expressed *in vitro* in the presence of disulfide bond enhancer, which helps hGH to form its disulfide bonds.

We also tested whether the natural *E. coli* Tat substrate SufI folds co- or posttranslationally. Surprisingly, a possible interaction of SufI with the Tat components was detected in the *in vivo* force profile as the native SufI-RR variant shows higher pulling force values compared to the force profile of the SufI-KK mutant. In addition, the *in vitro* expression in the absence of Tat components showed similar force profiles, suggesting a possible interaction of SufI with the Tat components during translation.

Export assays of co-translationally folding substrates revealed that the Tat system could export most of these substrates (SOD1, R16 and Protein G B1 domain). A small zinc finger binding protein, ADR1a, could not be exported according to immunoblot data. However, the substrate might have not been able to fold in the cytoplasm due to a lack of zinc, which is required for folding. Moreover, the substrate might have been exported but could not be detected because of the very low molecular weight of this protein (3.5 kDa). Nilsson *et al.*, (2015) demonstrated the force profile of ADR1a expressed in *vitro* in the presence of an access of zinc. Moreover, an additional N-terminal domain was added to the substrate for visualisation. We suggest that optimization of the export assays of ADR1a is necessary to confirm whether the Tat system has the ability to export a small zinc finger binding domain protein.

In this study, we identified when Tat substrates fold in the cytoplasm and at what stage they interact with the Tat substrates. Although this increased our understanding on the folding kinetics and Tat mediated translocation of Tat substrates, it also did raise additional questions. The data presented in this study suggests a possible interaction of the *E. coli* native Tat substrate with the Tat components. *In vivo* expression assays of the SufI-RR and SufI-KK variant in Δ*tat* cells will reveal whether interaction with Tat increases pulling force acting on the SufI-RR nascent chain. In addition, *in vivo* expression assays of the SufI-RR and SufI-KK variant in 'Tat Express' cells will also reveal whether interaction with Tat components takes place. Even though force acting on the nascent chain of BT6 was detected *in vitro,* the heme *b* cofactor was not present to complete folding. In order to further investigate co-translational folding of BT6 *in vitro* expression assays are needed in the presence of the heme *b* cofactor. Furthermore, the addition of chaperones, such as GroEL and trigger factor, to the *in vitro* expression assays of BT6 and hGH will confirm

that chaperones might inhibit folding in the *in vivo* assays performed in this study. Finally, it could be useful to perform electron cryomicroscopy (Cryo-EM) studies in order to confirm whether the tested Tat substrates fold and interact with Tat in a co-translational manner. Also, cross-linking experiments could be useful to see whether there is early co-translational interaction with Tat. This study will act as a platform for further research and provides a significant starting point for studying the Tat translocase system in the future.

It has to be mentioned that the pulling force measurement technique is a relatively new technique, which needs to be optimised. This technique has only been used to study two events in the cell (cotranslational protein folding and membrane protein insertion). Therefore it has to be optimised in order to study other events such as: protein-protein interaction. Different arrest peptides, linker lengths, expression methods, etc. need to be tested in order to study different events in the cell. This is important to continue investigating a possible interaction of Tat substrates with the Tat components using the pulling force measurement technique.

Proposed pathway for the translation, folding and proofreading of Tat substrates.

The aim of this study was to investigate what exactly Tat senses in order to reject or accept a substrate for translocation to the periplasm. Based on the data demonstrated in this study we propose the pathway Tat substrates undergo from translation by the ribosome to proofreading by the Tat components. In chapter 5 we demonstrate that, some Tat substrates can already start to fold co-translationally. Moreover a Tat substrate can possibly be targeted co-translationally hence the signal peptide can be recognised when the substrate is still synthesised.

Previous studies demonstrated that the signal peptide is recognised by the TatBC complex. Following substrate binding, the TatBC complex recruits TatA components upon which proofreading possibly takes place (Rocco *et al.,* 2012) In chapter 3 and 4 we have shown that the Tat proofreading mechanism has the ability to sense the conformational flexibility of the substrate. We suggested that the Tat components could sense flexible conformations by a cascade of interactions between the substrate and Tat components. The exact mechanism on how and when the Tat components can sense the conformational state remains unclear.

Because of the possible co-translational targeting, the Tat components can rapidly proofread the conformational flexibility of the substrate in order to translocate the substrate to the periplasm. This will increase the export efficiency as the co-translationally folded substrates can be exported in a faster manner.

Chapter 7:

References

Akiyama, Y. and Ito, K. (1987) 'Topology analysis of the SecY protein, an integral membrane protein involved in protein export in Escherichia coli.', *The EMBO journal*, 6(11), pp. 3465–3470. doi: 10.1002/j.1460-2075.1987.tb02670.x.

Alami, M. *et al.* (2003) 'Differential interactions between a twin-arginine signal peptide and its translocase in Escherichia coli', *Molecular Cell*, 12(4), pp. 937–946. doi: 10.1016/S1097-2765(03)00398-8.

Alanen, H. I. *et al.* (2015) 'Efficient export of human growth hormone, interferon α2b and antibody fragments to the periplasm by the Escherichia coli Tat pathway in the absence of prior disulfide bond formation', *Biochimica et Biophysica Acta - Molecular Cell Research*. Elsevier B.V., 1853(3), pp. 756–763. doi: 10.1016/j.bbamcr.2014.12.027.

Allen, S. C. H. *et al.* (2002) 'Essential cytoplasmic domains in the Escherichia coli tatC protein', *Journal of Biological Chemistry*, 277(12), pp. 10362–10366. doi: 10.1074/jbc.M109135200.

Angelini, S., Deitermann, S. and Koch, H. G. (2005) 'FtsY, the bacterial signal-recognition particle receptor, interacts functionally and physically with the SecYEG translocon', *EMBO Reports*, 6(5), pp. 476–481. doi: 10.1038/sj.embor.7400385.

Bacher, G. *et al.* (1996) 'Regulation by the ribosome of the GTPase of the signal-recognition particle during protein targeting', *Nature*, pp. 248–251. doi: 10.1038/381248a0.

Baglieri, J. *et al.* (2012) 'Structure of TatA paralog, TatE, suggests a structurally homogeneous form of Tat protein translocase that transports folded proteins of differing diameter', *Journal of Biological Chemistry*, 287(10), pp. 7335–7344. doi: 10.1074/jbc.M111.326355.

Barnett, J. P. *et al.* (2011) 'Expression of the bifunctional Bacillus subtilis TatAd protein in Escherichia coli reveals distinct TatA/B-family and TatB-specific domains', *Archives of Microbiology*, 193(8), pp. 583–594. doi: 10.1007/s00203-011-0699-4.

Beck, D. *et al.* (2013) 'Ultrastructural characterisation of Bacillus subtilis TatA complexes suggests they are too small to form homooligomeric translocation pores', *Biochimica et Biophysica Acta - Molecular Cell Research*. Elsevier B.V., 1833(8), pp. 1811–1819. doi: 10.1016/j.bbamcr.2013.03.028.

Bedouelle, H. *et al.* (1980) 'Mutations which alter the function of the signal sequence of the maltose binding protein of Escherichia coli', *Nature*, 285(5760), pp. 78–81. doi: 10.1038/285078a0.

Beena, K., Udgaonkar, J. B. and Varadarajan, R. (2004) 'Effect of Signal Peptide on the Stability and Folding Kinetics of Maltose Binding Protein', *Biochemistry*, 43(12), pp. 3608–3619. doi: 10.1021/bi0360509.

Behrendt, J. *et al.* (2004) 'Topological studies on the twin-arginine translocase component TatC', *FEMS Microbiology Letters*, 234(2), pp. 303–308. doi: 10.1016/j.femsle.2004.03.048.

Berks, B. C. (1996) 'A common export pathway for proteins binding complex redox cofactors?', *Molecular Microbiology*, 22(3), pp. 393–404. doi: 8939424.

Berks, B. C., Lea, S. M. and Stansfeld, P. J. (2014) 'Structural biology of tat protein transport', *Current Opinion in Structural Biology*. Elsevier Ltd, 27(1), pp. 32–37. doi: 10.1016/j.sbi.2014.03.003.

Berks, B. C., Palmer, T. and Sargent, F. (2003) 'The Tat protein translocation pathway and its role in microbial physiology', *Advances in Microbial Physiology*, 47(0), pp. 187–254.

Blaudeck, N. *et al.* (2005) 'Isolation and characterization of bifunctional Escherichia coli TatA mutant proteins that allow efficient Tat-dependent protein translocation in the absence of TatB', *Journal of Biological Chemistry*, 280(5), pp. 3426–3432. doi: 10.1074/jbc.M411210200.

Blümmel, A. S. *et al.* (2015) 'Initial assembly steps of a translocase for folded proteins', *Nature Communications*, 6. doi: 10.1038/ncomms8234.

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

Bogsch, E. G. *et al.* (1998) 'An Essential Component of a System with Homologues in Plastids and Mitochondria', *The Journal of Biological Chemistry*, 273(29), pp. 18003–18006. doi: 10.1074/jbc.273.29.18003.

Bolhuis, A. *et al.* (2001) 'TatB and TatC Form a Functional and Structural Unit of the Twin-arginine Translocase from Escherichia coli', *Journal of Biological Chemistry*, 276(23), pp. 20213–20219. doi: 10.1074/jbc.M100682200.

Bowers, C. and Lau, F. (2003) 'Secretion of LamB-LacZ by the signal recognition particle pathway of Escherichia coli', *Journal of bacteriology*, 185(19), pp. 5697–5705. doi: 10.1128/JB.185.19.5697.

Briggs, M. S. *et al.* (1986) 'Conformations of Signal peptides induced by lipids suggest initial steps in protein export', *Science,* 'p. 716. 22.', 233.

Browning, D. F. *et al.* (2017) 'Escherichia coli "TatExpress" strains super-secrete human growth hormone into the bacterial periplasm by the Tat pathway', *Biotechnology and Bioengineering*, 114(12), pp. 2828–2836. doi: 10.1002/bit.26434.

Brüser, T. and Sanders, C. (2003) 'An alternative model of the twin arginine translocation system', *Microbiological Research*, 158(1), pp. 7–17. doi: 10.1078/0944-5013-00176.

Buchanan, G., Sargent, F., *et al.* (2002) 'A genetic screen for suppressors of Escherichia coli Tat signal peptide mutations establishes a critical role for the second arginine within the twin-arginine motif', *Archives of Microbiology*, 177(1), pp. 107–112. doi: 10.1007/s00203-001-0366-2.

Buchanan, G., De Leeuw, E., *et al.* (2002) 'Functional complexity of the twin-arginine translocase TatC component revealed by site-directed mutagenesis', *Molecular Microbiology*, 43(6), pp. 1457–1470. doi: 10.1046/j.1365-2958.2002.02853.x.

Butkus, M. E., Prundeanu, L. B. and Oliver, D. B. (2003) 'Translocon "Pulling" of Nascent SecM Controls the Duration of Its Translational Pause and Secretion-Responsive secA Regulation', *Journal of Bacteriology*, 185(22), pp. 6719–6722. doi: 10.1128/JB.185.22.6719-6722.2003.

Carlo Maurer, Sascha Panahandeh, Anna-Carina Jungkamp, Michael Moser, and M. M.̈ller (2010) 'TatB Functions as an Oligomeric Binding Site for Folded Tat Precursor Proteins', *Molecular Biology of the Cell*, 21(1), pp. 4151–4161. doi: 10.1091/mbc.E10.

Casadaban, M. J. and Cohen, S. N. (1980) 'Analysis of gene control signals by DNA fusion and cloning in Escherichia coli', *Journal of Molecular Biology*, 138(2), pp. 179–207. doi: 10.1016/0022-2836(80)90283-1.

Cavenagh et al (2006) *Protein NMR spectroscopy*. 2nd edn.

Chan, C. S. *et al.* (2011) 'Towards understanding the Tat translocation mechanism through structural and biophysical studies of the amphipathic region of TatA from Escherichia coli', *Biochimica et Biophysica Acta - Biomembranes*. Elsevier B.V., 1808(9), pp. 2289–2296. doi: 10.1016/j.bbamem.2011.05.024.

Chan, C. S. and Turner, R. J. (2015) 'Biogenesis of escherichia coli dmso reductase: A network of participants for protein folding and complex enzyme maturation', *Advances in Experimental Medicine and Biology*, 883, pp. 215–234. doi: 10.1007/978-3-319-23603-2_12.

Cline, K. and Mori, H. (2001) 'Thylakoid pH-dependent precursor proteins bind to a cpTatC-Hcf106 complex before Tha4-dependent transport', *Journal of Cell Biology*, 154(4), pp. 719–729. doi: 10.1083/jcb.200105149.

Connolly, T., Rapiejko, P. and Gilmore, R. (1991) 'Requirement of GTP hydrolysis for dissociation of the signal recognition particle from its receptor', *Science*, 252(5009), pp. 1171–1173. doi: 10.1126/science.252.5009.1171.

Cristóbal, S. *et al.* (1999) 'Competition between Sec- and TAT-dependent protein translocation in Escherichia coli', *EMBO Journal*, 18(11), pp. 2982–2990. doi: 10.1093/emboj/18.11.2982.

Cunningham, K. and Wickner, W. (1989) 'Specific recognition of the leader region of precursor proteins is required for the activation of translocation ATPase of Escherichia coli.', *Proceedings of the National Academy of Sciences of the United States of America*, 86(November), pp. 8630–8634. doi: 10.1073/pnas.86.22.8630.

Cymer, F. and von Heijne, G. (2013) 'Cotranslational folding of membrane proteins probed by arrest-peptide–mediated force measurements', *Proceedings of the National Academy of Sciences*, 110(36), pp. 14640–14645. doi: 10.1073/pnas.1306787110.

Dalbey, R. E. and Kuhn, A. (2004) 'YidC family members are involved in the membrane insertion, lateral integration, folding, and assembly of membrane proteins', *Journal of Cell Biology*, 166(6), pp. 769–774. doi: 10.1083/jcb.200405161.

Dalbeys, R. E. and Wickner, W. (1986) 'The Role of the Polar, Carboxyl-terminal Domain of', 261(29), pp. 13844–13849.

Degrado, W. F. *et al.* (1989) 'Protein Design, a Minimalist Approach', 243(4891), pp. 622–628.

DeLisa, M. P. *et al.* (2002) 'Genetic analysis of the twin arginine translocator secretion pathway in bacteria.', *The Journal of biological chemistry*, 277(33), pp. 29825–29831. doi: 10.1074/jbc.M201956200.

DeLisa, M. P. *et al.* (2004) 'Phage Shock Protein PspA of Escherichia coli Relieves Saturation of Protein Export via the Tat Pathway', *Journal of Bacteriology*, pp. 366–373. doi: 10.1128/JB.186.2.366-373.2004.

DeLisa, M. P., Tullman, D. and Georgiou, G. (2003) 'Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway', *Proceedings of the National Academy of Sciences*, 100(10), pp. 6115–6120. doi: 10.1073/pnas.0937838100.

Dow, J. M. *et al.* (2013) 'Characterization of a pre-export enzyme-chaperone complex on the twin-arginine transport pathway', *Biochem J*, 452(1), pp. 57–66. doi: 10.1042/BJ20121832.

Dow, J. M. *et al.* (2014) 'Characterization of a periplasmic nitrate reductase in complex with its biosynthetic chaperone', *FEBS Journal*, 281(1), pp. 246–260. doi: 10.1111/febs.12592.

Driessen, a J. and Wickner, W. (1991) 'Proton transfer is rate-limiting for translocation of precursor proteins by the Escherichia coli translocase.', *Proceedings of the National Academy of Sciences of the United States of America*, 88(March), pp. 2471–2475. doi: 10.1073/pnas.88.6.2471.

Driessen, A. J. M. (2001) 'SecB, a molecular chaperone with two faces', *Trends in Microbiology*, 9(5), pp. 193–196. doi: 10.1016/S0966-842X(01)01980-1.

Driessen, A. J. M. and Nouwen, N. (2008) 'Protein Translocation Across the Bacterial Cytoplasmic Membrane', *Annual Review of Biochemistry*, 77(1), pp. 643–667. doi: 10.1146/annurev.biochem.77.061606.160747.

Du, W. *et al.* (2003) '1H NMR study of the molecular structure and magnetic properties of the active site for the cyanomet complex of O2-avid hemoglobin from the trematode Paramphistomum epiclitum', *European Journal of Biochemistry*, 270(13), pp. 2707–2720. doi: 10.1046/j.1432-1033.2003.03638.x.

Dubini, A. and Sargent, F. (2003) 'Assembly of Tat-dependent [NiFe] hydrogenases: Identification of precursor-binding accessory proteins', *FEBS Letters*, 549(1–3), pp. 141–146. doi: 10.1016/S0014-5793(03)00802-0.

Duong, F. and Wickner, W. (1997) 'Distinct catalytic roles of the SecYE, SecG and SecDFyajC subunits of preprotein translocase holoenzyme', *EMBO Journal*, 16(10), pp. 2756–2768. doi: 10.1093/emboj/16.10.2756.

Economou, A. (2010a) *Protein Secretion - Methods and Protocols*. doi: 10.1007/978-1-60327-412-8.

Economou, A. (2010b) 'Signal peptides are allosteric activators of the protein translocase', 462(7271), pp. 363–367. doi: 10.1038/nature08559.Signal.

Edelman, G. M. and Gally, J. A. (2001) 'Degeneracy and complexity in biological systems', *Proceedings of the National Academy of Sciences*, 98(24), pp. 13763–13768. doi: 10.1073/pnas.231499798.

Eimer, E. *et al.* (2015) 'TatE as a regular constituent of bacterial twin-arginine protein translocases', *Journal of Biological Chemistry*, 290(49), pp. 29281–29289. doi: 10.1074/jbc.M115.696005.

Eimer, E. *et al.* (2018) 'Unanticipated functional diversity among the TatA-type components of the Tat protein translocase', *Scientific Reports*, 8(1), pp. 1–12. doi: 10.1038/s41598-018-19640-3.

Farid, T. A. *et al.* (2014) 'NIH Public Access', 9(12), pp. 826–833. doi: 10.1038/nchembio.1362.Elementary.

Fincher, V., McCaffery, M. and Cline, K. (1998) 'Evidence for a loop mechanism of protein transport by the thylakoid Delta pH pathway', *FEBS Letters*, 423(1), pp. 66–70. doi: 10.1016/S0014-5793(98)00066-0.

Frauenfeld, J. *et al.* (2012) 'Cryo-EM structure of the ribosome-SecYE complex in the Membrane Environment', 18(5), pp. 614–621. doi: 10.1038/nsmb.2026.Cryo.

Fröbel, J. *et al.* (2012) 'Transmembrane insertion of twin-arginine signal peptides is driven by TatC and regulated by TatB', *Nature Communications*, 3, pp. 1–10. doi: 10.1038/ncomms2308.

Frobel, J., Rose, P. and Muller, M. (2012) 'Twin-arginine-dependent translocation of folded proteins', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1599), pp. 2246–2246. doi: 10.1098/rstb.2012.0227.

G. Blobel, B. D. (1975) 'transfer of proteins across membranes', 67, pp. 835–851.

Garza-Sánchez, F., Janssen, B. D. and Hayes, C. S. (2006) 'Prolyl-tRNAPro in the A-site of SecM-arrested ribosomes inhibits the recruitment of transfer-messenger RNA', *Journal of Biological Chemistry*, 281(45), pp. 34258–34268. doi: 10.1074/jbc.M608052200.

Gibson, D. G. *et al.* (2009) 'Enzymatic assembly of DNA molecules up to several hundred kilobases', *Nature Methods*, 6(5), pp. 343–345. doi: 10.1038/nmeth.1318.

Gohlke, U. *et al.* (2005) 'The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter.', *Proceedings of the National Academy of Sciences of the United States of America*, 102(30), pp. 10482–6. doi: 10.1073/pnas.0503558102.

Goosens, V. J. *et al.* (2013) 'Novel twin-arginine translocation pathway-dependent phenotypes of bacillus subtilis unveiled by quantitative proteomics', *Journal of Proteome Research*, 12(2), pp. 796–807. doi: 10.1021/pr300866f.

Goosens, V. J., Monteferrante, C. G. and Van Dijl, J. M. (2014) 'The Tat system of Gram-positive bacteria', *Biochimica et Biophysica Acta - Molecular Cell Research*. Elsevier B.V., 1843(8), pp. 1698–1706. doi: 10.1016/j.bbamcr.2013.10.008.

Gouffi, K. *et al.* (2004) 'Dual Topology of the Escherichia coli TatA Protein', *Journal of Biological Chemistry*, 279(12), pp. 11608–11615. doi: 10.1074/jbc.M313187200.

Grahl, S. *et al.* (2012) 'Overlapping transport and chaperone-binding functions within a bacterial twin-arginine signal peptide', *Molecular Microbiology*, 83(6), pp. 1254–1267. doi: 10.1111/j.1365-2958.2012.08005.x.

Grosset, A. M. *et al.* (2001) '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(18), pp. 5474–5487. doi: 10.1021/bi002504f.

Guymer, D. *et al.* (2010) 'Intrinsic GTPase activity of a bacterial twin-arginine translocation proofreading chaperone induced by domain swapping', *FEBS Journal*, 277(2), pp. 511–525. doi: 10.1111/j.1742-4658.2009.07507.x.

Hanada, M., Nishiyama, K. and Tokuda, H. (1996) 'SecG plays a critical role in protein translocation in the absence of the proton motive force as well as at low temperature.', *FEBS letters*, 381(1–2), pp. 25–8. doi: 10.1016/0014-5793(96)00066-X.

Hayashi, K. *et al.* (2006) 'Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110', *Molecular Systems Biology*, 2. doi: 10.1038/msb4100049.

Hedman, R. *et al.* (2015) 'Forces on nascent polypeptides during membrane insertion and translocation via the Sec translocon', pp. 1–21.

von Heijne, G. (1983) ‘Amino Acids near Signal-Sequence Cleavage Sites’, *Amino Acids*, 21, pp. 17–21.

von Heijne, G. (1985) ‘Signal sequences. The limits of variation’, *Journal of Molecular Biology*, 184(1), pp. 99–105. doi: 10.1016/0022-2836(85)90046-4.

von Heijne, G. (1987) ‘Membrane Biology 9’, *Glass*, 29, pp. 21–29.

Hessa, T. *et al.* (2005) ‘Recognition of transmembrane helices by the endoplsmic reticulum translocon’, *Nature*, 433(7024), pp. 377–381. doi: 10.1038/nature03216.

Hessa, T. *et al.* (2007) ‘Molecular code for transmembrane-helix recognition by the Sec61 translocon’, *Nature*, 450(7172), pp. 1026–1030. doi: 10.1038/nature06387.

Hicks, M. G. *et al.* (2003) ‘The Escherichia coli twin-arginine translocase: Conserved residues of TatA and TatB family components involved in protein transport’, *FEBS Letters*, 539(1–3), pp. 61–67. doi: 10.1016/S0014-5793(03)00198-4.

Hinsley, A. P. *et al.* (2001) ‘A naturally occurring bacterial Tat signal peptide lacking one of the “invariant” arginine residues of the consensus targeting motif’, *FEBS Letters*, 497(1), pp. 45–49. doi: 10.1016/S0014-5793(01)02428-0.

Houben, E. N. G. *et al.* (2005) ‘Early encounters of a nascent membrane protein: Specificity and timing of contacts inside and outside the ribosome’, *Journal of Cell Biology*, 170(1), pp. 27–35. doi: 10.1083/jcb.200503035.

Hwang, T. L. and Shaka, A. J. (1995) ‘Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients’, *Journal of Magnetic Resonance - Series A*, pp. 275–279. doi: 10.1006/jmra.1995.1047.

Ismail, N. *et al.* (2013) ‘A bi-phasic pulling force acts on transmembrane helices during translocon-mediated membrane integration’, 19(10), pp. 1018–1022. doi: 10.1038/nsmb.2376.A.

Ito, K., Chiba, S. and Pogliano, K. (2010) 'Divergent stalling sequences sense and control cellular physiology', *Biochemical and Biophysical Research Communications*. Elsevier Inc., 393(1), pp. 1–5. doi: 10.1016/j.bbrc.2010.01.073.

Ize, B. *et al.* (2002) 'In vivo dissection of the Tat translocation pathway in Escherichia coli', *Journal of Molecular Biology*, 317(3), pp. 327–335. doi: 10.1006/jmbi.2002.5431.

Jack, R. L. *et al.* (2004) 'Coordinating assembly and export of complex bacterial proteins', *EMBO Journal*, 23(20), pp. 3962–3972. doi: 10.1038/sj.emboj.7600409.

Jepson, B. J. N. *et al.* (2007) 'Spectropotentiometric and Structural Analysis of the Periplasmic Nitrate Reductase from *Escherichia coli*', *Journal of Biological Chemistry*, 282(9), pp. 6425–6437. doi: 10.1074/jbc.M607353200.

Jones, A. S. *et al.* (2016) 'Proofreading of substrate structure by the Twin-Arginine Translocase is highly dependent on substrate conformational flexibility but surprisingly tolerant of surface charge and hydrophobicity changes', *Biochimica et Biophysica Acta - Molecular Cell Research*. Elsevier B.V., 1863(12), pp. 3116–3124. doi: 10.1016/j.bbamcr.2016.09.006.

Jongbloed, J. D. H. *et al.* (2000) 'TatC is a specificity determinant for protein secretion via the twin-arginine translocation pathway', *Journal of Biological Chemistry*, 275(52), pp. 41350–41357. doi: 10.1074/jbc.M004887200.

Jongbloed, J. D. H. *et al.* (2004) 'Two minimal Tat translocases in Bacillus', *Molecular Microbiology*, 54(5), pp. 1319–1325. doi: 10.1111/j.1365-2958.2004.04341.x.

Kelley L.A, S. Mezulis, C.M. Yates, M.N. Wass, M.J.E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis, Nat. Protoc. 10 (2015) 845–858. [28]

Kneuper, H. *et al.* (2012) 'Molecular dissection of TatC defines critical regions essential for protein transport and a TatB-TatC contact site', *Molecular Microbiology*, 85(5), pp. 945–961. doi: 10.1111/j.1365-2958.2012.08151.x.

Kudva, R. *et al.* (2013) 'Protein translocation across the inner membrane of Gram-negative bacteria: The Sec and Tat dependent protein transport pathways', *Research in Microbiology*, 164(6), pp. 505–534. doi: 10.1016/j.resmic.2013.03.016.

Van Der Laan, M., Nouwen, N. P. and Driessen, A. J. M. (2005) 'YidC - An evolutionary conserved device for the assembly of energy-transducing membrane protein complexes', *Current Opinion in Microbiology*, 8(2), pp. 182–187. doi: 10.1016/j.mib.2005.02.004.

Lee, H. C. and Bernstein, H. D. (2001) 'The targeting pathway of Escherichia coli presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal', *Proceedings of the National Academy of Sciences*, 98(6), pp. 3471–3476. doi: 10.1073/pnas.051484198.

Lee, P. A. *et al.* (2002) 'Truncation analysis of TatA and TatB defines the minimal functional units required for protein translocation', *Journal of Bacteriology*, 184(21), pp. 5871–5879. doi: 10.1128/JB.184.21.5871-5879.2002.

Li, H. *et al.* (2010) 'DmsD, a Tat system specific chaperone, interacts with other general chaperones and proteins involved in the molybdenum cofactor biosynthesis', *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1804(6), pp. 1301–1309. doi: 10.1016/j.bbapap.2010.01.022.

Li, S. Y., Chang, B. Y. and Lin, S. C. (2006) 'Coexpression of TorD enhances the transport of GFP via the TAT pathway', *Journal of Biotechnology*, 122(4), pp. 412–421. doi: 10.1016/j.jbiotec.2005.09.011.

Lichtenstein, B. R. *et al.* (2013) 'Engineering oxidoreductases: maquette proteins designed from Scratch', *Biochem. Soc. Trans.*, 40(3), pp. 561–566. doi: 10.1042/BST20120067.Engineering.

Lill, R., Dowhan, W. and Wickner, W. (1990) 'The ATPase activity of secA is regulated by acidic phospholipids, secY, and the leader and mature domains of precursor proteins', *Cell*, 60(2), pp. 271–280. doi: 10.1016/0092-8674(90)90742-W.

Luirink, J. *et al.* (2004) 'An Alternative protein targeting pathway in Escherichia coli: studies on the role of FtsY', *Applied Health Economics & Health Policy*, 3(1), pp. 29–33. doi: 10.1002/j.1460-2075.1994.tb06511.x.

Luirink, J. *et al.* (2005) 'BIOGENESIS OF INNER MEMBRANE PROTEINS IN *ESCHERICHIA COLI*', *Annual Review of Microbiology*, 59(1), pp. 329–355. doi: 10.1146/annurev.micro.59.030804.121246.

Luirink, J. and Sinning, I. (2004) 'SRP-mediated protein targeting: Structure and function revisited', *Biochimica et Biophysica Acta - Molecular Cell Research*, 1694(1–3 SPEC.ISS.), pp. 17–35. doi: 10.1016/j.bbamcr.2004.03.013.

Magalon, A. and Böck, A. (2000) 'Dissection of the maturation reactions of the [NiFe] hydrogenase 3 from Escherichia coli taking place after nickel incorporation', *FEBS Letters*, 473(2), pp. 254–258. doi: 10.1016/S0014-5793(00)01542-8.

Marion, D., Ikura, M. and Bax, A. (1989) 'Improved solvent suppression in one- and two-dimensional NMR spectra by convolution of time-domain data', *Journal of Magnetic Resonance (1969)*, 84(2), pp. 425–430. doi: 10.1016/0022-2364(89)90391-0.

Matos, C. F. R. O. *et al.* (2014) 'Efficient export of prefolded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in Escherichia coli CyDisCo strains', *Biotechnology Progress*, 30(2), pp. 281–290. doi: 10.1002/btpr.1858.

Matos, C. F. R. O., Robinson, C. and Di Cola, A. (2008) 'The Tat system proofreads FeS protein substrates and directly initiates the disposal of rejected molecules', *EMBO Journal*, 27(15), pp. 2055–2063. doi: 10.1038/emboj.2008.132.

Matsumoto, G., Yoshihisa, T. and Ito, K. (1997) 'SecY and SecA interact to allow SecA insertion and protein translocation across the Escherichia coli plasma membrane', *EMBO Journal*, 16(21), pp. 6384–6393. doi: 10.1093/emboj/16.21.6384.

Maurer, C. *et al.* (2009) 'Impairment of twin-arginine-dependent export by seemingly small alterations of substrate conformation', *FEBS Letters*. Federation of European Biochemical Societies, 583(17), pp. 2849–2853. doi: 10.1016/j.febslet.2009.07.038.

Mehner, D. *et al.* (2012) 'The Tat system for membrane translocation of folded proteins recruits the membrane-stabilizing Psp machinery in Escherichia coli', *Journal of Biological Chemistry*, 287(33), pp. 27834–27842. doi: 10.1074/jbc.M112.374983.

Michiel Kleerebezem, W. C. and J. T. (1996) 'Involvement of stress protein PspA (phage shock protein A) of Escherichia coli in maintenance of the protonmotive force under stress conditions', 140(3562), pp. 24–26. doi: 10.1128/IAI.01900-14.

Monteferrante, C. G. *et al.* (2012) 'TatAc, the third TatA subunit of Bacillus subtilis, can form active twin-arginine translocases with the TatCd and TatCy subunits', *Applied and Environmental Microbiology*, pp. 4999–5001. doi: 10.1128/AEM.01108-12.

Monteferrante, C. G. *et al.* (2013) 'Mapping the twin-arginine protein translocation network of Bacillus subtilis', *Proteomics*, 13(5), pp. 800–811. doi: 10.1002/pmic.201200416.

Mori, H. and Cline, K. (2002) 'A twin arginine signal peptide and the pH gradient trigger reversible assembly of the thylakoid ??pH/Tat translocase', *Journal of Cell Biology*, 157(2), pp. 205–210. doi: 10.1083/jcb.200202048.

Mori, H. and Ito, K. (2001) 'An essential amino acid residue in the protein translocation channel revealed by targeted random mutagenesis of SecY.', *Proceedings of the National Academy of Sciences of the United States of America*, 98(9), pp. 5128–5133. doi: 10.1073/pnas.081617398.

Muñiz, M., Morsomme, P. and Riezman, H. (2001) 'Protein sorting upon exit from the endoplasmic reticulum', *Cell*, 104(2), pp. 313–320. doi: 10.1016/S0092-8674(01)00215-X.

Muto, H., Nakatogawa, H. and Ito, K. (2006) 'Genetically Encoded but Nonpolypeptide Prolyl-tRNA Functions in the A Site for SecM-Mediated Ribosomal Stall', *Molecular Cell*, 22(4), pp. 545–552. doi: 10.1016/j.molcel.2006.03.033.

Nakatogawa, H. and Ito, K. (2001) 'Secretion monitor, secM, undergoes self-translation arrest in the cytosol', *Molecular Cell*, 7(1), pp. 185–192. doi: 10.1016/S1097-2765(01)00166-6.

Nakatogawa, H. and Ito, K. (2002) 'The ribosomal exit tunnel functions as a discriminating gate', *Cell*, 108(5), pp. 629–636. doi: 10.1016/S0092-8674(02)00649-9.

Nakatogawa, H., Murakami, A. and Ito, K. (2004) 'Control of SecA and SecM translation by protein secretion', *Current Opinion in Microbiology*, 7(2), pp. 145–150. doi: 10.1016/j.mib.2004.01.001.

Natale, P., Brüser, T. and Driessen, A. J. M. (2008) 'Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane-Distinct translocases and mechanisms', *Biochimica et Biophysica Acta - Biomembranes*, 1778(9), pp. 1735–1756. doi: 10.1016/j.bbamem.2007.07.015.

Nesmeyanova, M. A. *et al.* (1997) 'Positively charged lysine at the N-terminus of the signal peptide of the Escherichia coli alkaline phosphatase provides the secretion efficiency and is involved in the interaction with anionic phospholipids', *FEBS Letters*. Federation of European Biochemical Societies, 403(2), pp. 203–207. doi: 10.1016/S0014-5793(97)00052-5.

Nilavongse, A. *et al.* (2006) 'The NapF protein of the Escherichia coli periplasmic nitrate reductase system: Demonstration of a cytoplasmic location and interaction with the catalytic subunit, NapA', *Microbiology*, 152(11), pp. 3227–3237. doi: 10.1099/mic.0.29157-0.

Nilsson, O. B. *et al.* (2015) 'Cotranslational Protein Folding inside the Ribosome Exit Tunnel', *Cell Reports*, 12(10), pp. 1533–1540. doi: 10.1016/j.celrep.2015.07.065.

Nilsson, O. B. *et al.* (2016) 'Trigger Factor Reduces the Force Exerted on the Nascent Chain by a Cotranslationally Folding Protein', *Journal of Molecular Biology*. Elsevier Ltd, 428(6), pp. 1356–1364. doi: 10.1016/j.jmb.2016.02.014.

Nilsson, O. B. *et al.* (2017) 'Cotranslational folding of spectrin domains via partially structured states', *Nature Structural and Molecular Biology*. Nature Publishing Group, 24(3), pp. 221–225. doi: 10.1038/nsmb.3355.

Nishiyama, K., Hanada, M. and Tokuda, H. (1994) '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(14), pp. 3272–7. doi: 10.1002/j.1460-2075.1994.tb06628.x.

Nishiyama, K. I., Suzuki, T. and Tokuda, H. (1996) 'Inversion of the membrane topology of SecG coupled with SecA-dependent preprotein translocation', *Cell*, 85(1), pp. 71–81. doi: 10.1016/S0092-8674(00)81083-1.

Oresnik, I. J., Ladner, C. L. and Turner, R. J. (2001) 'Identification of a twin-arginine leader-binding protein', *Molecular Microbiology*, 40(2), pp. 323–331. doi: 10.1046/j.1365-2958.2001.02391.x.

Overton, T. W. (2014) 'Recombinant protein production in bacterial hosts', *Drug Discovery Today*. Elsevier Ltd, 19(5), pp. 590–601. doi: 10.1016/j.drudis.2013.11.008.

Paetzel, M. *et al.* (2002) 'Signal peptidases', *Chemical Reviews*, 102(12), pp. 4549–4579. doi: 10.1021/cr010166y.

Palmer, T. and Berks, B. C. (2001) 'Constitutive Expression of Escherichia coli tat Genes Indicates an Important Role for the Twin-Arginine Translocase during Aerobic and Anearobic Growth', *Society*, 183(5), pp. 1801–1804. doi: 10.1128/JB.183.5.1801.

Palmer, T. and Berks, B. C. (2012) 'The twin-arginine translocation (Tat) protein export pathway', *Nature Reviews Microbiology*. Nature Publishing Group, 10(7), pp. 483–496. doi: 10.1038/nrmicro2814.

Papanikolau, Y. *et al.* (2007) 'Structure of Dimeric SecA, the Escherichia coli Preprotein Translocase Motor', *Journal of Molecular Biology*, 366(5), pp. 1545–1557. doi: 10.1016/j.jmb.2006.12.049.

Patel, R., Smith, S. M. and Robinson, C. (2014) 'Protein transport by the bacterial Tat pathway', *Biochimica et Biophysica Acta - Molecular Cell Research*. Elsevier B.V., 1843(8), pp. 1620–1628. doi: 10.1016/j.bbamcr.2014.02.013.

Peterson, J. H., Woolhead, C. A. and Bernstein, H. D. (2003) 'Basic Amino Acids in a Distinct Subset of Signal Peptides Promote Interaction with the Signal Recognition Particle', *Journal of Biological Chemistry*, 278(46), pp. 46155–46162. doi: 10.1074/jbc.M309082200.

Pohlschröder, M., Murphy, C. and Beckwith, J. (1996) 'In vivo analyses of interactions between SecE and SecY, core components of the Escherichia coli protein translocation machinery', *Journal of Biological Chemistry*, 271(33), pp. 19908–19914. doi: 10.1074/jbc.271.33.19908.

Pommier, J., Me, V. and Iobbi-nivol, C. (1998) 'TorD, A Cytoplasmic Chaperone That Interacts with the Unfolded Trimethylamine', *Molecular Biology*, 273(26), pp. 16615–16620.

Powers, T. and Walter, P. (1995) 'Reciprocal stimulation of GTP hydrolysis by two directly interacting GTpases', *Science*. doi: 10.1126/science.7660124.

Ramasamy, S. *et al.* (2013) 'The glove-like structure of the conserved membrane protein tatc provides insight into signal sequence recognition in twin-arginine translocation', *Structure*. Elsevier Ltd, 21(5), pp. 777–788. doi: 10.1016/j.str.2013.03.004.

Randall, L. L. and Hardy, S. J. S. (2002) 'SecB, one small chaperone in the complex milieu of the cell', *Cellular and Molecular Life Sciences*, 59(10), pp. 1617–1623. doi: 10.1007/PL00012488.

Rapoport, T. A. *et al.* (2004) 'Membrane-protein integration and the role of the translocation channel', *Trends in Cell Biology*, 14(10), pp. 568–575. doi: 10.1016/j.tcb.2004.09.002.

Rapoport, T. A. (2007) 'Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes', *Nature*, 450(7170), pp. 663–669. doi: 10.1038/nature06384.

Ray, N. *et al.* (2003) 'DmsD is required for the biogenesis of DMSO reductase in Escherichia coli but not for the interaction of the DmsA signal peptide with the Tat apparatus', *FEBS Letters*, 534(1–3), pp. 156–160. doi: 10.1016/S0014-5793(02)03839-5.

Richter, S. *et al.* (2007) 'Functional tat transport of unstructured, small, hydrophilic proteins', *Journal of Biological Chemistry*, 282(46), pp. 33257–33264. doi: 10.1074/jbc.M703303200.

Robinson, C. *et al.* (2011) 'Transport and proofreading of proteins by the twin-arginine translocation (Tat) system in bacteria', *Biochimica et Biophysica Acta - Biomembranes*. Elsevier B.V., 1808(3), pp. 876–884. doi: 10.1016/j.bbamem.2010.11.023.

Rocco, M. a., Waraho-Zhmayev, D. and DeLisa, M. P. (2012) 'Twin-arginine translocase mutations that suppress folding quality control and permit export of misfolded substrate proteins', *Proceedings of the National Academy of Sciences*, 109(33), pp. 13392–13397. doi: 10.1073/pnas.1210140109.

Rodrigue, A. *et al.* (1999) 'Co-translocation of a Periplasmic Enzyme Complex by a Hitchhiker Mechanism through the Bacterial Tat Pathway Co-translocation of a Periplasmic Enzyme Complex by a Hitchhiker Mechanism through the Bacterial Tat Pathway *', 274(19), pp. 13223–13228. doi: 10.1074/jbc.274.19.13223.

Rodriguez, F. *et al.* (2013) 'Structural model for the protein-translocating element of the twin-arginine transport system', *Proceedings of the National Academy of Sciences*, 110(12), pp. E1092–E1101. doi: 10.1073/pnas.1219486110.

Rollauer, S. E. *et al.* (2013) 'Structure of the TatC core of the twin-arginine protein transport system Europe PMC Funders Author Manuscripts Structure of TatC We targeted TatC proteins from thermophilic bacteria and archaea for structure', 492(7428), pp. 210–214. doi: 10.1038/nature11683.Structure.

Rose, P. *et al.* (2013) 'Substrate-Dependent Assembly of the Tat Translocase as Observed in Live Escherichia coli Cells', *PLoS ONE*, 8(8), pp. 1–17. doi: 10.1371/journal.pone.0069488.

Rose, R. W. *et al.* (2002) 'Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway', *Molecular Microbiology*, 45(4), pp. 943–950. doi: 10.1046/j.1365-2958.2002.03090.x.

Sargent, F. *et al.* (1998) 'Overlapping functions of components of a bacterial Sec-independent protein export pathway', *EMBO Journal*, 17(13), pp. 3640–3650. doi: 10.1093/emboj/17.13.3640.

Sargent, F. *et al.* (1999) 'Sec-independent Protein Translocation in Escherichia coli arginine " motif . The Tat system involves the integral', 274(51), pp. 36073–36082.

Schatz, P. J. *et al.* (1989) 'The secE gene encodes an integral membrane protein required for protein export in Escherichia coli', *Genes Dev.*, 3(7), pp. 1035–1044. doi: 10.1101/gad.3.7.1035.

Schatz, P. J. *et al.* (1991) '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(7), pp. 1749–57. doi: 10.1002/j.1460-2075.1991.tb07699.x.

Schiebel, E. *et al.* (1991) 'ΔμH+ and ATP function at different steps of the catalytic cycle of preprotein translocase', *Cell*, 64(5), pp. 927–939. doi: 10.1016/0092-8674(91)90317-R.

Schierle, C., Boyd, D. and Beckwith, J. (2003) 'The DsbA Signal Sequence Directs Ef cient, Cotranslational Export of Passenger Proteins to the', *Society*, 185(19), pp. 5706–5713. doi: 10.1128/JB.185.19.5706.

Schlegel, S. *et al.* (2014) 'Bacterial-based membrane protein production', *Biochimica et Biophysica Acta - Molecular Cell Research*. Elsevier B.V., 1843(8), pp. 1739–1749. doi: 10.1016/j.bbamcr.2013.10.023.

Shifman, J. M. *et al.* (1998) 'Functionalized de Novo designed proteins: Mechanism of proton coupling to oxidation/reduction in heme protein maquettes', *Biochemistry*, 37(47), pp. 16815–16827. doi: 10.1021/bi9816857.

Sijbrandi, R. *et al.* (2003) 'Signal recognition particle (SRP)-mediated targeting and Sec-dependent translocation of an extracellular Escherichia coli protein', *Journal of Biological Chemistry*, 278(7), pp. 4654–4659. doi: 10.1074/jbc.M211630200.

Stanley, N. R., Palmer, T. and Berks, B. C. (2000) 'The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in Escherichia coli', *Journal of Biological Chemistry*, 275(16), pp. 11591–11596. doi: 10.1074/jbc.275.16.11591.

Sutherland, G. *et al.* (2018) 'Probing the quality control mechanism of the Escherichia coli twin-arginine translocase using folding variants of a de novo-designed heme protein', *Jbc*, (1). doi: 10.1074/jbc.RA117.000880.

Torsten H. Walther, Stephan L. Grage, Nadine Roth, and A. S. U. (2010) 'Membrane Alignment of the Pore-Forming Component TatAd of the Twin-Arginine Translocase from Bacillus subtilis Resolved by Solid-State NMR Spectroscopy', *Journal of the American Chemical Society*, 132(45), pp. 15942–15944. doi: 10.1021/ja1053785.

Tottey, S. *et al.* (2008) 'Protein-folding location can regulate manganese-binding versus copper- or zinc-binding', *Nature*, 455(7216), pp. 1138–1142. doi: 10.1038/nature07340.

Tsirigotaki, A. *et al.* (2017) 'Protein export through the bacterial Sec pathway', *Nature Reviews Microbiology*. Nature Publishing Group, 15(1), pp. 21–36. doi: 10.1038/nrmicro.2016.161.

Tullman-Ercek, D. *et al.* (2007) 'Export pathway selectivity of Escherichia coli twin arginine translocation signal peptides', *Journal of Biological Chemistry*, 282(11), pp. 8309–8316. doi: 10.1074/jbc.M610507200.

Ulfig, A. and Freudl, R. (2018) 'The early mature part of bacterial twin-arginine translocation (Tat) precursor proteins contributes to TatBC receptor binding', *Journal of Biological Chemistry*, (10), p. jbc.RA118.002576. doi: 10.1074/jbc.RA118.002576.

Veenendaal, A. K. J., Van Der Does, C. and Driessen, A. J. M. (2004) 'The protein-conducting channel SecYEG', *Biochimica et Biophysica Acta - Molecular Cell Research*, 1694(1–3 SPEC.ISS.), pp. 81–95. doi: 10.1016/j.bbamcr.2004.02.009.

Walker, K. L., Jones, A. S. and Robinson, C. (2015) 'The Tat pathway as a biotechnological tool for the expression and export of heterologous proteins in *Escherichia coli*', *Pharmaceutical Bioprocessing*, 3(6), pp. 387–396. doi: 10.4155/pbp.15.21.

Walther, T. H. *et al.* (2013) 'Folding and self-assembly of the TatA translocation pore based on a charge zipper mechanism', *Cell*. Elsevier, 152(1–2), pp. 316–326. doi: 10.1016/j.cell.2012.12.017.

Wexler, M. *et al.* (2000) 'TatD is a cytoplasmic protein with DNase activity. No requirement for TatD family proteins in Sec-Independent protein export', *Journal of Biological Chemistry*, 275(22), pp. 16717–16722. doi: 10.1074/jbc.M000800200.

White, G. F. *et al.* (2010) 'Subunit organization in the TatA complex of the twin arginine protein translocase: A site-directed EPR spin labeling study', *Journal of Biological Chemistry*, 285(4), pp. 2294–2301. doi: 10.1074/jbc.M109.065458.

White, S. H. and von Heijne, G. (2008) 'How Translocons Select Transmembrane Helices', *Annual Review of Biophysics*, 37(1), pp. 23–42. doi: 10.1146/annurev.biophys.37.032807.125904.

Wickner, W. (1994) 'How ATP Drives Proteins Across Membranes', 266(November), pp. 1197–1199.

Winstone, T. M. L., Tran, V. A. and Turner, R. J. (2013) 'The hydrophobic region of the DmsA twin-arginine leader peptide determines specificity with chaperone DMsD', *Biochemistry*, 52(43), pp. 7532–7541. doi: 10.1021/bi4009374.

Van Der Wolk, J. P. W., De Wit, J. G. and Driessen, A. J. M. (1997) 'The catalytic cycle of the Escherichia coli SecA ATPase comprises two distinct preprotein translocation events', *EMBO Journal*, 16(24), pp. 7297–7304. doi: 10.1093/emboj/16.24.7297.

Xu, Z., Knafels, J. D. and Yoshino, K. (2000) 'Crystal structure of the bacterial protein export chaperone SecB', *Nature Structural Biology*, 7(12), pp. 1172–1177. doi: 10.1038/82040.

Yen, M. R. *et al.* (2001) 'Phylogenetic and structural analyses of the oxal family of protein translocases', *FEMS Microbiology Letters*, 204(2), pp. 223–231.

Yen, M. R. *et al.* (2002) 'Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system', *Archives of Microbiology*, 177(6), pp. 441–450. doi: 10.1007/s00203-002-0408-4.

Zhang, G. and Ignatova, Z. (2011) 'Folding at the birth of the nascent chain: Coordinating translation with co-translational folding', *Current Opinion in Structural Biology*. Elsevier Ltd, 21(1), pp. 25–31. doi: 10.1016/j.sbi.2010.10.008.

Zoufaly, S. *et al.* (2012) 'Mapping precursor-binding site on TatC subunit of twin arginine-specific protein translocase by site-specific photo cross-linking', *Journal of Biological Chemistry*, 287(16), pp. 13430–13441. doi: 10.1074/jbc.M112.343798.

Appendix

Constructs used for BT6 studies

Plasmid name	Description	Reference
pKRK7	pEXT22 TorA-hGH-His_6	(Alanen *et al.*, 2015)
pAJ21	pEXT22 TorA-BT6	(Sutherland *et al.,* 2018)
pGS01	pEXT22 TorA-BT6M1 (H53A)	(Sutherland *et al.,* 2018)
pAJ25	pEXT22 TorA-BT6M0 (H53A, H88A)	(Sutherland *et al.,* 2018)
pAJ26	pEXT22 TorA-KK-BT6	(Sutherland *et al.*, 2018)
pAJ27	pEXT22 TorA-KR-BT6	(Sutherland *et al.*, 2018)
pDM12	pEXT22 TorA-BT6-2Asp (K5D, K13D)	This study
pDM13	pEXT22 TorA-BT6-4Asp (K5D, K13D, K75D, K83D)	This study
pDM14	pEXT22 TorA-BT6-2Lys (D9K, D23K)	This study
pDM15	pEXT22 TorA-BT6-4Lys (D9K, D23K, D79K, D93K)	This study
pDM19	pEXT22 TorA-BT6-3NLeu (D9L, K13L, E16L)	This study
pDM18	pEXT22 TorA-BT6-C-term26 (Additional–SNVIIIIITNKDPNSSSVDLAAALE)	This study
pDM16	pEXT22 TorA-BT6-C-term13 (Additional – SNVIIIIITNKDP)	This study
pDM17	pEXT22 TorA-BT6-C-term6 (Additional – SNVIII)	This study

Constructs used for co-translation folding studies

Plasmid name	Description	Reference
pDM26	pEXT22 TorA-ADR1a	This study
pDM27	pEXT22 TorA-R16	This study
pDM28	pEXT22 TorA-SOD1	This study
pDM29	pEXT22 TorA-ProteinG	This study
pDM31	pEXT22 hGH TorA-KK	This study
pDM32	pEXT22 SufI-KK	This study
pDM33	pEXT22 BT6 TorA-KK	This study
pDM38	pING BT6h2 TorA-RR L26	This study
pDM39	pING BT6h2 TorA-RR L30	This study
pDM40	pING BT6h2 TorA-RR L35	This study
pDM41	pING BT6h2 TorA-RR L40	This study
pDM42	pING BT6h2 TorA-RR L45	This study
pDM43	pING BT6h2 TorA-RR L50	This study
pDM44	pING BT6h2 TorA-RR L55	This study
pDM45	pING BT6h2 TorA-RR L60	This study
pDM46	pING BT6h2 TorA-RR L30 FL	This study
pDM47	pING BT6h2 TorA-RR L30 A	This study
pDM48	pING BT6h2 TorA-RR L60 FL	This study
pDM49	pING BT6h2 TorA-RR L60 A	This study
pDM50	pING BT6h2 TorA-KK L26	This study
pDM51	pING BT6h2 TorA-KK L30	This study
pDM52	pING BT6h2 TorA-KK L35	This study
pDM53	pING BT6h2 TorA-KK L40	This study

Plasmid name	Description	Reference
pDM54	pING BT6h2 TorA-KK L45	This study
pDM55	pING BT6h2 TorA-KK L50	This study
pDM56	pING BT6h2 TorA-KK L55	This study
pDM57	pING BT6h2 TorA-KK L60	This study
pDM58	pING BT6h2 TorA-KK L30 FL	This study
pDM59	pING BT6h2 TorA-KK L30 A	This study
pDM60	pING BT6h2 TorA-KK L60 FL	This study
pDM61	pING BT6h2 TorA-KK L60 A	This study
pDM62	pING hGH TorA-RR L26	This study
pDM63	pING hGH TorA-RR L30	This study
pDM64	pING hGH TorA-RR L35	This study
pDM65	pING hGH TorA-RR L40	This study
pDM66	pING hGH TorA-RR L45	This study
pDM67	pING hGH TorA-RR L50	This study
pDM68	pING hGH TorA-RR L55	This study
pDM69	pING hGH TorA-RR L60	This study
pDM70	pING hGH TorA-RR L30 FL	This study
pDM71	pING hGH TorA-RR L30 A	This study
pDM72	pING hGH TorA-RR L60 FL	This study
pDM73	pING hGH TorA-RR L60 A	This study
pDM74	pING hGH TorA-KK L26	This study
pDM75	pING hGH TorA-KK L30	This study
pDM76	pING hGH TorA-KK L35	This study
pDM77	pING hGH TorA-KK L40	This study

Plasmid name	Description	Reference
pDM78	pING hGH TorA-KK L45	This study
pDM79	pING hGH TorA-KK L50	This study
pDM80	pING hGH TorA-KK L55	This study
pDM81	pING hGH TorA-KK L60	This study
pDM82	pING hGH TorA-KK L30 FL	This study
pDM83	pING hGH TorA-KK L30 A	This study
pDM84	pING hGH TorA-KK L60 FL	This study
pDM85	pING hGH TorA-KK L60 A	This study
pDM86	pING SufI RR L26	This study
pDM87	pING SufI RR L30	This study
pDM88	pING SufI RR L35	This study
pDM89	pING SufI RR L40	This study
pDM90	pING SufI RR L45	This study
pDM91	pING SufI RR L50	This study
pDM92	pING SufI RR L55	This study
pDM93	pING SufI RR L60	This study
pDM94	pING SufI RR L30 FL	This study
pDM95	pING SufI RR L30 A	This study
pDM96	pING SufI RR L60 FL	This study
pDM97	pING SufI RR L60 A	This study
pDM98	pING SufI KK L26	This study
pDM99	pING SufI KK L30	This study
pDM100	pING SufI KK L35	This study
pDM101	pING SufI KK L40	This study

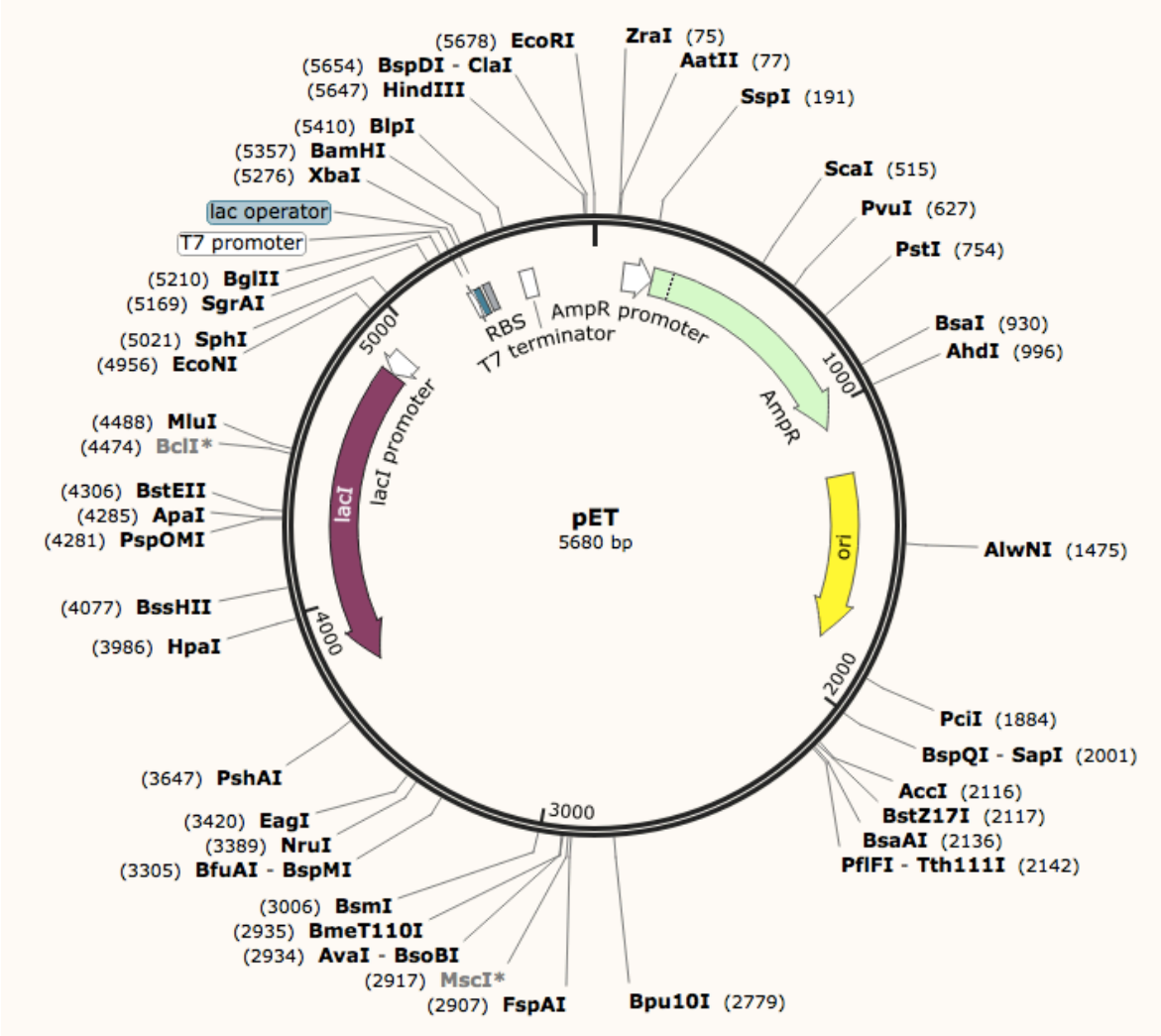
Plasmid name	Description	Reference
pDM102	pING SufI KK L45	This study
pDM103	pING SufI KK L50	This study
pDM104	pING SufI KK L55	This study
pDM105	pING SufI KK L60	This study
pDM106	pING SufI KK L30 FL	This study
pDM107	pING SufI KK L30 A	This study
pDM108	pING SufI KK L60 FL	This study
pDM109	pING SufI KK L60 A	This study
pDM110	pET BT6h2 TorA-RR L26	This study
pDM111	pET BT6h2 TorA-RR L30	This study
pDM112	pET BT6h2 TorA-RR L35	This study
pDM113	pET BT6h2 TorA-RR L40	This study
pDM114	pET BT6h2 TorA-RR L45	This study
pDM115	pET BT6h2 TorA-RR L50	This study
pDM116	pET BT6h2 TorA-RR L55	This study
pDM117	pET BT6h2 TorA-RR L60	This study
pDM118	pET BT6h2 TorA-RR L30 FL	This study
pDM119	pET BT6h2 TorA-RR L30 A	This study
pDM120	pET BT6h2 TorA-RR L60 FL	This study
pDM121	pET BT6h2 TorA-RR L60 A	This study
pDM122	pET BT6h2 TorA-KK L26	This study
pDM123	pET BT6h2 TorA-KK L30	This study
pDM124	pET BT6h2 TorA-KK L35	This study
pDM125	pET BT6h2 TorA-KK L40	This study

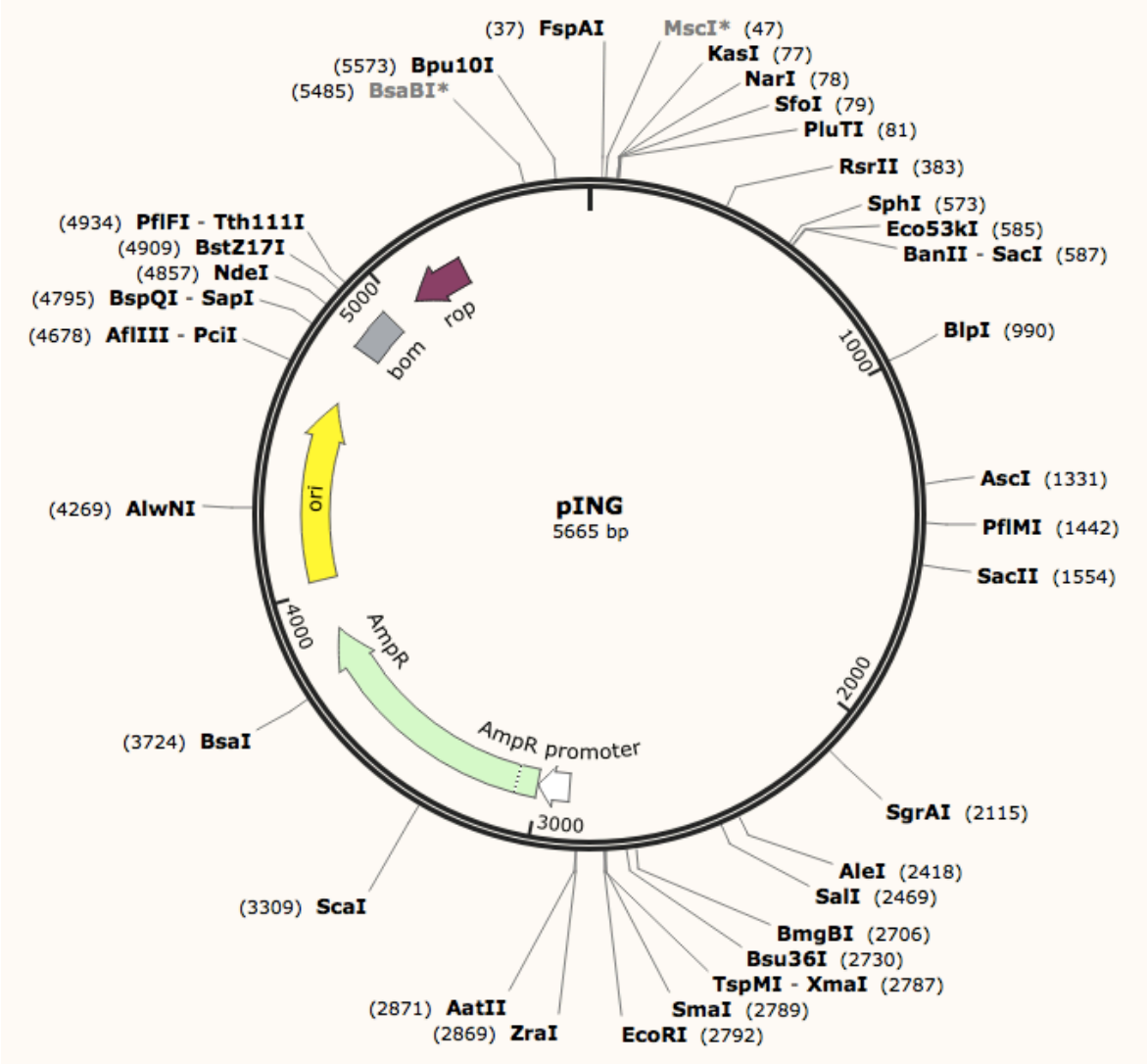
Plasmid name	Description	Reference
pDM126	pET BT6h2 TorA-KK L45	This study
pDM127	pET BT6h2 TorA-KK L50	This study
pDM128	pET BT6h2 TorA-KK L55	This study
pDM129	pET BT6h2 TorA-KK L60	This study
pDM130	pET BT6h2 TorA-KK L30 FL	This study
pDM131	pET BT6h2 TorA-KK L30 A	This study
pDM132	pET BT6h2 TorA-KK L60 FL	This study
pDM133	pET BT6h2 TorA-KK L60 A	This study
pDM134	pET hGH TorA-RR L26	This study
pDM135	pET hGH TorA-RR L30	This study
pDM136	pET hGH TorA-RR L35	This study
pDM137	pET hGH TorA-RR L40	This study
pDM138	pET hGH TorA-RR L45	This study
pDM139	pET hGH TorA-RR L50	This study
pDM140	pET hGH TorA-RR L55	This study
pDM141	pET hGH TorA-RR L60	This study
pDM142	pET hGH TorA-RR L30 FL	This study
pDM143	pET hGH TorA-RR L30 A	This study
pDM144	pET hGH TorA-RR L60 FL	This study
pDM145	pET hGH TorA-RR L60 A	This study
pDM146	pET hGH TorA-KK L26	This study
pDM147	pET hGH TorA-KK L30	This study
pDM148	pET hGH TorA-KK L35	This study
pDM149	pET hGH TorA-KK L40	This study

Plasmid name	Description	Reference
pDM150	pET hGH TorA-KK L45	This study
pDM151	pET hGH TorA-KK L50	This study
pDM152	pET hGH TorA-KK L55	This study
pDM153	pET hGH TorA-KK L60	This study
pDM154	pET hGH TorA-KK L30 FL	This study
pDM155	pET hGH TorA-KK L30 A	This study
pDM156	pET hGH TorA-KK L60 FL	This study
pDM157	pET hGH TorA-KK L60 A	This study
pDM158	pET SufI RR L26	This study
pDM159	pET SufI RR L30	This study
pDM160	pET SufI RR L35	This study
pDM161	pET SufI RR L40	This study
pDM162	pET SufI RR L45	This study
pDM163	pET SufI RR L50	This study
pDM164	pET SufI RR L55	This study
pDM165	pET SufI RR L60	This study
pDM166	pET SufI RR L30 FL	This study
pDM167	pET SufI RR L30 A	This study
pDM168	pET SufI RR L60 FL	This study
pDM169	pET SufI RR L60 A	This study
pDM170	pET SufI KK L26	This study
pDM171	pET SufI KK L30	This study
pDM172	pET SufI KK L35	This study
pDM173	pET SufI KK L40	This study

Plasmid name	Description	Reference
pDM174	pET SufI KK L45	This study
pDM175	pET SufI KK L50	This study
pDM176	pET SufI KK L55	This study
pDM177	pET SufI KK L60	This study
pDM178	pET SufI KK L30 FL	This study
pDM179	pET SufI KK L30 A	This study
pDM180	pET SufI KK L60 FL	This study
pDM181	pET SufI KK L60 A	This study


(6201) SpeI
(6148) XbaI
(6102) SexAI*
(6090) PaeR7I - XhoI
(6033) BamHI
(6018) EcoRI
lac operator
tac promoter
(5932) SspI
(5702) PshAI
(5641) AfeI
(5465) MauBI
AarI (111)
EagI (593)
NeoR/KanR
NcoI (1119)
BstBI (1369)
BmgBI (2442)
AleI (2641)
AflII (2709)
PasI (2783)
SbfI (3121)
(4389) BsmBI
Myc
FLAG
rrnB T2 terminator
lacI
lacIq promoter
pEXT22
6257 bp
600
1000
2000
3000
4000
5000


pET
5680 bp
(5678) EcoRI
(5654) BspDI - ClaI
(5647) HindIII
(5410) BlpI
(5357) BamHI
(5276) XbaI
lac operator
T7 promoter
(5210) BglII
(5169) SgrAI
(5021) SphI
(4956) EcoNI
(4488) MluI
(4474) BclI*
(4306) BstEII
(4285) ApaI
(4281) PspOMI
(4077) BssHII
(3986) HpaI
(3647) PshAI
(3420) EagI
(3389) NruI
(3305) BfuAI - BspMI
(3006) BsmI
(2935) BmeT110I
(2934) AvaI - BsoBI
(2917) MscI*
(2907) FspAI
Bpu10I (2779)
PciI (1884)
BspQI - SapI (2001)
AccI (2116)
BstZ17I (2117)
BsaAI (2136)
PflFI - Tth111I (2142)
AlwNI (1475)
BsaI (930)
AhdI (996)
PstI (754)
PvuI (627)
ScaI (515)
ZraI (75)
AatII (77)
SspI (191)
RBS
T7 terminator
AmpR promoter
AmpR
ori
lacI
lacI promoter
5000
1000
2000
3000
4000


pING
5665 bp
(37) FspAI
MscI* (47)
KasI (77)
NarI (78)
SfoI (79)
PluTI (81)
RsrII (383)
SphI (573)
Eco53kI (585)
BanII - SacI (587)
BlpI (990)
AscI (1331)
PflMI (1442)
SacII (1554)
SgrAI (2115)
AleI (2418)
SalI (2469)
BmgBI (2706)
Bsu36I (2730)
TspMI - XmaI (2787)
SmaI (2789)
EcoRI (2792)
(2869) ZraI
(2871) AatII
(3309) ScaI
(3724) BsaI
(4269) AlwNI
(4678) AflIII - PciI
(4795) BspQI - SapI
(4857) NdeI
(4909) BstZ17I
(4934) PflFI - Tth111I
(5485) BsaBI*
(5573) Bpu10I
1000
2000
3000
4000
5000
ori
bom
rop
AmpR
AmpR promoter

The *Escherichia coli* Tat system senses localized unfolded regions of a *de novo* - designed maquette protein.

Daphne M.J. Mermans[1], Alexander S. Jones[1], Gary S. Thompson[1], C. Neil Hunter[2], P. Leslie Dutton[3], Colin Robinson[1*]

[1]School of Biosciences, University of Kent, Canterbury CT2 7NJ, United Kingdom

[2]Department of Molecular Biology and Biotechnology, University of Sheffield , Sheffield, S10 2TN, United Kingdom

[3]Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

* Corresponding author.

Email c.robinson-504@kent.ac.uk

Tel: +44 1227 823443

Abstract

The *Escherichia coli* Tat system transports a range of folded proteins across the plasma membrane. It preferentially exports correctly folded proteins, so clearly has the ability to discriminate between folding states, but the underlying proofreading mechanism is poorly understood. We used a simple, *de novo* designed substrate to test the importance of surface charge and structural dynamics. A 4-helix 'maquette' termed BT6 has been previously shown to be exported by Tat, but only when correctly folded around two heme *b* molecules; a partially folded 1-heme variant (BT6M1) was very poorly exported, while a zero-heme form (BT6M0) was not exported at all. Here, we show that significant changes to the BT6 surface charge, involving substitution of 2 or 4 Asp residues by Lys (creating BT6-2Lys and BT6-4Lys), or 2 Lys residues by Asp (BT6-2Asp), had little effect on export. This demonstrates that extensive surface charge variations do not influence the proofreading of this protein by Tat. Exported BT6-2Lys and BT6-4Lys were each shown to contain 2 heme molecules, suggesting that they are properly folded, whereas BT6-2Asp contains only 1 heme. Surprisingly, BT6-2Asp is exported far more efficiently than the other single-heme form, BT6M1, and indeed more efficiently than the 'wild type' BT6 2-heme protein. 1D-NMR analysis demonstrates that BT6M1 and BT6-2Asp are folded to very similar extents in overall terms, with any differences confined to specific localisations. Overall, BT6 is considerably more folded than either. These data show that the Tat system's proofreading mechanism does not sense a substrate's overall folding status, and we propose instead that Tat prefentially

senses the folding status of localised regions. We further show that even small sequences appended to the C-terminus can block transport.

Introduction

The Tat system is an unusual protein transport system that operates in most bacterial cytoplasmic membranes. In *E. coli* the system minimally consists of three integral membrane proteins, TatA, TatB and TatC, and Tat dependent substrates are characterised by an N-terminal signal peptide containing a conserved twin arginine (RR) motif. After translocation of the precursor protein to the periplasm, the signal peptide is cleaved by signal peptidase (reviewed by xxx). A key feature of the system is its ability to transport fully folded proteins, and the system is indeed able to preferentially transport correctly folded proteins; numerous studies have shown that the Tat system possesses a 'proofreading' capability that enables it to reject incorrectly folded proteins (Robinson et al., 2011, Maurer *et al.,* 2009, DeLisa *et al*., 2003).

The mechanism of this proofreading activity remains poorly understood. It has been shown that the Tat system tolerates minor structural changes as long as a near-native tertiary structure is present (Alanen *et al.*, 2015). Furthermore, it has been suggested that the system senses hydrophobicity rather than the folded state of substrates (Richter *et al.*, 2007). However, other studies demonstrated that the Tat system can tolerate significant changes in substrate surface hydrophobicity in a folded heterologous substrate (Jones *et al.*, 2016).

The aim of this study was to study the Tat export system's proofreading ability, using a simple model substrate to introduce (i) changes in the surface characteristics, (ii) alterations to the folding state (iii) additional unfolded domains. Export assays were used to see which variants are exported and which are rejected. To analyse the proofreading capacity we chose a *de novo*-designed 'maquette' protein substrate (Farid *et al.*, 2014). The structure of the 'BT6' variant of the maquette is shown in Figure 1A. BT6 shares no sequence identity with natural substrates and the complexity was minimized to increase engineering freedom (Farid *et al.*, 2014). The protein consists of four α-helices that enclose a hydrophobic cavity that accommodates cofactors. The BT6 interior contains *bis*-histidine ligation sites for binding 2 molecules of heme *b*.

We have previously shown that the Tat translocase is able to recognize and distinguish between folded and unfolded BT6 maquette substrates (Sutherland *et al.*, 2018). That study tested whether there is a difference in translocation of fully folded (BT6), partially folded (BT6M1), and unfolded BT6 (TorA-BT6M0) variants. The folding state of the BT6 maquettes was affected by reducing the number of bound hemes. The resulting hypothesis is that the Tat translocase must sens differences in structural flexibility of the proteins.

In this study, we examined the Tat systems proofreading ability by using BT6 maquette. BT6 substrates were mutated in order to change the surface characteristics to probe the tolerance of the Tat systems proofreading mechanism. The results suggest that the Tat proofreading system does not

sense a global unfolded state of the substrate but instead sneses the folding state of specific, localised regions of the protein.

Materials and methods

Plasmids and bacterial strains

Table 1 shows the constructs that were synthesised with a C-terminal 6xhistidine tag by GeneArt (Life Technologies, Thermo Fisher Scientific, UK). Constructs were ligated into pEXT22 by restriction cloning. Restriction digest was carried out in a 50 µl reaction with 1 µg of template DNA, 1 µl Buffer 3.1 (New England Biolabs, UK), 1.5 µl NdeI and incubated at 37 °C for 30 min before adding 1 µl BamHI and incubating for another 1.5 h at 37 °C. Digestion products where separated on 1% (w/v) agarose (Bio-Rad Laboratories Ltd., USA) gels by electrophoresis for 30 min at 150 V. Bands were visualised upon UV exposure and bands of interest were excised using a scalpel blade before being purified with QIAprep Gel Extraction kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. Purified DNA fragments were ligated into vector pEXT22 using T4 DNA ligase (Roche). Insert and vector were mixed in a 3:1 ratio with 1 µl ligase buffer and 1 µl T4 ligase, before incubation overnight in Biometra T3 Thermocycler. After ligation 5 µl ligation product was used to transform 50 µl *E. coli* DH5α and W3110 competent cells.

Cell culture and fractionation

5 ml LB medium (10 g/L sodium chloride, 10 g/L tryptone, 5 g/L yeast extract) containing 100 µg/mL kanamycin was inoculated with a single colony of the *E. coli* strains carrying the respective plasmids and grown overnight at 37 °C, 200 rpm. 50 mL of fresh LB medium containing 100 µg/mL kanamycin was then inoculated with the overnight culture to OD_{600} = 0.05. Cultures were then grown at 37 °C, 200 rpm to OD_{600} = 0.5 before induction with 0.5 mM IPTG. 3 hours post induction an amount equivalent to OD_{600} = 10 was collected and fractionated into Cytoplasm (C), Membrane (M), and Periplasm (P) fractions. The periplasmic fraction was collected using the EDTA/lysozyme/cold osmotic shock and centrifugation method (Matos *et al.*, 2014). The spheroplast pellet was then washed by resuspension in buffer containing 50 mM Tris-Acetate (pH 8.2), 250 mM sucrose and 10 mM $MgSO_4$ and centrifuged at 14,000 rpm, 4 °C, 5 min. The resulting pellet was then resuspended in buffer containing 50mM Tris-Acetate, 2.5 mM EDTA (pH 8.2) and sonicated for 6 x 10 s, 8 µm amplitude (Soniprep 150plus, Sanyo Gallenkamp, Loughborough, UK). The sonicated samples were then centrifuged at 70,000 rpm, 4 °C, 30 min. The supernatant was taken as the cytoplasmic fraction (C) and the pellet was resuspended in buffer containing 50 mM Tris-Acetate, 2.5 mM EDTA (pH 8.2) to obtain the Membrane fraction (M). All cell fractions were stored frozen at – 20 °C.

Western blotting

The fractionation protein samples were transferred to PVDF-membranes (GE Healthcare, Buckinghamshire, UK). The PVDF membranes were then blocked overnight at 4 °C with PBS-T containing 5% (w/v) dried skimmed milk and immunoblotted with C-terminal His antibodies. The membranes were then washed with 1x PBS-tween20 (0.1%) before incubation with 3.5 µl 6x-His C-terminal Tag HRP antibody (New England Biolabs, UK) in 20 mL 1x PBS tween20 (0.1%) for 1 hour. Membranes were washed and proteins were visualised using an ECL (enhanced chemiluminescence) kit (Biorad, Herts, UK) according to the manufacturer's instructions. A BioRad chemiluminescence imager and corresponding software was used to develop the membranes.

Export assay for purification of BT6 variants

For the purification of the BT6 variants the pEXT22 vectors BT6, BT6M1, TorA-BT6m0, BT6-2Lys, BT6-4Lys and BT6-2Asp were transformed into the *E. coli* W3110 'TatExpress' cell line (2). Transformed cells were grown in 50 ml LB medium (10g/L sodium chloride, 10g/L tryptone, 5 g/L yeast extract) containing 100 µg/mL kanamycin was inoculated with a single colony and grown overnight at 37 °C, 200 rpm. 400 mL fresh LB medium and 100 µg/mL kanamycin were inoculated with overnight culture to OD_{600} = 0.05. Cultures were then grown at 30 °C, 200 rpm to OD_{600} = 0.5 before induction of plasmid with 0.5mM IPTG and cultures were incubated for 24 h, and harvested by centrifugation (4000 rpm, 20

min, 4 °C). To obtain the periplasmic fraction cells were resuspended in 10 mL buffer containing 100 mM Tris-acetate pH 8.2, 500 mM sucrose, 5 mM EDTA pH 8.0) and 10 mL ice-cold dH_2O before addition of 800 µl hen egg white lysozyme (1 mg/mL) followed by incubation on ice for 10 min. To stabilise the inner membrane 800 µl 1 M $MgSO_4$ was added to the solution before centrifugation (14,000 rpm, 20 min, 4 °C). The supernatant was taken to collect the periplasmic fraction. The spheroplast pellet was resuspended in 10 mL 0.5 M Sodiumphosphate buffer (pH 7.3) and protease inhibitor (cOmplete Mini Protease Inhibitor, Roche) was added. Cells were sonicated for 6 x 30 s, 8 µm amplitude (Soniprep 150plus, Sanyo Gallenkamp, Loughborough, UK) and centrifuged (14,000 rpm, 20 min, 4 °C). The supernatant was collected as the spheroplast fraction.

Purification of Tat-exported proteins from E. coli periplasm

IMAC Sepharose Fast Flow Column (GE Healthcare) was used to purify the BT6 variants. After equilibration of the column with 20 mM Sodium Phosphate, periplasmic and spheroplastic fractions were applied to the column. The column was washed with 20 mL of buffer containing 25 mM imidazole, 0.5M NaCl, 20mM Sodiumphosphate, pH 7.3. BT6 variants were eluted with 9 mL buffer containing a gradient of 100 – 200 mM imidazole, 50 mM EDTA, 20 mM Sodiumphosphate pH 7.3. The elution fractions were collected and concentrated using Vivaspin centrifugal concentrators (Sartorius).

1D-1H-NMR spectra

1D-1H NMR spectra of BT6, BT6M1, TorA-BT6M0 and BT6-2Asp were acquired on a Bruker Advance III 600MHz spectrometer with a cryogenically cooled probe at 298K. Water supressed spectra were acquired using a Double Pulse Field Gradient Selected Excitation Experiment (DPFGSE) (Hwang and Shaka, 1995). All experiments were acquired with the same parameters. Spectra were acquired using 2048 complex points with a spectral width of 9615Hz, an inter-scan delay of 1s and 2048 repetitions for improved signal to noise. Data were processed in topspin 3.5pl7. All spectra were zero filled once, apodised with an exponential function (LB=3) and the residual water signal was supressed with a convolution based solvent filter (quadrature Gaussian window of width 0.2ppm) (Marion *et al.*, 1989).

Results

BT6 is an ideal model substrate for studies on Tat proofreading because the protein is simple, tightly folded and efficiently exported in *E. coli* when a Tat-specific TorA signal peptide is attached. The protein's conformation is tightly constrained by ligation to 2 heme molecules (at both ends of the protein as shown in Fig. 1). The 2-heme protein (BT6) is efficiently exported by Tat whereas a variant that binds only 1 heme (BT6M1) is very poorly exported, and a no-heme variant (BT6M0) is not exported at all (Sutherland et al, 2018). Clearly, Tat senses the less tightly folded status of the mutated forms. In this study we

created a series of new mutated forms in order to further analyse this proofreading capability, and where possible we purified the proteins to determine whether they contain 0, 1 or 2 hemes. For the purposes of this study we label 2 heme-containing variants as 'folded'. Figure 1 shows the structure, surface electrostatis potential and surface hydrophobicity map of the fully folded two-heme binding, BT6. The outer surface is predicted to be relatively hydrophilic with the hydrophilic surface residues displayed in purple and the hydrophobic surface residues in yellow.

The Tat system tolerates changes in substrate surface charge

We first tested whether changes in the substrate's outer surface electrostatic potential (Figure 1B) affects export to the periplasm by the Tat translocase. In a study using an scFv as substrate, it was suggested that the Tat system is highly tolerant towards surface charge changes (Jones *et al.*, 2016). However, the Tat translocase might sense certain outer surface characteristics as unfolded, and we therefore created four BT6 mutants with altered outer surface characteristics (Figure 2A). Two mutants were designed with 2 or 4 Lys residues substituted by Asp residues (denoted BT6-2Asp and BT6-4Asp); these substitutions thus change original BT6 protein's charge by -4 and -8, respectively. In contrast, another two mutants were designed with 2 or 4 Asp residues substituted by Lys (denoted BT6-2Lys and BT6-4Lys). The proteins were expressed in *E. coli* with an N-terminal TorA signal peptide TorA and a C-terminal 6x His-tag, and Figure 2B shows BT6-2Lys, BT6-4Lys and BT6-2Asp are all efficiently exported with a

mature size band observed in the periplasm (P). Note that the BT6-2Asp mature protein has a different mobility (it migrates more slowly than the 17 kDa BT6) due to the charge difference. There is no mature sized band in the periplasmic fraction with BT6-4Asp, showing that these mutations completely block export.

Efficient export usually indicates that the protein contains 2 heme molecules and is folded, but it was considered important to test this point so we purified BT6-2Lys, BT6-4Lys and BT6-2Asp from the periplasmic fraction by immobilized metal affinity chromatography (IMAC). Figure 2C shows the color of the purified BT6 variants. Previously, it has been demonstrated that purified BT6 has an intense red color, the partially folded BT6M1 has an orange color and an unfolded (no heme) BT6M0 variant is colorless (Sutherland *et al*., 2018). After purification, the intense red color indicated that BT6-2Lys and BT6-4Lys contain two hemes and are therefore folded (the absorption spectra are shown in supplementary Figure 1). In contrast, BT6-2Asp had an orange color, which indicates that this BT6 variant only accommodates one heme.

We confirmed this point by comparing absorbance spectra of purified BT6, BT6M1, BT6M0 and BT6-2Asp, normalized at 280nm, using the procedure adopted by Sutherland et al (2018). The BT6-2Asp and BT6M1 spectra are essentially identical (and very different to that of BT6), confirming that both contain only one heme (Figure 3). Unfortunately, the cytoplasmic BT6-4Asp protein could not be purified in sufficient quantity for analysis and we therefore

cannot determine whether the lack of export of this variant is due to changes on the surface of the protein or effects on folding.

This results obtained with BT6-2Asp are unexpected because this protein was exported with high efficiency (Figure 2), yet contains only a single heme, which implies partial unfolding. To address this point in more detail we directly compared the export kinetics of the BT6, BT6-2Asp and the previously-characterised single-heme BT6M1 precursor forms in a time-course analysis. The cells were induced with IPTG and the periplasmic fraction was harvested after 60, 120, and 300 min. Figure 4 confirms that TorA-BT6 is exported much more efficiently than TorA-BT6M1, as observed by Sutherland et al. (2018); the BT6 periplasmic band is clearly detectable even at the earlier time points whereas the BT6M1 band is barely detectable at any stage of the time course. However, TorA-BT6-2Asp is exported even more efficiently that the non-mutated BT6 form, with export essentially complete at the earliest time point and the band far stronger on the immunoblot. BT6-2Asp is certainly exported far more efficiently than the other single-heme protein, BT6M1.

1D NMR indicates that BT6-2Asp is generally less tightly folded than BT6, but has a specific region that is more tightly folded

We purified the BT6 variants and subjected them to 1D-NMR analysis because the nature of the technique is very sensitive to the structure and dynamics of molecules at the atomic scale on timescales of ms-ns. However, the spectra of

the maquettes are complicated because of the presence of heme groups with a high spin Fe^{III} present (low spin heme groups only occur when O_2 is bound to the hemes Fe^{III} ion). This leads to two processes occurring as hemes are bound to the BT6 proteins. The first is that the presence of a heme group in the maquette may increase the 'foldedness' or order of the protein leading to an increase in chemical shift dispersion and a broadening of the proteins peaks to the line width expected for a protein of this size. In the case of the maquettes (mass ~ 21 kDa) we may predict a linewidth of somewhere around 18Hz for an amide or 0.03 ppm at 600 MHz (Cavenagh *et al*, 2006) assuming that only dipole-dipole relaxation is present and making the approximation that the protein is approximately spherical.

The second effect is that the lone pair from the paramagnetic centre in the heme can lead to extreme shifts for nuclei both within the porphyrin ring that encloses it and also in nearby residues in the protein. On top of this the paramagnetic centre present in the heme can also cause broadening due to increased T2 relaxation. In interpretation of the results it is assumed based on the observed data and previous analysis that BT6M0 contains no heme and so is not broadened, but that BT6, BT6M1 and BT6-2Asp are all bound to heme to some extent and so show broadening and extreme shifts. Examples of these extreme shifts can be seen in Figure 5A for the low field region of their spectra, with broad dispersed peaks being observed between 9.5 and 10.5 ppm.

A comparison of the 1D NMR specta for BT6M1 and BT6-2Asp shows that they are almost identical, while the BT6 and BT6-2Asp spectra show clear differences.

As reported by Sutherland et al (2018), BT6 and BT6M1 both show broadening of the envelope for the amide and aliphatic side chain protons compared with TorA-BT6M0 showing that Fe^{III} heme groups are bound and also possibly suggesting an increase in the structuring of the protein (BT6M1 is somewhat sharper but still well dispersed, most probably because there is less relaxation from the single bound heme-Fe^{III} group). The mutant protein BT6-2Asp also shows similar broadening of the aliphatic and amide envelopes similar to BT6M1 suggesting that the structure present is very similar in most respects to that in BT6M1.

However, analysis of the extremely shifted peaks in the proteins above 9.5 ppm (Figure 5B) may suggest that a more subtle and interesting analysis can be made. For both BT6M1 and especially BT6 a number of broad peaks are observed in this region which could be attributed to highly shifted peaks from the heme and the protein due to aromatic ring currents and hyperfine shifts from the heme-Fe^{III} moiety in a similar fashion to those observed in previous studies (Du *et al.*, 2003). However, the observed peaks are considerably broader than those observed in Du et al leading to the conclusion that there is some movement of the heme within the structure of the maquette leading to the heme taking up multiple orientations which leads to further broadening due to exchange effects. It can also be seen that these peaks cannot be attributed to just the tryptophan side chains which are also present in this region as the area of the observed peaks appears to be too large to be consistent with the 4 tryptophan peaks present (the area of the region 9.5-10.5 ppm is $1/32^{nd}$ of the amide and aromatic ring region; 6.1-9.5 ppm whereas it should be 1 ~70th). However, analysis of the same region

in the protein BT6-2Asp shows two additional peaks at 10.05 ppm, which are entirely consistent with two of the tryptophan residues present in the protein. As these peaks are not observed in the spectra of any other proteins studied this suggests that there is an ordering around two of the tryptophans (if it was an ordering of the heme group it would be expected that many more signals would be observed). Furthermore, the linewidths observed (10 Hz) are not inconsistent with the peaks expected from the tryptophans of a protein of mass. It could be suggested, as an alternative that the peaks could be caused by a decrease of order as this can also sharpen peaks, however, as there is no other sign of sharpening in the overall envelope of the protein this seems unlikely. Therefore, overall it can be suggested that a specific region in BT6-2Asp is more ordered and structured than BT6M1 and possibly BT6.

Addition of even a short C-terminal unfolded domain is not tolerated by the Tat system

Previous studies suggested that the addition of an unfolded element to a folded substrate blocks export by the Tat system (Jones *et al.*, 2016); a 26-residue unfolded domain was added to the C-terminus of a folded scFv protein and the protein was completely rejected. In this study, we capitalised on the highly folded nature of BT6 to test whether the addition of variably-sized unfolded polypeptide chains to a fully folded BT6 maquette affects export. We created three BT6 variants in which 26 residues (SNVIIIIITNKDPNSSSVDLAAALE), 13 residues (SNVIIIIITNKDP) or 6 residues (SNVIII) were added to the C-terminus of TorA-

BT6. As described previously, an N-terminal Tat signal peptide TorA and a C-terminal 6x His-tag were fused to the BT6 variants. Figure 6 shows that there is no mature sized band of (17 kDa) observed for any of the three BT6 variants in the periplasmic fraction (P). The Tat substrates are degraded in the cytoplasm when export to the periplasm does not take place.

The C-terminal extensions were not predicted to affect folding and we attempted to test this directly. The low levels of cytoplasmic protein precluded purification of large quantities but the 3 proteins were partially purified from cytoplasmic extracts and Fig. 6B shows absorption spectra. Although the spectra are not as clean as those obtained for the periplasmic proteins (Fig. 3), each shows a peak at 412 nm that is indicative of heme binding. The data strongly suggest that Tat can detect the presence of an additional 6-, 13- or 26-residue C-terminal unfolded domain and block the translocation of these BT6 maquette variants.

Discussion

The Tat system exports folded and cofactor-containing proteins from the cytoplasm to the periplasm, and it somehow senses the folding state of a substrate and prevents premature export to the periplasm. The details of this process are, however, lacking and we are particularly interested knowing what Tat recognizes as 'unfolded' before rejecting a substrate. Does Tat sense folding dynamics, and reject proteins that are insufficiently rigid, and if so does it somehow sample the entire protein or only certain regions? We also wanted to

address whether significant changes in surface charge distribution could have an effect. In this study, we investigated this remarkable proofreading system using a *de novo* co-factor binding 'maquette' protein. We designed a variety of BT6 variants with alterations on the substrate's outer surface, tested whether they were fully or partially folded, and used export assays to determine whether the mutants were accepted or rejected by the Tat components.

The initial export assays with BT6-2Lys and BT6-4Lys showed that the Tat system tolerated substantial changes in charge with no detectable effect on export efficiency. The 4 mutations on the surface of BT6-4Lys result in a substantial change in charge (a net change of +8 within a small protein) and these data suggest that major changes in surface charge can be tolerated. The BT6-4Asp mutant was not exported, but we cannot determine whether this is due to associated changes in folding.

The BT6-2Asp variant yielded unexpected information. As with BT6-2Lys and BT6-4Lys, this variant was efficiently exported, but whereas the former were shown to bind two heme molecules it is clear that BT6-2Asp binds only one, and this protein should therefore be partially unfolded, as was clearly shown for the other 1-heme mutant, BT6M1 (Sutherland et al., 2018). 1D-NMR data confirm that this is the case: BT6-2Asp and BT6M1 have nearly identical spectra which shows that at a general level their folding states are very similar indeed. We conclude that the Tat proofreading system does not assess the overall folding status of the substrate, but rather must respond to the folding state of localised

domains in the substrate. With BT6-2Asp and BT6M1, we believe that one of these important domains is one of the few regions that differ in terms of conformational flexibility, and the NMR data furthermore raise the possibility that this domain may include unspecified Trp residues in the BT6 protein. The inference is that this domain is particularly rigid in the BT6-2Asp variant, and that the Tat system accepts this variant as folded because the domain is one of the primary interaction sites during the binding/ proofreading process.

Finally, we tested whether the Tat components will reject a folded substrate that contains an unfolded domain of varying length. Three BT6 mutants were designed with an additional 6-, 13- or 26-residue C-terminal unfolded 'tail' (BT6-Cterm26, BT6-Cterm13 and BT6-Cterm6). Tat rejected all of these proteins, strongly suggesting that in this case, the Tat system has the ability to sense relatively short unfolded domains attached to a small α-helical protein. Analysis of the BT6 structure (Fig. 1) shows that the N- and C-termini are in close proximity, and these C-terminal extensions are therefore likely to be close to the Tat apparatus when the N-terminal signal peptide mediates binding. This may partly explain why the extensions have such a ramatic impact.

Acknowledgements

CR and DMJM were supported by the Marie Curie Initial Training Network grant (Horizon 2020, ProteinFactory, 642863). AJ was supported by Biotechnology and Biological Sciences Research Council ‘Bioprocessing Research Industry Club’

grant BB/ K011219/1.We would like to thank Kevin Howland for carrying out the mass spectrometry analysis and the University of Kent Welcome Trust Biomolecular NMR Facility.

References

Alami, M., Lüke, I., Deitermann, S., Eisner, G., Koch, H. G., Brunner, J., & Müller, M. (2003). Differential interactions between a twin-arginine signal peptide and its translocase in Escherichia coli. *Molecular Cell*, *12*(4), 937–946. https://doi.org/10.1016/S1097-2765(03)00398-8

Alanen, H. I., Walker, K. L., Lourdes Velez Suberbie, M., Matos, C. F. R. O., Bönisch, S., Freedman, R. B., … Robinson, C. (2015). Efficient export of human growth hormone, interferon α2b and antibody fragments to the periplasm by the Escherichia coli Tat pathway in the absence of prior disulfide bond formation. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1853*(3), 756–763. https://doi.org/10.1016/j.bbamcr.2014.12.027

Berks, B. C. (1996). A common export pathway for proteins binding complex redox cofactors? *Molecular Microbiology*, *22*(3), 393–404. https://doi.org/8939424

Cavenagh Et al. (2006). *Protein NMR spectroscopy* (2nd ed.).

Du, W., Xia, Z., Dewilde, S., Moens, L., & La Mar, G. N. (2003). 1H NMR study of the molecular structure and magnetic properties of the active site for the cyanomet complex of O2-avid hemoglobin from the trematode Paramphistomum epiclitum. *European Journal of Biochemistry*, *270*(13), 2707–2720. https://doi.org/10.1046/j.1432-1033.2003.03638.x

Farid, T. A., Kodali, G., Solomon, L. A., Lichtenstein, B. R., Molly, M., Fry, B. A., … Dutton, P. L. (2014). NIH Public Access, *9*(12), 826–833. https://doi.org/10.1038/nchembio.1362.Elementary

Hwang, T. L., & Shaka, A. J. (1995). Water Suppression That Works. Excitation

Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *Journal of Magnetic Resonance - Series A*. https://doi.org/10.1006/jmra.1995.1047

Jones, A. S., Austerberry, J. I., Dajani, R., Warwicker, J., Curtis, R., Derrick, J. P., & Robinson, C. (2016). Proofreading of substrate structure by the Twin-Arginine Translocase is highly dependent on substrate conformational flexibility but surprisingly tolerant of surface charge and hydrophobicity changes. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1863*(12), 3116–3124. https://doi.org/10.1016/j.bbamcr.2016.09.006

Marion, D., Ikura, M., & Bax, A. (1989). Improved solvent suppression in one- and two-dimensional NMR spectra by convolution of time-domain data. *Journal of Magnetic Resonance (1969)*, *84*(2), 425–430. https://doi.org/10.1016/0022-2364(89)90391-0

Matos, C. F. R. O., Robinson, C., Alanen, H. I., Prus, P., Uchida, Y., Ruddock, L. W., … Keshavarz-Moore, E. (2014). Efficient export of prefolded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in Escherichia coli CyDisCo strains. *Biotechnology Progress*, *30*(2), 281–290. https://doi.org/10.1002/btpr.1858

Matos, C. F. R. O., Robinson, C., & Di Cola, A. (2008). The Tat system proofreads FeS protein substrates and directly initiates the disposal of rejected molecules. *EMBO Journal*, *27*(15), 2055–2063. https://doi.org/10.1038/emboj.2008.132

Maurer, C., Panahandeh, S., Moser, M., & Müller, M. (2009). Impairment of twin-arginine-dependent export by seemingly small alterations of substrate conformation. *FEBS Letters*, *583*(17), 2849–2853. https://doi.org/10.1016/j.febslet.2009.07.038

Richter, S., Lindenstrauss, U., Lücke, C., Bayliss, R., & Brüser, T. (2007). Functional tat transport of unstructured, small, hydrophilic proteins. *Journal of Biological Chemistry*, *282*(46), 33257–33264. https://doi.org/10.1074/jbc.M703303200

Robinson, C., Matos, C. F. R. O., Beck, D., Ren, C., Lawrence, J., Vasisht, N., & Mendel, S. (2011). Transport and proofreading of proteins by the twin-arginine translocation (Tat) system in bacteria. *Biochimica et Biophysica Acta - Biomembranes*, *1808*(3), 876–884. https://doi.org/10.1016/j.bbamem.2010.11.023

Stanley, N. R., Palmer, T., & Berks, B. C. (2000). The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in Escherichia coli. *Journal of Biological Chemistry*, *275*(16), 11591–11596. https://doi.org/10.1074/jbc.275.16.11591

Sutherland, G., Grayson, K., Mermans, D., Jones, A., Roberston, A., Auman, D., … Hunter, C. N. (2018). Probing the quality control mechanism of the Escherichia coli twin-arginine translocase using folding variants of a de novo-designed heme protein. *Jbc*, (1). https://doi.org/10.1074/jbc.RA117.000880

Table 1

Table showing plasmids used to express BT6 variants for this study

Name	Reference
TorA-BT6	(G. Sutherland *et al.*, 2018)
TorA-BT6M1 (H53A)	(G. Sutherland *et al.*, 2018)
TorA-BT6M0 (H53A, H88A)	(G. Sutherland *et al.*, 2018)
TorA-BT6-2Asp (K5D, K13D)	This study
TorA-BT6-4Asp (K5D, K13D, K75D, K83D)	This study
TorA-BT6-2Lys (D9K, D23K)	This study
TorA-BT6-4Lys (D9K, D23K, D79K, D93K)	This study
TorA-BT6-3NLeu (D9L, K13L, E16L)	This study
TorA-BT6-C-term26 (Additional–SNVIIIIITNKDPNSSSVDLAAALE)	This study
TorA-BT6-C-term13 (Additional – SNVIIIIITNKDP)	This study
TorA-BT6-C-term6 (Additional – SNVIII)	This study

Figure legends

Figure 1. Secondary and tertiary structure of the two-heme binding maquette BT6 protein. (A) Ribbon diagram of the overall structure of the two-heme binding maquette BT6 protein shows four identical α-helix loops (green). The loops accommodate four histidine residues which ligate two heme *b* molecules (red). (B) The electrostatic potential map of the BT6 outer surface shows positively charged (red), negatively charged (blue) and uncharged (grey) residues. (C) Hydrophilic (purple) and hydrophobic (yellow) outer surface residues are shown in the hydrophobicity map.

Figure 2. Changes in substrate surface electrostatic potential of BT6 maquette (WT) are tolerated by the Tat system. (A). Representation of the altered outer surface residues of the BT6 maquette variants: BT6-2Lys, BT6-4Lys, BT6-2Asp and BT6-4Asp. The substituted residues (negative = red and blue = positive) are shown on the ribbon diagram. Two of the variants had 2 or 4 Asp residues substituted by 2 and 4 Lys residues (denoted BT6-2Lys and BT6-4Lys). In contrast, the 2 or 4 Lys residues of the other variants were substituted by 2 or 4 Asp residues (denoted BT6-2Asp and BT6-4Asp). (B). The four variants, bearing TorA signal peptides, were expressed in *E. coli* and cells were fractionated into cytoplasm, membrane and periplasm (C, M, P). The samples were analysed by SDS-PAGE and immunoblotting using antibodies to the 6-His tag on the Ctermini of the proteins. (C). The BT6 variants (BT6-2Lys, BT6-4Lys and BT6-2Asp) were purified from the periplasmic fraction by immobilized metal

affinity chromatography (IMAC). The intense red color indicates that BT6-2Lys and BT6-4Lys accommodate two hemes and are fully folded. BT6-2Asp has an orange color, which indicates that this BT6 variant accommodates only one heme.

Figure 3. UV-visible absorbance spectra of purified BT6-2Asp confirms the presence of a single heme. The absorbance spectra of BT6-2Asp (orange) compared to the previously published absorbance spectra of BT6 (black), BT6M1 (red), TorA-BT6M0 (blue)(G. A. Sutherland et al., 2018). There is high similarity between the spectrum of BT6M1 (one heme binder) and BT6-2Asp, indicating that BT6-2Asp accommodates one molecule of heme.

Figure 4. BT6-2Asp is exported far more efficiently than BT6M1, and slightly more efficiently than BT6. TorA-BT6, TorA-BT6M1 and TorA-BT6-2Asp were expressed in *E. coli* as detailed in Figure 2, and the periplasmic fraction was prepared after 60, 120, and 300 min induction with IPTG). Samples were analysed by SDS-PAGE and immunoblotting.

Figure 5. 1D-1H-NMR spectra suggest that folding of a restricted domain in BT6-2Asp enhances export efficiency. 1D 1H NMR spectra of BT6, BT6M1, TorA-BT6M0, and BT6-2Asp measured at 298K and 600MHz in 90% H_2O /10% D_2O using a DPFGSE water suppressed pulse sequence. Panel A: comparison of the spectra for BT6 and BT6-2Asp, and BT6-2Asp and BT6M1; the peak

marked with a * is presumed to be from a small molecule contaminant. Panel B: blow-up of the region between 9.5 and 12.5 ppm for each of the 1D spectra.

Figure 6. An addition of an unfolded domain to folded BT6 maquette (WT) is rejected by the Tat translocase. A: 3 variants of TorA-BT6 were expressed expressed in W3110 cells, containing C-terminal extensions of 6, 13 and 26 residues (termed BT6-Cterm6, BT6-Ctermm13 and BT6-Cterm26). Expression was induced with IPTG for 3 hours and cells were then fractionated to yield cytoplasmic (C), membrane (M) and periplasmic (P) fractions. The fractions were immunoblotted using a C-terminal His antibody. Molecular mass markers (kDa) are shown on the left side of the blot. B: cytoplasmic extracts were prepared from cells expressing the 3 variants, and were subjected to IMAC chromatography. The eluates were concentrated and absorbance spectra were obtained as in Fig. 3.

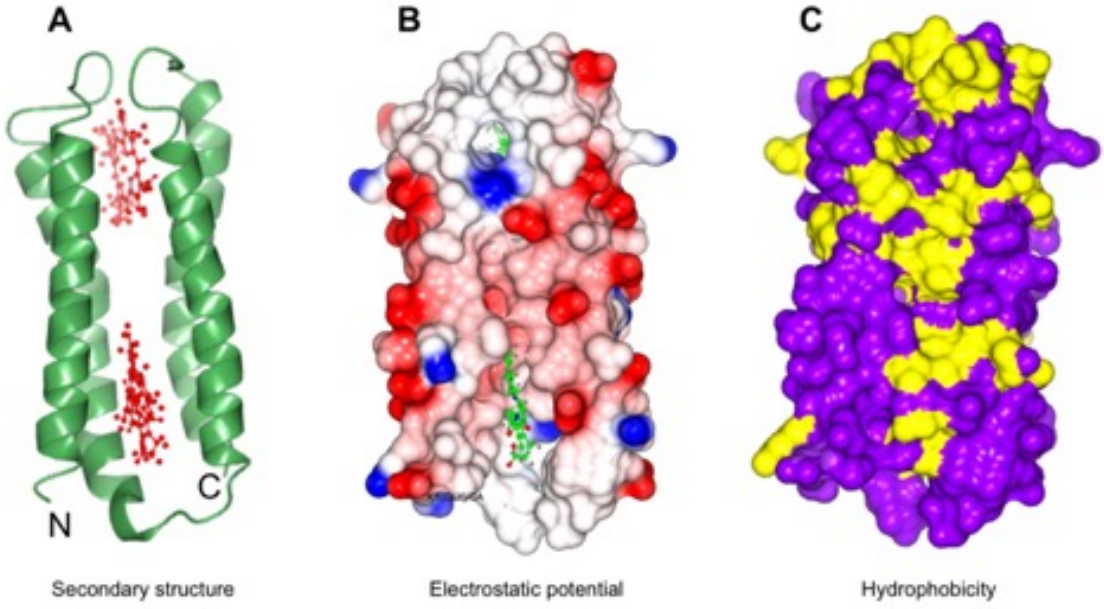


Figure 1

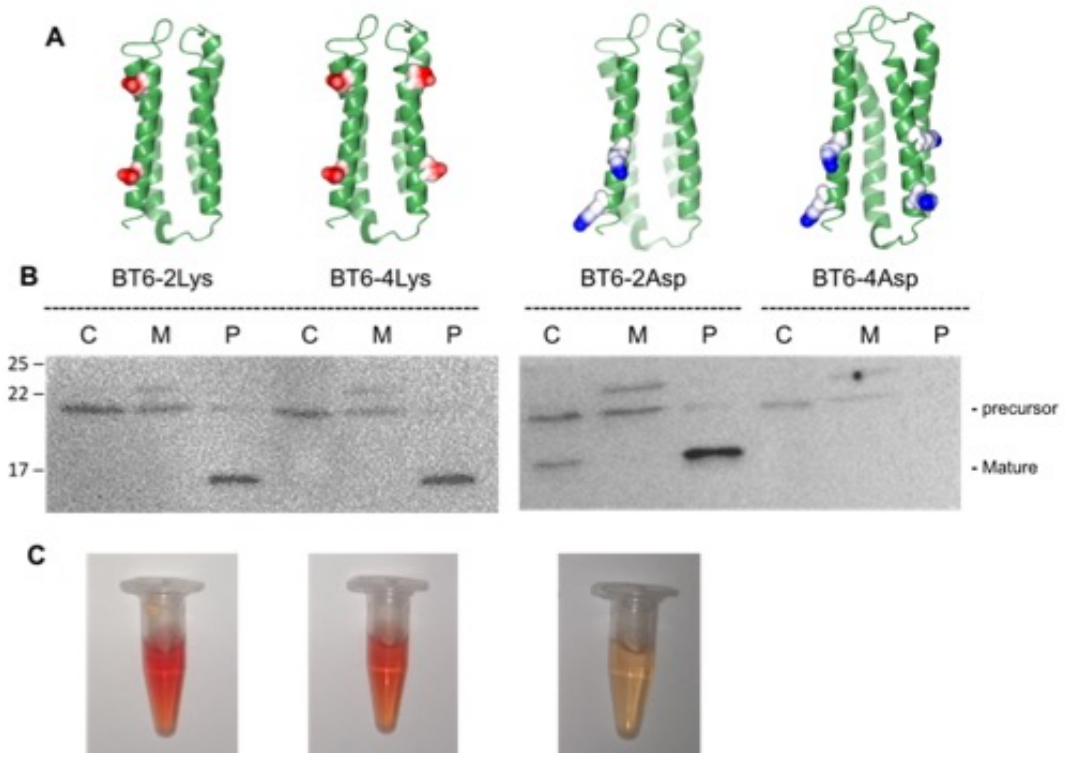


Figure 2

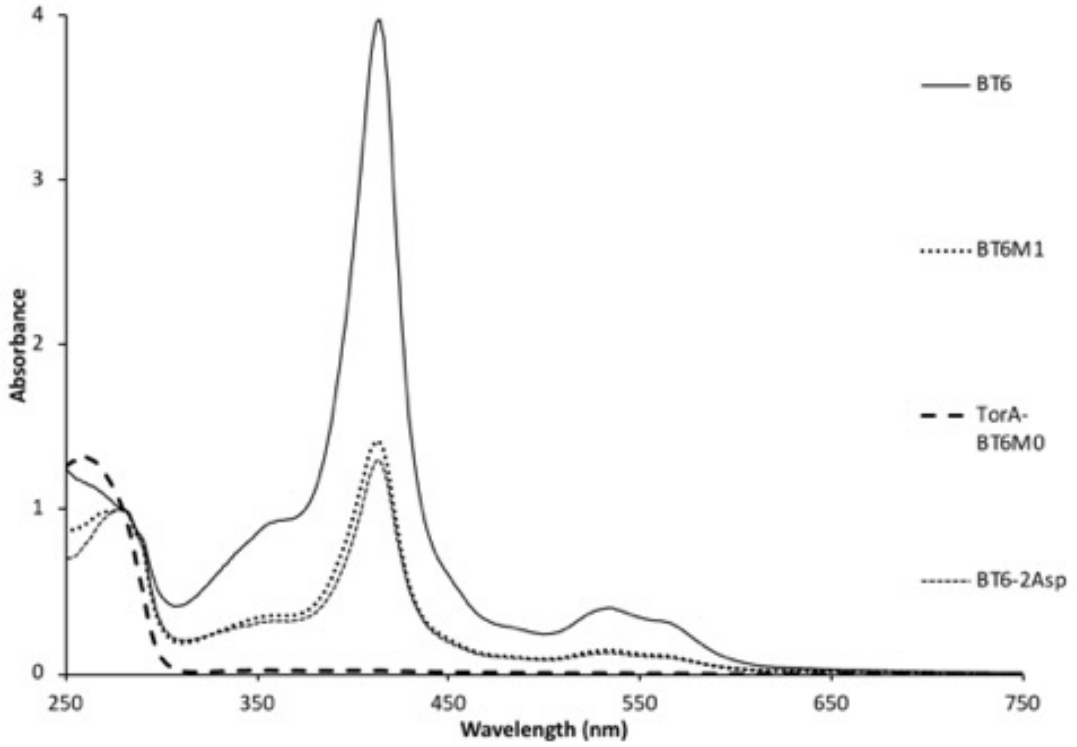


Figure 3

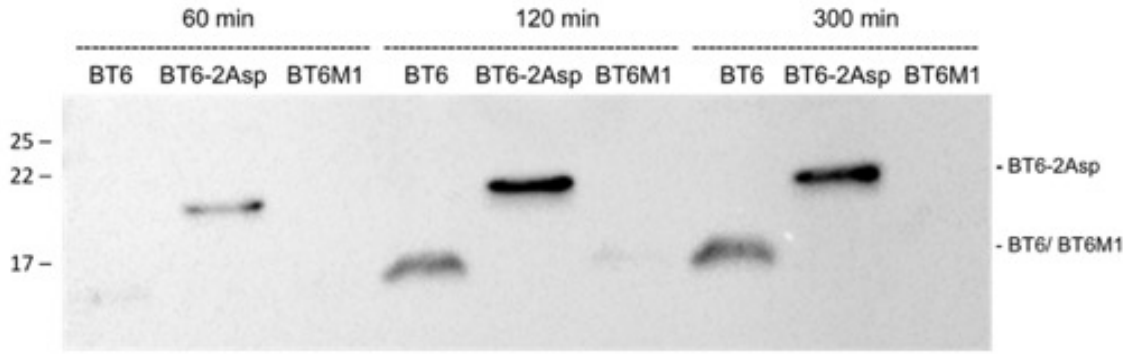


Figure 4

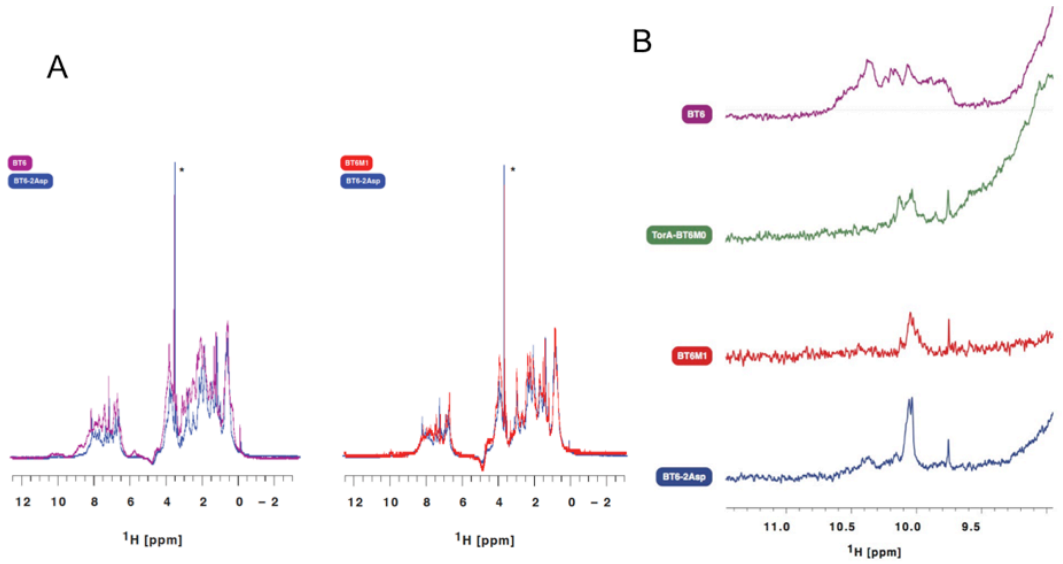
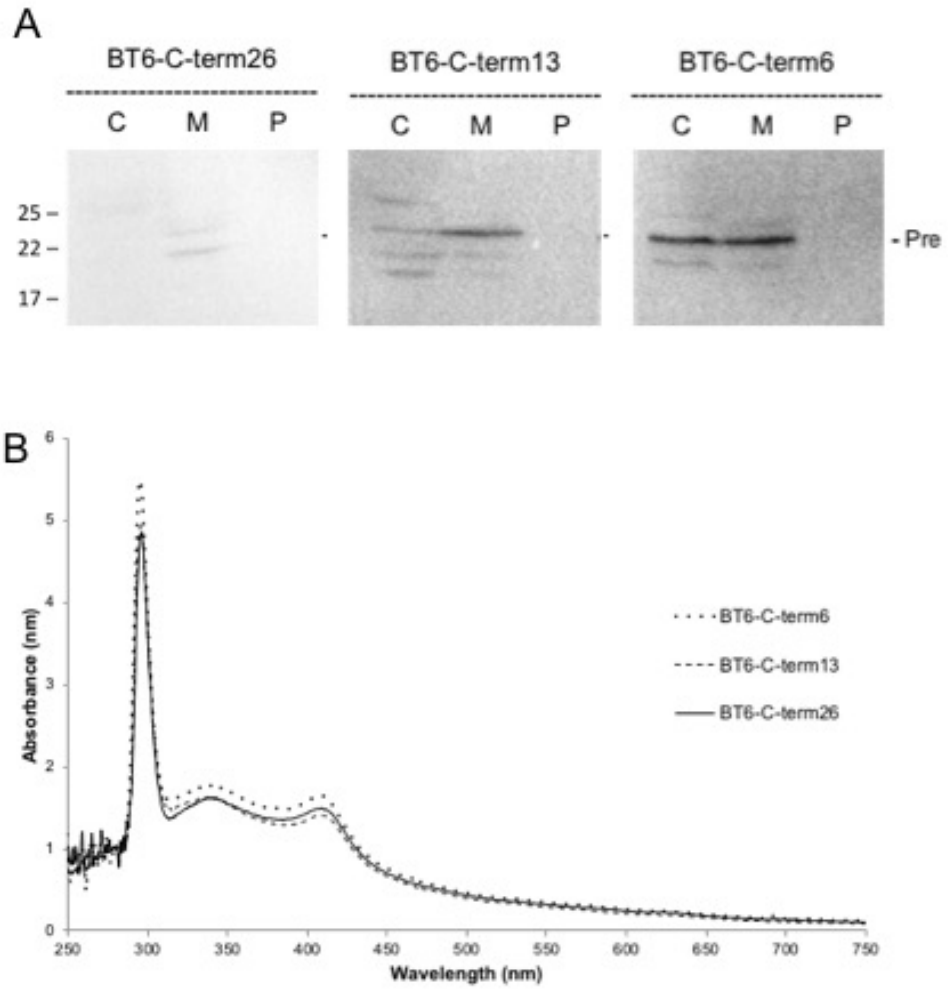

A
B
BT6
BT6-2Asp
BT6M1
TorA-BT6M0
12 10 8 6 4 2 0 −2
^{1}H [ppm]
11.0 10.5 10.0 9.5

Figure 5


A
BT6-C-term26
BT6-C-term13
BT6-C-term6
C M P
25 –
22 –
17 –
Pre
B
Absorbance (nm)
Wavelength (nm)
BT6-C-term6
BT6-C-term13
BT6-C-term26

JBC ARTICLE

Author's Choice

Probing the quality control mechanism of the *Escherichia coli* twin-arginine translocase with folding variants of a *de novo*–designed heme protein

Received for publication, November 9, 2017, and in revised form, March 15, 2018 Published, Papers in Press, March 20, 2018, DOI 10.1074/jbc.RA117.000880

George A. Sutherland[‡1], Katie J. Grayson[‡2], Nathan B. P. Adams[‡3], Daphne M. J. Mermans[§4], Alexander S. Jones[§5], Angus J. Robertson[‡1], Dirk B. Auman[¶6], Amanda A. Brindley[‡3], Fabio Sterpone[‖], Pierre Tuffery[**], Philippe Derreumaux[‖], P. Leslie Dutton[¶7], Colin Robinson[§4], Andrew Hitchcock[‡3], and C. Neil Hunter[‡3,8]

From the [‡]Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN, United Kingdom, the [§]School of Biosciences, University of Kent, Canterbury CT2 7NJ, United Kingdom, the [¶]Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, the [‖]Laboratoire de Biochimie Théorique, UPR 9080 CNRS, Université Paris Diderot, Sorbonne Paris Cité, 75005 Paris, France, and [**]INSERM U973, Université Paris Diderot, Sorbonne Paris Cité, 75013 Paris, France

Edited by Chris Whitfield

Protein transport across the cytoplasmic membrane of bacterial cells is mediated by either the general secretion (Sec) system or the twin-arginine translocase (Tat). The Tat machinery exports folded and cofactor-containing proteins from the cytoplasm to the periplasm by using the transmembrane proton motive force as a source of energy. The Tat apparatus apparently senses the folded state of its protein substrates, a quality-control mechanism that prevents premature export of nascent unfolded or misfolded polypeptides, but its mechanistic basis has not yet been determined. Here, we investigated the innate ability of the model *Escherichia coli* Tat system to recognize and translocate *de novo*–designed protein substrates with experimentally determined differences in the extent of folding. Water-soluble, four-helix bundle maquette proteins were engineered to bind two, one, or no heme *b* cofactors, resulting in a concomitant reduction in the extent of their folding, assessed with temperature-dependent CD spectroscopy and one-dimensional ^{1}H NMR spectroscopy. Fusion of the archetypal N-terminal Tat signal peptide of the *E. coli* trimethylamine-*N*-oxide (TMAO) reductase (TorA) to the N terminus of the protein maquettes was sufficient for the Tat system to recognize them as substrates. The clear correlation between the level of Tat-dependent export and the degree of heme *b*–induced folding of the maquette protein suggested that the membrane-bound Tat machinery can sense the extent of folding and conformational flexibility of its substrates. We propose that these artificial proteins are ideal substrates for future investigations of the Tat system's quality-control mechanism.

The transport of proteins across membranes is one of the great challenges faced by the cell. In prokaryotes, two major pathways are used to achieve protein translocation across the cytoplasmic (inner) membrane. The general secretion (Sec)[9] pathway transports proteins in an unfolded configuration using energy provided by both ATP hydrolysis and the transmembrane proton gradient (1). In contrast, the twin-arginine translocase (Tat) system transports fully folded proteins (2, 3) and is energized solely by the transmembrane proton gradient.

The majority of Tat substrates are cofactor-containing proteins that require assembly in the cytoplasm (4–6) including those that fold too quickly for Sec transport (7) and those that assemble into oligomeric complexes (8). Proteins translocated by the Tat pathway have an N-terminal signal sequence characterized by a twin-arginine (RR) motif (9); the signal sequence is cleaved from the precursor protein during or immediately after translocation, releasing the mature protein into the periplasm.

How components of the Tat machinery assess the folding state of a protein substrate remains poorly understood (6, 10, 11). A quality control or proofreading mechanism could exist to prevent futile export of misfolded or misassembled proteins (12–14). In *Escherichia coli* the Tat apparatus comprises TatA/B family proteins and the TatC protein. Single point mutations in either TatA or TatC were identified that allow

The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

This article contains Tables S1 and S2 and Figs. S1–S7.

[1] Supported by Faculty of Science Studentships from The University of Sheffield.

[2] Supported by a doctoral studentship from the Engineering and Physical Sciences Research Council (UK). Present address: School of Biochemistry, University of Bristol, Bristol BS8 1TD, United Kingdom.

[3] Supported by Research Grant BB/M000265/1 from the Biotechnology and Biological Sciences Research Council (UK).

[4] Supported by the Marie Curie Initial Training Network Grant (Horizon 2020, ProteinFactory) 642863.

[5] Supported by a University of Kent Graduate Training Award studentship.

[6] Supported by National Institutes of Health Graduate Fellowship T32 GM008275, the Structural Biology & Molecular Biophysics Training Program.

[7] Supported by French "DYNAMO" Grant ANR-11-LABX-0011-01 for sabbatical year 2015–2016 at UPR9080 CNRS, IBPC.

[8] Supported in part by Advanced Award 338895 from the European Research Council. To whom correspondence should be addressed. Tel.: 44-114-222-4191; Fax: 44-114-222-2711; E-mail: c.n.hunter@sheffield.ac.uk.

[9] The abbreviations used are: Sec, general secretion; Tat, twin-arginine translocase; IPTG, isopropyl β-D-1-thiogalactopyranoside; BisTris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol (systematic).

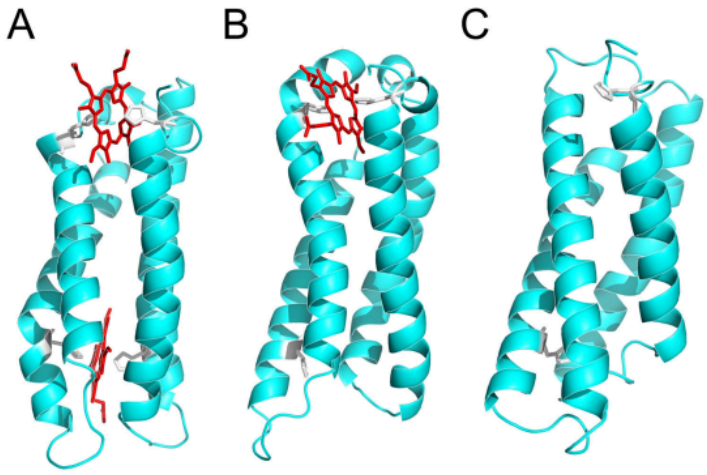


Figure 1. Structural models of the BT6 maquette proteins used in this study. *A,* BT6 (see Ref. 28) coordinates two heme *b* molecules (*red*) using four histidine ligands (*white*). *B,* in BT6M1 the H53A substitution means the protein can only coordinate one heme *b*. *C,* in BT6M0 the double H53A/H88A substitution prevents heme binding. All images were taken from 50-ns trajectories.

translocation of unfolded substrates, and point mutations in both TatB and TatC enabled export of a broader range of unfolded substrates, suggesting that the TatABC subunits cooperatively assess the folding state of proteins independently of protein translocation (15).

Richter *et al.* (16) showed that small, unstructured hydrophilic FG repeat proteins could be exported by the Tat system, and that the presence of hydrophobic surface patches was sufficient to abort transport, raising the possibility that the Tat system screens proteins based on their surface hydrophobicity. It has been reported that the length of the unstructured FG repeat polypeptide dramatically affects Tat export, with longer regions abolishing Tat export altogether (17). Conversely, Jones *et al.* (18) recently reported that the Tat system was surprisingly tolerant of hydrophobic patches on the surface of structured single-chain variable fragment proteins, and export efficiency was increased with greater structural rigidity. Chaperones may also prevent export of a protein until cofactor insertion has taken place (19–21), and mutants incapable of cofactor binding are rapidly degraded once in contact with the Tat machinery (22).

To further investigate the Tat quality-control mechanism, we used maquettes, which are simple, repetitive protein structures designed *de novo* from first principles with minimal reference to natural protein structures (23–26). As such structures contain unnecessary complexity, accumulated from perpetual rounds of blind natural selection (27), altering protein residues can have unpredictable effects on protein structure and dynamics. In contrast, the role of each amino acid in the simple maquette structure has been rationalized at the outset, so changes to structure and function become more predictable.

The maquettes used in this study are based on the BT6 maquette developed by Farid *et al.* (28), and consist of loops linking four largely identical α-helices enclosing a water-excluding cavity that can accommodate cofactors (Fig. 1*A–C*). Histidine residues within the maquette cavity ligate hemes, producing artificial proteins resembling *b*-type cytochromes and myoglobin (29, 30).

Here, three different maquette variants were utilized, each with a different heme *b* binding capacity (Fig. 1). Using nuclear magnetic resonance (NMR) and circular dichroism (CD) spectroscopy we show that binding two, one, or no hemes (Fig. 1, *A–C*, respectively) imparts changes in the extent of folding of the maquette variants. The archetypal trimethylamine-*N*-oxide reductase (TorA) Tat signal peptide was fused to the N terminus of the three proteins and the *E. coli* Tat machinery was challenged to differentiate between the folding variants. We show that the Tat apparatus is able to sense the conformational flexibility of the different maquette substrates, and that increasingly well-folded maquettes are exported with enhanced efficiency.

Results

Production of heme-reconstituted maquettes

The aim of this study was to test whether the *E. coli* Tat system could recognize and export a *de novo*–designed di-heme protein, and then to use variants of this protein with experimentally confirmed differences in conformational flexibility to test if the Tat apparatus selectively processes the more folded proteins.

As described above and depicted in the computationally-generated structures presented in Fig. 1, three variants of the BT6 maquette were produced, facilitating the incorporation of two (BT6), one (BT6M1), or no (BT6M0) heme *b* cofactors. The requirement for bis-histidine ligation of heme *b* in the artificial constructs enabled the generation of the one and no heme-binding variants through H53A, and H53A/H88A (histidine to

Table 1
Amino acid sequences of maquettes and signal peptides used in this study

Name	Sequence[a]	Details
BT6	MGGDGENLYFQG EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**H**EDALQKFEEALNQFEDLK	Di-heme binding
BT6M1	MGGDGENLYFQG EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQAEDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**H**EDALQKFEEALNQFEDLK	Single heme binding due to removal of 1 coordinating histidine residues
BT6M0	MGGDGENLYFQG EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQAEDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQAEDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**H**EDALQKFEEALNQFEDLK	No heme binding due to removal of 2 coordinating histidine residues
TorA	MNNNDLFQAS***RR***RFLAQLGGLTVAGMLGPSLLTPRRATAAQAA	TorA signal peptide for periplasmic localization by the Tat system
TorA R12/13K	MNNNDLFQAS***KK***RFLAQLGGLTVAGMLGPSLLTPRRATAAQAA	R12K/R13K TorA signal peptide
PelB	MKYLLPTAAAGLLLLAAQPAMA	PelB signal peptide for periplasmic localization by Sec system

[a] Heme coordinating histidine residues are shown in bold. The twin-arginine motif of the Tat signal peptide is shown in bold italics.

alanine) point mutations, respectively (Table 1). The computational models in Fig. 1 extend the theoretical structure of the BT6 maquette outlined by Farid *et al.* (28). They are included to illustrate the apo- and heme-maquette designs used as Tat substrates, and are not intended as a replacement for atomically-accurate experimental structures.

Maquettes are typically overproduced in *E. coli* in large amounts, with bound heme largely absent following purification (see "Experimental procedures") because the native tetrapyrrole biosynthesis pathway cannot keep pace with the induced synthesis of maquette (Fig. S1*A*). Apoproteins were reconstituted *in vitro* with an excess of heme and unbound pigment was removed by anion exchange chromatography. The Soret absorption bands for the BT6, BT6M1, and BT6M0 maquettes were normalized for protein concentration (absorbance at 280 nm) and had maxima at 413, 412, and 396 nm, respectively (Fig. 2*A*), with the amplitudes corresponding to binding two, one, and no hemes, respectively. The weak, blue-shifted absorption of the BT6M0 maquette suggested a low level of adventitiously bound heme. We compared heme absorption in the maquette-bound, and solvated (buffer or DMSO) states, which had respective maxima at 413, 404, and 384 nm (Fig. S1*B*). The blue-shifted absorption maxima for solvated heme are consistent with the weak, blue-shifted absorption for BT6M0 (Fig. 2*A*) arising from residual, weakly bound heme; ligation into the internal cavity of BT6 and BT6M1 causes a red-shift absorption of the heme.

Thermostability and folding of maquette variants

Temperature-dependent CD spectroscopy was performed to assess the effect of heme ligation on the thermal stability of each maquette scaffold. At 15 °C the CD spectra of all three maquette variants, with or without bound heme, were typical of α-helical structures (Fig. S2, *black lines*). Likewise, irrespective of heme binding heating to 80 °C resulted in spectra typical of that of a denatured protein (31) (Fig. S2, from *black* to *red* to *gray lines*).

Ellipticity at 222 nm was monitored during temperature cycling from 15 to 80 to 15 °C to observe denaturing and subsequent re-folding of protein (Fig. 2, *B–D*). For BT6, heme incorporation displaced the midpoint of the melting curve from 22 to 56 °C, the 34 °C difference indicating increased stabilization following heme binding. The same procedure with BT6M1 increased stabilization by only 16 °C (25 to 41 °C), and had very little effect on the BT6M0 scaffold (32 to 35 °C) (Table 2). Thus, bis-histidine ligation of two hemes within the maquette cavity of BT6 significantly stabilizes the four-helix bundle structure, and there is a smaller effect with only one bis-histidine ligation in BT6M1. In the case of BT6M0, where heme ligation is not possible, there is correspondingly no significant stabilization of maquette structure in the presence of heme.

One-dimensional proton (^{1}H) NMR spectroscopy was used to assess conformational changes in tertiary structure upon heme binding to maquettes. A ^{1}H NMR spectrum characteristic of a protein with limited tertiary structure was observed for all maquette scaffolds in the absence of heme (Fig. 3*A*, Fig. S3). Addition of heme to BT6 showed greater dispersion of resonances in the amide proton region, with a notable increase in the number of peaks at around 10 ppm, suggesting that the heme is binding to the scaffold and stabilizing the protein tertiary structure (Fig. 3*B*). BT6M1 showed a small increase in resonance dispersion following heme addition (Fig. 3*C*), whereas BT6M0 showed no significant change in amide proton dispersion (Fig. 3*D*). The increase in the chemical shift dispersion observed in the methyl proton region (at around 1.0–0.0 ppm) for BT6, BT6M1, and BT6M0 mirrored the behavior observed for the amide proton resonances (Fig. S4). Together, the increase in ^{1}H NMR dispersion in both backbone amide and methyl regions confirm the ligation of heme into the BT6 and BT6M1 variants and indicate an increase in protein folding upon ligation. These results are consistent with a previous NMR study that showed poor dispersion for a related apo-maquette, and progressive structuring of the maquette when 1 then 2 eq of heme were added (28).

In vivo Tat export assays

The *E. coli* TorA signal peptide is sufficient to direct green fluorescent protein (GFP) to the *E. coli* periplasm via the Tat system (13). Constructs in which the sequence encoding the TorA signal peptide (residues 1–39) and first 4 amino acids of the mature TorA protein (residues 40–43) were added in-frame to the 5′ terminus of the genes encoding the BT6

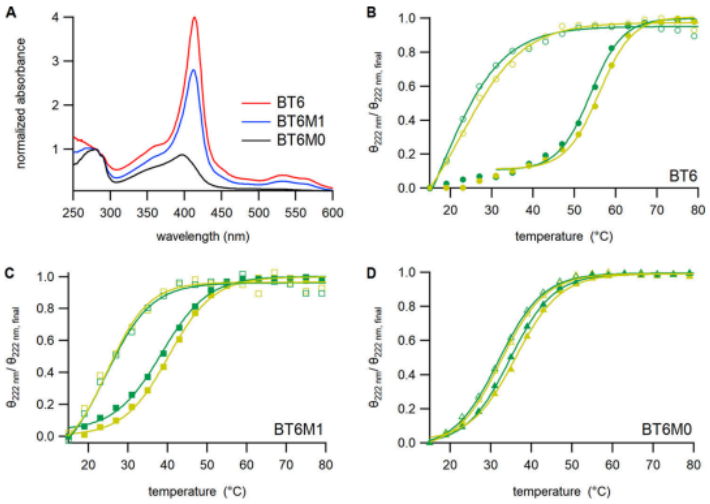


Figure 2. Spectroscopic analysis of apo- and heme-reconstituted maquette proteins. *A*, UV-visible absorption spectra of heme-reconstituted maquette samples normalized to absorbance at 280 nm. *B–D*, normalized melting (*yellow*) and refolding (*green*) of secondary structure measured as the reduction in CD ellipticity at 222 nm across a temperature gradient in the absence (*open markers*) or presence (*solid markers*) of heme. Ellipticity was recorded every 1 °C, but only every third data point is shown for clarity. *Lines* are theoretical and described by a Boltzmann distribution (see "Experimental procedures"). Melting (T_m) and refolding (T_f) temperatures are reported in Table 2.

Table 2
Melting temperatures (T_m) and folding temperatures (T_f) of maquettes in the absence (apo) or presence of heme

Maquette	Apo		+Heme	
	T_m[a]	T_f	T_m	T_f
	°C		°C	
BT6	22	17	56	54
BT6M1	25	24	41	39
BT6M0	32	32	35	36

[a] The T_m and T_f (temperature at which 50% of the protein is unfolded or folded, respectively) values were determined from Boltzman distribution fits as shown in Fig. 2, *B–D*.

maquette variants followed by a C-terminal His_6 tag (Fig. S5) were synthesized and cloned into the pEXT22 vector (Table S1). In the TorA–BT6 variants residues His-95 and His-130 correspond to His-53 and His-88 in the maquette scaffolds presented in Fig. 1.

The resulting plasmids were co-transformed into *E. coli* BL21(DE3) along with empty pET-21a(+). The pEXT22 vector was used as it allows isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible expression of the synthetic gene under the control of a tightly regulated *tac* promoter. The low copy number of pEXT22 (R100 origin of replication, 1–1.5 copies per chromosome) (32, 33) limits production of recombinant protein, to avoid overwhelming the *E. coli* Tat apparatus with substrate.

Following growth and protein production, periplasmic fractions were prepared from harvested cells as described under "Experimental procedures" and analyzed alongside cell-free extracts prepared from identically grown cells. The C-terminal His_6 tag allows immunodetection of the recombinant protein; the predicted molecular mass of the unprocessed precursor protein is ~21 kDa, whereas the Tat-processed mature protein is ~17 kDa. Anti-GroEL was used to probe the degree of cytoplasmic contamination in the periplasmic preparations (Fig. 4*A*); for all samples only a faint signal was detectable in the periplasm compared with that detected in cell-free extracts, indicating the level of contamination was very low. The Tat export assays were performed in cells that also maintained the pET-21a(+) vector, allowing immunoblot signals to be normalized by immunodetection of β-lactamase (Fig. 4*B*), allowing direct comparison of the degree of Tat transport between samples.

For TorA–BT6, a 17-kDa signal was present for both periplasmic and cell-free extract (Fig. 4*C*) preparations with no signal detected at 21 kDa, indicating complete processing and export of TorA–BT6. To confirm the periplasmic localization was Tat-dependent, a R12K/R13K (KK) variant of the TorA signal peptide was generated and fused to BT6 (Table 1); it has previously been demonstrated that the KK motif completely abolishes Tat-transport (12, 34). When this construct was tested, a 21-kDa signal corresponding to the unprocessed apoprotein is observed in blots of total lysates and no processed protein was detectable in the periplasm (Fig. 4*C*). Together these data indicate that the BT6 scaffold is able to acquire heme from the native biosynthetic pathway and fold *in vivo*, and that this folding event can be detected by the Tat system, with the TorA signal peptide sufficient for Tat-mediated recognition and transport of heme-loaded BT6 to the *E. coli* periplasm.

To probe the proofreading ability of the Tat machinery to recognize and efficiently export only folded proteins, similar constructs were generated in which the TorA signal peptide was fused to the BT6M1 and BT6M0 maquettes, which, in the

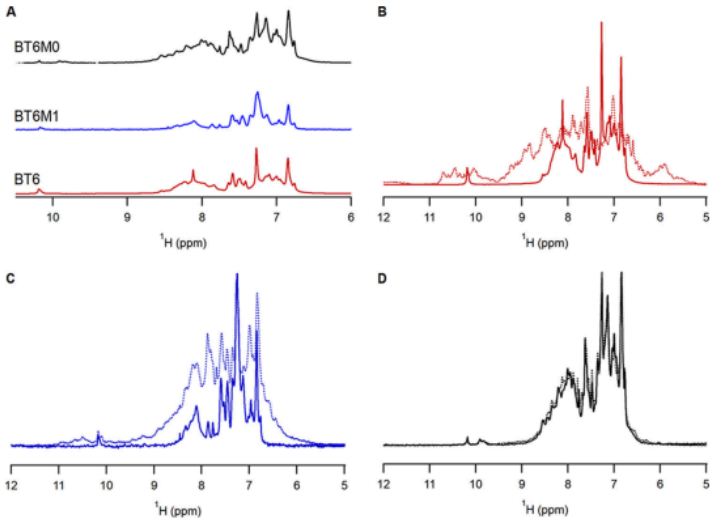


Figure 3. Proton NMR resonances of maquettes with and without heme cofactor. *A–D*, the amide proton region for apo-BT6 (*red*), apo-BT6M1 (*blue*), and apo-BT6M0 (*black*) (*A*), and comparisons in the absence (*solid line*) and presence (*dashed line*) of heme for BT6 (*B*), BT6M1 (*C*), and BT6M0 (*D*). For BT6 and BT6M1, the presence of heme induces changes in the amide proton resonance dispersion, whereas for BT6M0 negligible changes are observed. These chemical shift changes are consistent with a heme binding event coupled with a change in protein conformation.

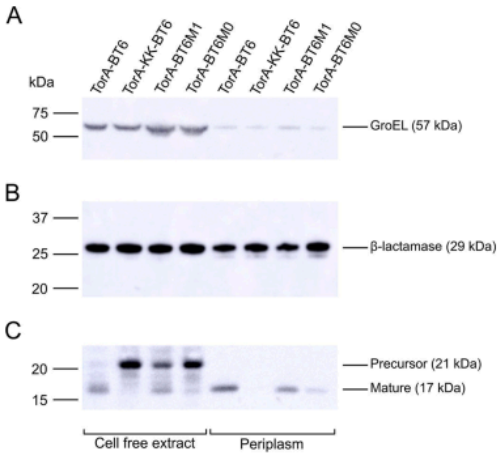


Figure 4. Differential export of BT6 variants by the Tat system determined by immunoblotting. *A*, GroEL was used as a cytoplasmic marker to confirm only very minor cytoplasmic contamination of periplasmic preparations. *B*, β-lactamase was used to confirm equal loadings of cell-free extracts or periplasm samples. *C*, the C-terminal His tag on the BT6 maquettes was used to determine the degree of Tat-dependent periplasmic localization and the level of the unprocessed precursor proteins in cell-free extracts. For all panels the positions of molecular weight markers and the expected size of proteins are indicated alongside the blots. Each blot is representative of at least 3 independent experiments.

presence of heme, are either partially folded (BT6M1) or unfolded (BT6M0) relative to BT6 (Figs. 2 and 3). Compared with BT6, considerably less protein was found in the periplasm for BT6M1, and almost none for BT6M0 (Fig. 4*C*). Conversely, blots of cell-free extracts of the same strains reveal a significant signal for the unprocessed maquettes for BT6M0, and to a lesser extent BT6M1, as observed for the TorA–KK–BT6 control, showing that the preprotein accumulates in the cytoplasm when it is less well folded (Fig. 4*B*).

The cellular levels of maquette variants should be very similar as all the proteins are identical apart from 1 or 2 point mutations and were expressed from the same plasmid in the same cell line under the same conditions. Indeed, the data presented in Fig. 4*C* shows that the total signal (preprotein plus exported protein where relevant) for all three maquettes is similar. However, to further confirm that the differential periplasmic targeting of the three BT6 variants by the Tat system was due to differences in their heme loading and associated folding and not in expression, synthesis, or stability, we performed two additional control experiments. First we produced each maquette under the same conditions as for the Tat export assays but without the TorA signal peptide, and immunoblotting confirmed approximately equal levels of each (Fig. S6*A*), ruling out that BT6M1 and BT6M0 are produced at lower levels or are more rapidly degraded in the cytoplasm than BT6.

We also performed Sec system-mediated export assays of BT6, BT6M1, and BT6M0. As the Sec system transports unfolded proteins across the cytoplasmic membrane, differences in folding, mediated in the case of BT6 by heme binding, will not affect transport. Maquette genes were cloned into the pET-22b(+) vector (Table S1) in-frame with an N-terminal PelB (*Erwinia carotovora* pectate lyase B) Sec leader peptide (22 amino acids) (35, 36) (Table 1). Stop codons were omitted so

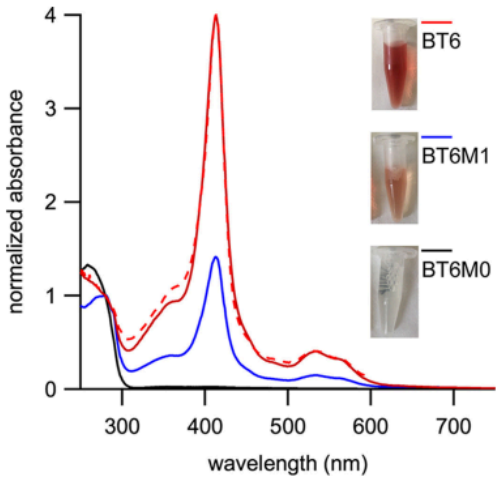


Figure 5. Spectroscopic analysis of purified maquette proteins. UV-visible absorption spectra of maquettes purified from the periplasm of *E. coli* (BT6, *red line*; BT6M1, *blue line*; BT6M0, *black line*) compared with BT6 reconstituted with heme *in vitro* (*red dashed line*). Spectra are normalized to absorbance at 280 nm. *Inset panels* show the purified proteins. See Fig. S7 for SDS-PAGE analysis of the purified maquettes.

proteins had a C-terminal His_6 tag for immunodetection; the unprocessed proteins have a predicted molecular mass of ~19 kDa and the mature maquettes would be ~17 kDa following cleavage of the PelB signal sequence. Fig. S6*B* shows that normalized loadings of cell lysates yield single, similarly intense signals for BT6, BT6M1, and BT6M0, indicating all three apo-maquettes are equally synthesized and transported to the periplasm when directed through the Sec system.

Purification of maquettes from the periplasm

To demonstrate that heme ligation into BT6 was responsible for its Tat-dependent localization in the periplasm, production of His-tagged maquettes was performed on a larger scale using the recently described *E. coli* W3110-TatExpress cell line (37). Cells were fractionated and maquette proteins were purified from the periplasmic fraction by immobilized metal affinity chromatography. Eluates were concentrated and striking differences in pigmentation between the maquette variants were observed, with a deep red color for BT6, weaker pigmentation for BT6M1, and no visible color for BT6M0 (Fig. 5, *inset panels*). SDS-PAGE of the concentrated eluates revealed the same pattern as the Tat exports presented in Fig. 4, with more BT6 than BT6M1 and only a trace amount of BT6M0 (Fig. S7). This also ruled out any leakage of unprocessed protein across the cytoplasmic membrane as the larger precursor species were not purified (Fig. S7). The absorbance spectra of Tat-exported BT6, BT6M1, and BT6M0 samples normalized at 280 nm are shown in Fig. 5, along with *in vitro* heme-reconstituted BT6. There is a close match between the spectra of the Tat-exported heme-loaded BT6 purified from the periplasm and the reconstituted BT6 sample (Fig. 5, *solid red versus dashed red lines*). The high affinity of BT6 for heme (28) ensures full occupancy of the two heme-binding sites, so the Tat-transported BT6 purified from the periplasm is also fully heme bound. These data, alongside the export assays shown in Fig. 4, reveal that the TorA–BT6 maquette is able to sequester heme *b* from the native *E. coli* biosynthetic pathway, correctly fold around its substrate, and retain the pigment following export by the Tat-apparatus.

Discussion

The Tat pathway is present in the cytoplasmic membranes of most prokaryotic organisms and is evolutionarily conserved in the thylakoids of plant chloroplasts and some mitochondrial membranes (reviewed in Refs. 3 and 7). As well as a biosynthetic role in numerous important cellular processes, the Tat pathway is required for colonization and virulence of globally significant human pathogens (38, 39). The Tat system differs from the universally conserved Sec pathway, which translocates unstructured polypeptides, as it transports folded, typically cofactor-containing proteins. How the Tat apparatus is able to determine that a protein is folded, cofactor loaded, and suitable for export is not understood.

Here, we have used artificial heme maquette proteins to investigate the ability of the archetypal *E. coli* Tat machinery to discriminate between folding states of protein substrates. The maquettes used in this study are four-helix bundle proteins enclosing a hydrophobic cavity that can accommodate heme cofactors ligated by histidine residues. Point mutations to the BT6 maquette, affecting only single histidine ligands, generated the BT6M1 and BT6M0 maquettes, and were sufficient to reduce the number of bound hemes, verified by absorption spectroscopy (Fig. 2*a*). The number of bound hemes therefore provides a simple, well-defined method to alter the stability of a protein substrate. The consequent effects of heme content on the thermostability and folding of each maquette were measured using temperature-dependent CD spectroscopy (Fig. 2*b*, Table 2) and ^{1}H NMR (Fig. 3).

When directed for export through the Tat export machinery, there were clear differences in the extent of translocation for BT6, BT6M1, and BT6M0 maquettes (Fig. 4) that correlate with the *in vitro* stability and folding experiments for heme-bound constructs. Export efficiency decreased with increasing conformational flexibility; the unstructured BT6M0 maquette is largely rejected for export, whereas limited export was observed in the intermediate case of the single heme-binding BT6M1 maquette.

The export data shown would only be obtained if the BT6 maquettes were able to ligate heme, *in vivo*, as heme coordination is a pre-requisite for protein stability and folding. Evidence for *in vivo* heme ligation was shown with the purification of heme-containing maquettes from the periplasmic fraction of *E. coli* (Fig. 5), the spectra of which strongly resemble those of heme proteins generated *in vitro* (Fig. 2*A*).

Although the native *E. coli* Tat translocase will never have encountered the artificial protein substrates described here, our data show that it is able to recognize and distinguish between them, even to the extent of processing the intermediate state of BT6M1 differently from the folded BT6 and the least structured BT6M0. Thus, the Tat proofreading process must involve a generic form of discrimination; the correlation between the structural flexibility of the Tat substrate and its suitability for export

suggests that there are initial encounters between Tat components and the substrate at the membrane surface. Such interactions might sense flexible motions of the substrate that are transmitted to other components of the Tat machinery, preventing transport across the membrane.

This proof-of-concept study shows that an artificial protein, engineered to bind heme *b* from a native biosynthetic pathway, can be exported from the cell in its correctly folded state. Development of this concept may contribute a significant advance in biotechnology, where the principle could be applied to other organisms and to important biomolecules and protein-cofactor complexes, particularly those that may cause toxicity to the host.

Experimental procedures

Computational prediction and molecular dynamics refinement of protein structures

To generate illustrative structural models for BT6, the amino acid sequence was run through the PEP-FOLD structure prediction algorithm (40) for a total of 100 simulations, from which the five best-scoring structures were selected for further analysis. All five structures were four-helix bundles in agreement with the schematic structure described by Farid *et al.* (28). From these, a structure with both pairs of histidines positioned closest to their respective heme-ligating positions was selected as a starting structure. From this structure, mutants BT6M1 and BT6M0 were generated using PyMOL version 1.7 (PyMOL Molecular Graphics System Version 1.7, Schrödinger, LLC). Hemes (0, 1, or 2) were manually docked into the bis-His sites of the structures.

To relax the starting structure into an energy-minimized conformation, molecular dynamics (MD) pre-processing, production runs, and post-processing were performed with Gromacs version 4.6 (41). MD simulations were performed using the CHARMM27 force field for proteins and TIP3P-CHARMM model for water. For the heme cofactor, parameters were used for reduced, deprotonated, bis-His–ligated heme included in the CHARMM27 force field. Covalent bonds were explicitly specified between the heme iron and the ϵ-nitrogen of the relevant histidine residues. NaCl was added to a total concentration of 150 mM. Each structure was equilibrated for 100 ps in the NVT ensemble, followed by 100 ps in the NPT ensemble. Temperature and pressure/density plots (following NVT and NPT equilibrations, respectively) were checked for convergence before proceeding to production MD runs of 50 ns. Following production runs, trajectories were corrected for periodicity and centered on the protein. Snapshots of the trajectories at 50 ns were exported to .pdb files. Electrostatics and solvent-accessible surface calculations were performed using APBS (42). Visualization of calculations and structures was performed with the PyMOL Molecular Graphics System version 1.7.

DNA manipulation

Plasmids and primers used in this study are provided in Tables S1 and S2, respectively. Synthetic genes encoding maquette protein variants codon optimized for expression in *E. coli* were purchased from DNA2.0 (now ATUM) or Integrated DNA Technologies. Point mutations were generated using the QuikChange II Site-directed Mutagenesis Kit (Agilent). The TorA Tat signal peptide was amplified from *E. coli* genomic DNA and joined in-frame to maquette constructs by overlap extension-PCR using Q5 High-Fidelity DNA Polymerase (New England Biolabs). All plasmids were sequence verified by automated DNA sequencing (GATC Biotech). Competent *E. coli* JM109 (Promega) was used for cloning and was grown in Luria-Bertani (LB) broth/agar supplemented with the appropriate antibiotic(s) (Table S1).

Production and purification of untagged maquettes

The high copy number pJexpress414 plasmid was used for recombinant protein production under control of an IPTG-inducible T7 promoter. *E. coli* BL21(DE3) containing the desired plasmid was grown with shaking (230 rpm) at 37 °C in LB broth with 100 μg ml^{-1} ampicillin to an absorbance at 600 nm (A_{600}) of ~0.6. At this point IPTG was added to a final concentration of 1 mM to induce expression and the cultures were incubated for a further 16 h at 37 °C. Cells were harvested by centrifugation (4,400 × *g*, 15 min, 4 °C) and resuspended in buffer A (50 mM HEPES, pH 7.4, 500 mM NaCl, 5 mM imidazole). Cells were lysed by sonication on ice and the lysate was clarified by centrifugation (53,000 × *g*, 30 min, 4 °C). The supernatant was filtered through a 0.45-μm filter and applied to a Chelating Sepharose Fast Flow column (GE Healthcare) pre-equilibrated with 10 mg ml^{-1} of nickel sulfate. The column was washed with 5 column volumes of buffer A with the flow-through and wash was collected and pooled. The pooled sample was buffer exchanged into buffer B (50 mM HEPES, pH 7.4) and further purified by ion exchange chromatography on a Fast Flow Q-Sepharose column (GE Healthcare) with a linear gradient of 0–1 M NaCl in buffer B. Where required maquettes were further purified by size exclusion chromatography on a Superdex 200 Increase column (GE Healthcare) in buffer C (50 mM HEPES, pH 7.4, 200 mM NaCl). Where necessary protein was concentrated using Vivaspin centrifugal concentrators (Sartorius).

Heme reconstitution into apo-maquettes

Hemin (Sigma) stocks (1 mg ml^{-1}) were prepared in 100% DMSO. Protein concentrations were calculated by absorbance at 280 nm following the method described by Gill and von Hippell (43), and using the experimentally determined extinction coefficient of 32.6 mM^{-1} cm^{-1}. Reconstitutions were conducted with a 10-fold molar excess of hemin in buffer D (50 mM HEPES, pH 7.4, 200 mM NaCl, 20% (v/v) DMSO) and incubated for 45 min at 25 °C, before being transferred to ice and buffer exchanged into buffer B using Vivaspin centrifugal concentrators. Unbound cofactor was removed by ion exchange chromatography on a DEAE-Sepharose (Sigma) column.

UV-visible absorption spectroscopy

Protein samples were buffer exchanged into buffer E (5 mM sodium phosphate buffer, pH 7.4) and UV-visible absorption was measured in a 1-cm path length UV cuvette in a Cary 60 UV-visible spectrophotometer (Agilent) at room temperature.

CD spectroscopy

Mean residue ellipticity ($[\theta]_{MRW}$) of protein samples was measured in a 1-mm path length quartz cuvette on a Jasco J-810 spectropolarimeter with a Jasco PFD-425S Peltier to enable temperature control. Spectra were obtained from 15 to 80 °C at 5 °C intervals. Spectra were recorded continuously at a scan speed of 100 nm min^{-1}, with 1-nm resolution and a 4-s response with 4 accumulations. Ellipticity ($[\theta]$) at 222 nm was measured every 1 °C from 15 to 80 °C at 1 °C min^{-1} with a 4-s response. Melting temperatures (T_m) and refolding temperatures (T_f) are the temperature at which 50% of the protein is unfolded or folded, respectively, as determined by fitting melting data to a sigmoidal Boltzmann distribution according to the following equation, where T is the temperature, T_m is the melting (or folding) temperature, and T_0 is the initial temperature of the experiment.

$$[\theta]_{222nm} = [\theta]^{base}_{222nm} + \left(\frac{[\theta]^{max}_{222nm}}{1 + \exp\left(\frac{T_m - T_0}{\frac{\delta T}{\delta [\theta]}} \right)} \right) \quad \text{(Eq. 1)}$$

Proton NMR spectroscopy

Spectra were recorded at 298 K on 0.2–0.5 mM protein samples in buffer E, with the addition of 10% D_2O (spectrometer lock), and 1 mM trimethylsilyl propanate (reference standard). ^1H NMR spectra were recorded using a Bruker Avance 800 MHz spectrometer fitted with a 5-mm QXI room temperature probe, equipped with z axis gradients. One-dimensional experiments were acquired as accumulations of 4096 transients over a spectral width of 24.038 kHz, corresponding to a proton spectral width of 30.0 ppm. All data were processed using an EM window function and 5-Hz line broadening, without linear prediction in TopSpin (Bruker). Spectra were referenced to trimethylsilyl propanate at 0 ppm prior to overlay and analysis.

E. coli fractionation

To avoid overloading the Tat system for *in vivo* transport assays, maquette genes were cloned into the KpnI and XbaI sites of the low copy number pEXT22 plasmid for expression from a tightly controlled *tac* promoter (33). The pEXT22 constructs (Table S1) were co-transformed into *E. coli* BL21(DE3) cells along with empty pET-21a(+). Cultures (50 ml) were grown at 37 °C with shaking (230 rpm) in LB medium with 30 μg ml^{-1} of kanamycin and 100 μg ml^{-1} of ampicillin in 250-ml Erlenmeyer flasks. At an A_{600} of ~0.6 expression was induced with 0.5 mM IPTG for 2 h. Cells were harvested by centrifugation (3,900 × *g*, 30 min, 4 °C) and washed with buffer F (100 mM Tris acetate, pH 8.2, 500 mM sucrose, 5 mM EDTA). To prepare cell-free lysates, cells were re-suspended to A_{600} ~2 in 2 ml of chilled Buffer C and lysed by sonication on ice. The supernatant, following clarification by centrifugation (16,600 × *g*, 30 min 4 °C), was collected as the cell-free extract. Periplasmic fractions were obtained using a procedure based on the EDTA/lysozyme/osmotic shock method described by Randall and Hardy (44). Briefly, cells were re-suspended to A_{600} ~10 in 500 μl of chilled Buffer F followed by the addition of 5 μl of 1 mg ml^{-1} lysozyme. 500 μl of chilled QH_2O was added and cells were incubated on ice for 5 min prior to the addition of 20 μl of 1 M $MgSO_4$. Spheroplasts were pelleted by centrifugation at 16,600 × *g* for 30 min in a pre-chilled (4 °C) microcentrifuge and the supernatant was collected as the periplasm.

Immunoblotting

Immunoblotting was performed essentially as described previously (45). Briefly, proteins were separated by SDS-PAGE on 12% BisTris gels (Invitrogen), transferred to polyvinylidene difluoride membranes (Invitrogen), and incubated with anti-His_6 (Bethyl Laboratories, Inc.), anti-β-lactamase (Abcam), or anti-GroEL (Sigma) primary antibodies followed by an appropriate horseradish peroxidase-conjugated secondary antibody (Sigma). Chemiluminescence was detected using the WESTAR ETA C 2.0 chemiluminescent substrate (Cyanagen) on an Amersham Biosciences Imager 600 (GE Healthcare).

Purification of Tat-exported proteins from E. coli periplasm

The *E. coli* W3110 TatExpress cell line (37) was transformed with the pEXT22 vector containing TorA–BT6, TorA–BT6M1, or TorA–BT6M0 (Table S1). 500-ml cultures were grown in 2-liter Erlenmeyer flasks at 30 °C with 220 rpm agitation. At an A_{600} of ~0.6, protein production was induced with 0.5 mM IPTG and cultures were incubated for 24 h, after which cells were harvested by centrifugation (3,900 × *g*, 30 min, 4 °C). To obtain periplasmic fractions, cells were resuspended in 10 ml of chilled buffer F. 10 ml of chilled milliQH_2O was added followed by 800 μl of 1 mg ml^{-1} of lysozyme and samples were incubated on ice for 10 min. 800 μl of 1 M $MgSO_4$ was added and the solution was centrifuged (16,600 × *g*, 30 min, 4 °C) with the supernatant collected as the periplasm.

Periplasmic fractions were applied to a Chelating Sepharose Fast Flow column (GE Healthcare) pre-equilibrated with 10 mg ml^{-1} of nickel sulfate. The column was washed with 20 ml of buffer A and 20 ml of buffer G (50 mM HEPES, pH 7.4, 500 mM NaCl, 50 mM imidazole). Protein was eluted with 10 ml of buffer H (50 mM HEPES, pH 7.4, 100 mM NaCl, 400 mM imidazole) and the elution fractions were collected.

Author contributions—G. A. S., K. J. G., N. B. P. A., D. M. J. M., A. S. J., A. J. R., D. B. A., A. A. B., F. S., P. T., P. D., P. L. D., C. R., A. H., and C. N. H. data curation; G. A. S., K. J. G., N. B. P. A., D. M. J. M., A. S. J., A. J. R, D. B. A., A. A. B., F. S., P. T., P. D., P. L. D., C. R., A. H., and C. N. H. formal analysis; G. A. S., K. J. G., N. B. P. A., D. M. J. M., A. S. J., A. J. R, D. B. A., A. A. B., F. S., P. T., P. D., P. L. D., C. R., A. H., and C. N. H. methodology; G. A. S., K. J. G., A. H., and C. N. H. writing-original draft; G. A. S., N. B. P. A., A. H., and C. N. H. writing-review and editing; K. J. G., A. S. J., F. S., P. T., P. D., P. L. D., C. R., A. H., and C. N. H. investigation; N. B. P. A., F. S., P. T., and P. D. software; K. J. G., A. S. J., F. S., P. T., P. D., P. L. D., C. R., A. H., and C. N. H. conceptualization; C. N. H. project administration.

References

1. de Keyzer, J., van der Does, C., and Driessen, A. J. (2003) The bacterial translocase: a dynamic protein channel complex. *Cell Mol. Life Sci.* **60,** 2034–2052 CrossRef Medline
2. Clark, S. A., and Theg, S. M. (1997) A folded protein can be transported across the chloroplast envelope and thylakoid membranes. *Mol. Biol. Cell* **8,** 923–934 CrossRef Medline
3. Berks, B. C. (2015) The twin-arginine protein translocation pathway. *Annu. Rev. Biochem.* **84,** 843–864 CrossRef Medline
4. Berks, B. C. (1996) A common export pathway for proteins binding complex redox cofactors? *Mol. Microbiol.* **22,** 393–404 CrossRef Medline
5. Santini, C. L., Ize, B., Chanal, A., Müller, M., Giordano, G., and Wu, L. F. (1998) A novel Sec-independent periplasmic protein translocation pathway in *Escherichia coli*. *EMBO J.* **17,** 101–112 Medline
6. Natale, P., Brüser, T., and Driessen, A. J. M. (2008) Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane: distinct translocases and mechanisms. *Biochim. Biophys. Acta* **1778,** 1735–1756 CrossRef
7. Palmer, T., and Berks, B. C. (2012) The twin-arginine translocation (Tat) protein export pathway. *Nat. Rev. Microbiol.* **10,** 483–496 CrossRef Medline
8. Rodrigue, A., Chanal, A., Beck, K., Müller, M., and Wu, L. (1999) Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial tat pathway. *J. Biol. Chem.* **274,** 13223–13228 CrossRef Medline
9. Chaddock, A. M., Mant, A., Karnauchov, I., Brink, S., Herrmann, R. G., Klösgen, R. B., and Robinson, C. (1995) A new type of signal peptide: central role of a twin-arginine motif in transfer signals for the ΔpH-dependent thylakoidal protein translocase. *EMBO J.* **14,** 2715–2722 Medline
10. Sargent, F., Stanley, N. R., Berks, B. C., and Palmer, T. (1999) Sec-independent protein translocation in *Escherichia coli*. *J. Biol. Chem.* **274,** 36073–36082 Medline
11. Patel, R., Smith, S. M., and Robinson, C. (2014) Protein transport by the bacterial Tat pathway. *Biochim. Biophys. Acta* **1843,** 1620–1628 CrossRef
12. Halbig, D., Wiegert, T., Blaudeck, N., Freudl, R., and Sprenger, G. A. (1999) The efficient export of NADP-containing glucose-fructose oxidoreductase to the periplasm of *Zymomonas mobilis* depends both on an intact twin-arginine motif in the signal peptide and on the generation of a structural export signal induced by cofactor bind. *Eur. J. Biochem.* **263,** 543–551 CrossRef Medline
13. Thomas, J. D., Daniel, R. A., Errington, J., and Robinson, C. (2001) Export of active green fluorescent protein to the periplasm by the twin-arginine translocase (Tat) pathway in *Escherichia coli*. *Mol. Microbiol.* **39,** 47–53 Medline
14. DeLisa, M. P., Tullman, D., and Georgiou, G. (2003) Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 6115–6120 CrossRef Medline
15. Rocco, M. A., Waraho-Zhmayev, D., and DeLisa, M. P. (2012) Twin-arginine translocase mutations that suppress folding quality control and permit export of misfolded substrate proteins. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 13392–13397 CrossRef Medline
16. Richter, S., Lindenstrauss, U., Lücke, C., Bayliss, R., and Brüser, T. (2007) Functional tat transport of unstructured, small, hydrophilic proteins. *J. Biol. Chem.* **282,** 33257–33264 CrossRef Medline
17. Lindenstrauss, U., and Brüser, T. (2009) Tat transport of linker-containing proteins in *Escherichia coli*. *FEMS Microbiol. Lett.* **295,** 135–140 Medline
18. Jones, A. S., Austerberry, J. I., Dajani, R., Warwicker, J., Curtis, R., Derrick, J. P., and Robinson, C. (2016) Proofreading of substrate structure by the twin-arginine translocase is highly dependent on substrate conformational flexibility but surprisingly tolerant of surface charge and hydrophobicity changes. *Biochim. Biophys. Acta* **1863,** 3116–3124 CrossRef
19. Oresnik, I. J., Ladner, C. L., and Turner, R. J. (2001) Identification of a twin-arginine leader-binding protein. *Mol. Microbiol.* **40,** 323–331 CrossRef Medline
20. Ray, N., Oates, J., Turner, R. J., and Robinson, C. (2003) DmsD is required for the biogenesis of DMSO reductase in *Escherichia coli* but not for the interaction of the DmsA signal peptide with the Tat apparatus. *FEBS Lett.* **534,** 156–160 CrossRef Medline
21. Jack, R. L., Buchanan, G., Dubini, A., Hatzixanthis, K., Palmer, T., and Sargent, F. (2004) Coordinating assembly and export of complex bacterial proteins. *EMBO J.* **23,** 3962–3972 CrossRef Medline
22. Matos, C. F., Robinson, C., and Di Cola, A. (2008) The Tat system proofreads FeS protein substrates and directly initiates the disposal of rejected molecules. *EMBO J.* **27,** 2055–2063 CrossRef Medline
23. DeGrado, W. F., Wasserman, Z. R., and Lear, J. D. (1989) Protein design, a minimalist approach. *Science* **243,** 622–628 CrossRef Medline
24. Shifman, J. M., Moser, C. C., Kalsbeck, W. A., Bocian, D. F., and Dutton, P. L. (1998) Functionalized *de novo* designed proteins: mechanism of proton coupling to oxidation/reduction in heme protein maquettes. *Biochemistry* **37,** 16815–16827 CrossRef Medline
25. Grosset, A. M., Gibney, B. R., Rabanal, F., Moser, C. C., and Dutton, P. L. (2001) 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,** 5474–5487 CrossRef Medline
26. Lichtenstein, B. R., Farid, T. A., Kodali, G., Solomon, L. A., Anderson, J. L., Sheehan, M. M., Ennist, N. M., Fry, B. A., Chobot, S. E., Bialas, C., Mancini, J. A., Armstrong, C. T., Zhao, Z., Esipova, T. V., Snell, D., *et al.* (2012) Engineering oxidoreductases: maquette proteins designed from scratch. *Biochem. Soc. Trans.* **40,** 561–566 CrossRef Medline
27. Edelman, G. M., and Gally, J. A. (2001) Degeneracy and complexity in biological systems. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 13763–13768 CrossRef Medline
28. Farid, T. A., Kodali, G., Solomon, L. A., Lichtenstein, B. R., Sheehan, M. M., Fry, B. A., Bialas, C., Ennist, N. M., Siedlecki, J. A., Zhao, Z., Stetz, M. A., Valentine, K. G., Anderson, J. L. R., Wand, A. J., Discher, B. M., Moser, C. C., and Dutton, P. L. (2013) Elementary tetrahelical protein design for diverse oxidoreductase functions. *Nat. Chem. Biol.* **9,** 826–833 CrossRef Medline
29. Choma, C. T., Lear, J. D., Nelson, M. J., Dutton, P. L., Robertson, D. E., and DeGrado, W. F. (1994) Design of a heme-binding four-helix bundle. *J. Am. Chem. Soc.* **116,** 856–865 CrossRef
30. Koder, R. L., Anderson, J. L., Solomon, L. A., Reddy, K. S., Moser, C. C., and Dutton, P. L. (2009) Design and engineering of an O_2 transport protein. *Nature* **458,** 305–309 CrossRef Medline
31. Greenfield, N. J. (2006) Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* **1,** 2527–2535 Medline
32. Liu, C. P., Churchward, G., and Caro, L. (1983) The repA2 gene of the plasmid R100.1 encodes a repressor of plasmid replication. *Plasmid* **10,** 148–155 CrossRef Medline
33. Dykxhoorn, D. M., St. Pierre, R., and Linn, T. (1996) A set of compatible tac promoter expression vectors. *Gene* **177,** 133–136 CrossRef Medline
34. Stanley, N. R., Palmer, T., and Berks, B. C. (2000) The twin-arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in *Escherichia coli*. *J. Biol. Chem.* **275,** 11591–11596
35. Lei, S. P., Lin, H. C., Wang, S. S., Callaway, J., and Wilcox, G. (1987) Characterization of the *Erwinia carotovora pelB* gene and its product pectate lyase. *J. Bacteriol.* **169,** 4379–4383 CrossRef Medline
36. Yoon, S., Kim, S., and Kim, J. (2010) Secretory production of recombinant proteins in *Escherichia coli*. *Recent Pat. Biotechnol.* **4,** 23–29 CrossRef
37. Browning, D. F., Richards, K. L., Peswani, A. R., Roobol, J., Busby, S. J. W., and Robinson, C. (2017) *Escherichia coli* "TatExpress" strains super-secrete human growth hormone into the bacterial periplasm by the Tat pathway. *Biotechnol. Bioeng.* **114,** 2828–2836 CrossRef Medline
38. Hitchcock, A., Hall, S. J., Myers, J. D., Mulholland, F., Jones, M. A., and Kelly, D. J. (2010) Roles of the twin-arginine translocase and associated chaperones in the biogenesis of the electron transport chains of the human pathogen *Campylobacter jejuni*. *Microbiology* **156,** 2994–3010 Medline
39. De Buck, E., Lammertyn, E., and Anné, J. (2008) The importance of the twin-arginine translocation pathway for bacterial virulence. *Trends Microbiol.* **16,** 442–453 CrossRef Medline

40. Shen, Y., Maupetit, J., Derreumaux, P., and Tufféry, P. (2014) Improved PEP-FOLD approach for peptide and miniprotein structure prediction. *J. Chem. Theory Comput.* **10,** 4745–4758 CrossRef Medline
41. Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: algorithms for highly efficient, load balanced, and scalable molecular simulations. *J. Chem. Theory Comput.* **4,** 435–447 CrossRef Medline
42. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 10037–10041 CrossRef Medline
43. Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182,** 319–326 CrossRef Medline
44. Randall, L. L., and Hardy, S. J. (1986) Correlation of competence for export with lack of tertiary structure of the mature species: a study *in vivo* of maltose-binding protein in *E. coli*. *Cell* **46,** 921–928 CrossRef Medline
45. Grayson, K. J., Faries, K. M., Huang, X., Qian, P., Dilbeck, P., Martin, E. C., Hitchcock, A., Vasilev, C., Yuen, J. M., Niedzwiedzki, D. M., Leggett, G. J., Holten, D., Kirmaier, C., and Neil Hunter, C. (2017) Augmenting light coverage for photosynthesis through YFP-enhanced charge separation at the *Rhodobacter sphaeroides* reaction centre. *Nat. Commun.* **8,** 13972 CrossRef Medline

Probing the quality control mechanism of the *Escherichia coli* twin-arginine translocase with folding variants of a *de novo*–designed heme protein

George A. Sutherland, Katie J. Grayson, Nathan B. P. Adams, Daphne M. J. Mermans, Alexander S. Jones, Angus J. Robertson, Dirk B. Auman, Amanda A. Brindley, Fabio Sterpone, Pierre Tuffery, Philippe Derreumaux, P. Leslie Dutton, Colin Robinson, Andrew Hitchcock and C. Neil Hunter

J. Biol. Chem. 2018, 293:6672-6681.
doi: 10.1074/jbc.RA117.000880 originally published online March 20, 2018
