The Dissertation Committee for Danielle Tullman Ercek Certifies that this is the approved version of the following dissertation:

Characterization and Engineering of the Twin-Arginine Translocation Pathway of *Escherichia coli*

Committee:

George Georgiou, Supervisor

Brent Iverson

Charles Earhart

Brian Korgel

Charles Buddie Mullins

Characterization and Engineering of the Twin-Arginine Translocation Pathway of *Escherichia coli*

by

Danielle Tullman Ercek, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2006

Dedication

For my parents, Maureen and Edward

Characterization and Engineering of the Twin-Arginine Translocation Pathway of *Escherichia coli*

Publication No._____

Danielle Tullman Ercek, Ph.D.

The University of Texas at Austin, 2006

Supervisor: George Georgiou

The twin-arginine translocation (Tat) pathway of *Escherichia coli* provides a novel method for the export of proteins from the cytoplasm to the periplasm. Remarkably, it allows for large, folded proteins to cross the inner membrane, with no apparent effect on cell viability. Protein export from the cytoplasm is employed in a variety of biotechnological applications including manufacturing and protein engineering. However, until the discovery of the Tat pathway, such applications relied on the Sec pathway, in which proteins transverse the lipid bilayer membrane in an extended form and protein folding takes place after export. Many proteins of biotechnological interest are not compatible with export via the "classical" Sec pathway. Thus, the export of such

Sec-incompatible proteins via the Tat pathway could open the way for new biotechnology applications.

This work explores several mechanistic, physiological and technology-related aspects of Tat export In all organisms, proteins are secreted by virtue of a peptide extension, or signal peptide, comprising 15-45 amino acids. The signal peptide serves as a "zip code" and is cleaved after export. *E. coli* contains 29 putative Tat-specific signal peptides but their ability to mediate export via the Tat pathway has not been confirmed experimentally. The export pathway (Tat or Sec) utilized by this set of 29 signal peptides was characterized using fusions to protein reporters. The reporter proteins chosen for this study are functional only when translocated across the membrane either via the Sec or Tat pathways. Surprisingly, it was found that while 11/29 signal peptides are Tat-specific and 2/29 are Sec-specific, a set of 16/29 signal peptides were able to direct export via both the Tat and Sec pathways. Interestingly, increasing the charge of the region surrounding the cleavage site – particularly the N-terminus of the mature protein – resulted in Tat specificity.

In separate studies we showed that in addition to an appropriate signal sequence, proteins destined for export via the Tat pathway must complete their folding in the cytoplasm. Partially folded proteins are not competent for export via this pathway. The requirement for folding in the cytoplasm prior to export was demonstrated by using an elegant system whereby the conformation of the polypeptide chain in the cytoplasm could be controlled using conditions that supported or prevented the formation of disulfide

bonds. These results led us to propose that the Tat pathway contains an intrinsic folding quality control mechanism, a concept that has since been widely adopted in the literature.

Finally, a new methodology was developed for the engineering of proteins that require the cytoplasmic machinery to fold but must then be exported into the bacterial periplasmic space. Specifically, we created a Tat-based system to enable the display of proteins on filamentous phage, a prerequisite for the high-throughput screening of protein libraries. This system relies on the forced dimerization of the phage coat protein p3, which is exported into the periplasm via the Sec pathway, and the protein that is to be displayed, which is exported via the Tat pathway. Forced dimerization of p3 and the desired protein in the periplasm was mediated by coiled-coil interactions. We further demonstrated that this Tat-dependent display platform shows promise for use in engineering ligand-binding loops into green fluorescent protein (GFP) for sensor applications.

Table of Contents

Table of Tables

Table of Figures

Abbreviations

aa	amino acid
ATP	adenosine triphosphate
BLAST	basic local alignment search tool
BPTI	bovine pancreatic trypsin inhibitor
BSA	bovine serum albumin
cfu	colony forming units
$C_{H/L}$	constant domain of the heavy (H) or light (L) chain of an antibody
C:ox	strain with an oxidizing cytoplasm
cpTat	chloroplast twin-arginine translocation pathway
DHFR	dihydrofolate reductase
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
ELISA	enzyme-linked immunosorbent assay
F_{AB}	antibody fragment
FACS	fluorescence-activated cell sorting
FLAG tag	peptide sequence D-Y-K-D-D-D-D-K
GFP	green fluorescent protein
HRP	horseradish peroxidase
IgG	immunoglobulin
IPTG	isopropyl-β-D-thiogalactopyranoside
kDa	kilo Dalton

LB	Luria-Bertani (Miller) growth medium
MBP	maltose binding protein
NE	no expression detected
OD	optical density
PCR	polymerase chain reaction
PhoA	alkaline phosphatase
PEG	polyethylene glycol
PBS	phosphate-buffered saline
pmf	proton motive force
pNPP	para-nitrophenyl phosphate
REMP	redox enzyme maturation protein
scFv	single-chain variable fragment
SDS	sodium dodecyl sulfate
PAGE	polyacrylamide gel electrphoresis
Sec	general Secretory pathway
SRP	signal recognition particle pathway
ss	signal sequence/signal peptide
Tat	twin-arginine translocation pathway
TBS/TBST	tris-buffered saline, tris-buffered saline with Tween-20
TMAO	trimethylamine-N-oxide
$V_{H/L}$	variable domain of the heavy (H) or light (L) chain of an antibody
wt	wild type

Chapter 1

Introduction, Background and State of the Art

1.1 Introduction

Bacterial protein expression and engineering play a major role in the biotechnology industry. In particular, recombinant protein production in bacteria – specifically, *Escherichia coli* – accounts for about 40% of the total biopharmaceutical protein market to date, a market generating over $30 billion per year in revenue (Walsh 2005) and projected to reach anywhere from $50 billion to over $65 billion per year by the end of the decade (Walsh 2005; Manheim *et al.* 2006). Bacteria are often preferred for the expression of proteins for both therapeutic and non-therapeutic applications (Maynard and Georgiou 2000; Blazek and Celer 2003; Mabry *et al.* 2006). Despite rapid growth in bioengineering research over the past decades, there is still much progress to be made in understanding, utilizing, and manipulating bacterial systems for both research and industrial purposes.

1.2 Protein translocation and biotechnology

Gram-negative bacterial cells have both an outer and inner cellular membrane, with the inner membrane separating the cytoplasm from the outer, periplasmic compartment. Similarly, the outer membrane serves as a barrier between the periplasm

and the extracellular surroundings. The inner membrane is a phospholipid bilayer, while the outer membrane has a phospholipid leaflet and an outer lipopolysaccharide leaflet. A thin peptidoglycan wall is also present, and is connected via lipoproteins to the outer membrane on the periplasmic side. Both membranes also contain embedded proteins. The cytoplasm maintains a reducing environment while the periplasm maintains oxidizing conditions that allow for the formation of disulfide bonds.

Expressing a large quantity of protein in the cytoplasmic compartment often leads to difficulties in downstream processing. For instance, protein aggregation is a problem when overexpressing a heterologous protein in the cytoplasm (Villaverde and Carrio 2003); in such cases, solubilization, denaturing, and refolding of the polypeptide chain are required to obtain the biologically active form of the protein (Walsh 2005). Moreover, many proteases are present in the cytoplasm that can degrade the produced protein and therefore reduce the product yield. Finally, disruption of the inner membrane and outer cell wall is necessary to harvest any proteins produced in the cytoplasm, but is a much more difficult process than either collecting proteins that have been secreted into the growth medium, or retrieving periplasmic proteins by removing the outer membrane of the cell.

For all of the reasons described above, it is desirable to engineer the transport of proteins out of the cytoplasmic compartment in which they are synthesized (Georgiou and Segatori 2005). Such a task is not trivial, and involves translocating a hydrophilic protein across the hydrophobic phospholipid bilayer (the inner membrane). This energetically-unfavorable process is carried out by a complicated cellular machinery that

constitutes a protein export pathway. Multiple protein export pathways are found in bacteria. The pathways discovered to date share some basic characteristics. For example, each pathway makes use of an energy source (e.g. ATP) to drive the translocation process. Additionally, with the notable exception of the so-called Type I export pathway, all proteins to be exported are synthesized with an N-terminal polypeptide tag, or signal peptide. The signal peptide not only marks a protein for export, it also dictates the appropriate export pathway to be used, and will be described in further detail in Section 1.3.1. Finally, regardless of the export pathway utilized, the export process must occur without allowing the leakage of ions or other molecules into or out of the cytoplasmic compartment – the membrane must continue to serve its function as a barrier while permitting the transport of these large polypeptides.

1.2.1 The Sec pathway

The General Secretory (Sec) pathway is the most well-studied protein export pathway in bacteria, and therefore has been utilized in most biotechnological applications to date. It exports a newly synthesized polypeptide across the inner membrane in a largely unfolded state. Export occurs through a protein pore (the Sec pore) which is formed by the SecYEG proteins. The ATPase SecA is also required for Sec export. The SecA protein serves as a nanomachine that couples the energy of ATP hydrolysis to the insertion of the polypeptide into the Sec pore, with approximately 20 amino acids crossing the pore per cycle of ATP (Uchida *et al.* 1995). The polypeptide is threaded through the Sec pore in a ratchet-like motion (Economou and Wickner 1994; Pohlschroder *et al.* 2005). Upon reaching the periplasm the signal peptide is cleaved by

signal peptidase I and the protein is folded. Since the Sec pore is quite narrow, it cannot accommodate proteins that have already folded (Eilers and Schatz 1986; Schatz and Dobberstein 1996). The protein SecB is sometimes required to maintain the polypeptide in an unfolded state until it reaches the Sec translocase (Kim and Kendall 2000).

Sec-dependent signal peptides, e.g. ssOmpA, ssPhoA, ssPelB, etc., have been widely used to direct the export of biotechnologically-relevant polypeptides. However, it has been known for over 20 years that many heterologous proteins, as well as some native bacterial cytoplasmic proteins, fail to be exported via the Sec pathway, in part due to folding considerations (Nouwen *et al.* 1996). Other proteins are compatible with transport but once in the periplasm are subjected to extensive degradation before they can reach their more stable, native conformation. Yet another group of proteins, exemplified by green fluorescent protein (GFP), are unable to fold properly once exported from the cytoplasm (Feilmeier *et al.* 2000), and are thus incompatible with the Sec pathway since it cannot export folded polypeptides.

1.2.1 The Tat pathway

Nearly ten years ago, the bacterial twin-arginine translocation (Tat) pathway was discovered to be capable of transporting native proteins that are not compatible with Sec export. Early on it was suggested that the Tat pathway exports folded proteins (Rodrigue *et al.* 1999). This finding stimulated interest in the potential practical applications of the Tat pathway for biotechnology purposes, namely the production and engineering of proteins that cannot be exported in an active form from the bacterial cytoplasm via the Sec pathway. For example, Tinker *et al.* (2005) used the Tat pathway for the export of

GFP fused to the A subunit of cholera toxin or the heat labile toxins LTI and LTIIb. Concomitant export of the respective B subunits, via the Sec apparatus, resulted in the formation of active and fluorescent holotoxin fusions in the periplasm. Previous attempts at exporting such fluorescent toxin molecules using only the Sec pathway had been unsuccessful. The authors suggest that this is because the other proteins they fused to the holotoxin may have been incompatible with export via the Sec pathway (Tinker *et al.* 2005). Additionally, it is known that GFP is nonfluorescent when exported to the periplasm via the Sec pathway (Feilmeier *et al.* 2000). There have also been reports of using the Tat pathway to export proteins of commercial interest (e.g. organophosphorous hydrolase in *E. coli*, Kang *et al.* 2005). Recently, DeLisa and coworkers demonstrated that the Tat pathway could be used to uncover and even engineer proteins that are both correctly folded and have increased solubility (Fisher *et al.* 2006). This clever screen relied on the fusion of a target protein to a Tat signal peptide and the reporter protein β-lactamase, which confers resistance the antibiotic ampicillin (Fisher *et al.* 2006). If the target protein domain of the fusion protein cannot fold, then the fusion is not compatible with Tat export. Consequently, β-lactamase is not exported into the periplasm and cannot protect the cells from the action of the antibiotic. This system capitalizes on the discovery that the Tat pathway only mediates the export of proteins that have folded in the cytoplasm, which is discussed in detail in Chapter 3.

1.3 Background: The Tat pathway

The Tat pathway serves the unique and remarkable role of transporting folded proteins across energy transducing membranes (Wickner and Schekman 2005). It is widespread within the microbial world and in plants, as it is found in archaea, bacteria, chloroplasts and plant mitochondria (Yen *et al.* 2002). In bacteria, the Tat pathway catalyzes the export of proteins from the cytoplasm across the inner membrane. In chloroplasts, the Tat components are found in the thylakoid membrane and direct the import of proteins from the stroma. The discovery of the Tat pathway dates to the early 1990s when it was noticed that a subset of polypeptides in chloroplasts could be translocated independently of ATP hydrolysis, unlike the Sec pathway which uses ATP hydrolysis as the driving force for the export process. Instead, the export of these chloroplast substrates relied solely on the proton gradient (Mould and Robinson 1991; Cline *et al.* 1992). The proton gradient results in two distinct effects – a chemical gradient resulting from the differing concentrations of solute on either side of the membrane, and an electrical gradient due to the charge differential across the membrane (Campbell 2001). Collectively, the potential energy of such a gradient is known as the proton motive force (pmf), and it can be tapped for energy by allowing the protons to flow in a controlled manner across the membrane, much like water through a hydroelectric plant (Campbell 2001). For this reason, the Sec-independent pathway of chloroplasts was initially designated the "ΔpH pathway" but more recently has been termed the cpTat pathway. In 1995, the first evidence was presented that the cpTat

pathway enables the translocation of pre-folded proteins (Creighton *et al.* 1995). Shortly after, it was observed that a group of bacterial periplasmic proteins containing various cofactors share a unique type of signal peptide containing a consensus 'double arginine' motif also found in substrates of the chloroplast pathway (Berks 1996). The existence of a bacterial pathway analogous to the one in chloroplasts was thus established; it was initially termed mtt for membrane_targeting and translocation (Weiner *et al.* 1998) and then later Tat for twin-arginine translocation (Sargent *et al.* 1998).

1.3.1 Signal peptides

With the exception of proteins that are exported across membranes via specialized systems (e.g. Type I secretion), a signal peptide is necessary in order to initiate export (Wickner and Schekman 2005). Signal peptides consist of three domains: the positively charged N-terminal domain (or n-region), the hydrophobic domain (h-region), and the C-terminal domain (c-region) (Figure 1.1). Sequence analysis revealed that signal peptides capable of targeting the Tat translocon contain the consensus motif Ser/Thr-**Arg**-**Arg**-X-Phe-Leu-Lys (where X is any polar amino acid). The nearly invariant twin arginine dipeptide in this motif gave rise to the pathway's name. In addition, the h-region of Tat signal peptides is typically less hydrophobic than that of Sec-specific signal peptides due to the presence of more glycine and threonine residues (Cristobal *et al.* 1999). Tat signal peptides tend to be longer than their Sec counterparts (primarily due to an extended n-region) with average lengths of 38 and 24 amino acids, respectively (Cristobal *et al.* 1999).

Figure 1.1. Features of a typical Tat signal peptide, ssTorA from *E. coli*. The n-region is shown in purple, the h-region in green, and the c-region in gold. The vertical dashed line indicates the site of cleavage by signal peptidase I.

The first arginine of the Arg-Arg dipeptide in the consensus motif can be substituted with Lys, and the second one can be substituted with Gln or Asn as well as Lys, with varying efficiencies (Stanley *et al.* 2000; Buchanan *et al.* 2001; DeLisa *et al.* 2002; Ize *et al.* 2002). The substitution of both arginines to lysines blocks the export of physiological Tat substrate proteins (Stanley *et al.* 2000). To date, only two natural Tat substrates are known to lack the arginine dipeptide: the pre-propenicillin amidase of *E. coli* (containing an Arg-Asn-Arg; ref. Ignatova *et al.* 2002), and tetrathionate reductase from *Salmonella enterica* (Lys-Arg; ref. Hinsley *et al.* 2001). The other residues within the consensus motif are more amenable to amino acid substitutions. For example, the Phe and Leu residues can be substituted with other highly hydrophobic amino acids with little effect on export, although efficiency does drop upon substitution with less hydrophobic amino acids, such as Tyr (Stanley *et al.* 2000). Also, neither the Ser nor the Lys within the consensus are essential for Tat targeting (Stanley *et al.* 2000).

1.3.2 The Tat proteins

The minimal set of components required for Tat translocation in *Escherichia coli* consists of three integral membrane proteins, one from each of the TatA, TatB, and TatC families (Tha4, Hcf106 and cpTatC in chloroplasts, respectively). Integral membrane proteins contain at least one highly hydrophobic alpha helix transversing the phospholipid bilayer of the inner membrane; this domain is therefore known as the transmembrane helix. TatC is the most conserved of the Tat proteins and sequence conservation is particularly strong within the transmembrane domains (Bogsch *et al.* 1998; Buchanan *et al.* 2002). In a phylogenetic tree of TatC, the cyanobacterial TatC proteins cluster with the chloroplast TatC proteins, highlighting the likely evolutionary origin of chloroplasts. TatA and TatB are very similar in terms of structure and sequence and therefore it is difficult to clearly delineate TatA from TatB homologues in some organisms. Phylogenetic analysis indicates that TatA homologues cluster more tightly than TatB proteins, suggesting that the latter may have diverged in sequence more rapidly than TatA (Yen *et al.* 2002). It appears that a very early gene duplication event gave rise to *tatA* and *tatB,* but since then only *tatA* has been duplicated (e.g. to give *tatE* in *E. coli*).

The *E. coli* TatC is a 258 amino acid polytopic membrane protein with six transmembrane helices and both the N- and C- termini located on the cytoplasmic side of the membrane (Figure 1.2; refs. Behrendt *et al.* 2004; Ki *et al.* 2004). The cytoplasmic side of TatC seems to be particularly important for function. Several residues within cytoplasmic loops are completely or functionally conserved across prokaryotes, chloroplasts, and plant mitochondria and site-directed mutagenesis of a number of these

residues has revealed that they are important or essential for TatC function (Allen *et al.* 2002; Buchanan *et al.* 2002; Barrett and Robinson 2005). *In vitro* biochemical studies have revealed that TatC serves as the initial docking site for Tat signal peptides. Signal peptides from bacteria and chloroplasts have been shown to interact with TatC either alone or in a complex with TatB (Cline and Mori 2001; de Leeuw *et al.* 2002; Alami *et al.* 2003). It is likely that more than one TatC monomer is required for each targeting event, since two mutant versions of TatC that alone blocked Tat-dependent transport were shown to support transport when expressed together (Buchanan *et al.* 2002). The interaction between TatC and the signal peptide *in vitro* requires the twin arginine motif (Cline and Mori 2001; Alami *et al.* 2003).

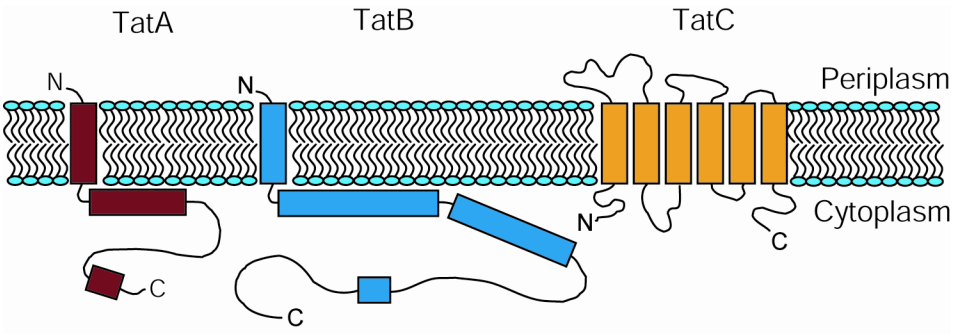

Figure 1.2. The predicted structure and topology of the *E. coli* Tat components. Predicted helical regions are shown as boxes.

The *E. coli* TatB protein consists of 171 amino acids, with an N-terminal transmembrane helix followed by a short hinge region leading to a predicted amphipathic helix that may lie along the cytoplasmic side of the inner membrane (De Leeuw *et al.* 2001). The region of TatB following the amphipathic helix is predicted to remain helical in nature to approximately residue 75, after which point TatB is considered largely

unstructured (Figure 1.2). This latter region of TatB was shown through truncation analysis to be non-essential for Tat transport. However, once TatB is truncated to the amphipathic helix, activity is lost (Lee *et al.* 2002). TatB was shown *in vitro* to contact the entire length of the signal peptide and also the mature protein more than 20 residues past the signal peptide cleavage site (Alami *et al.* 2003). Such interactions were only seen when TatC was present, suggesting that substrate targeting involves a series of sequential interactions, with TatC forming the primary recognition site before the substrate is transferred to TatB. In turn TatB could be considered as a mediator between TatC and TatA, contacting the substrate after initial recognition by TatC and then potentially being involved in transfer to a complex consisting mainly of the TatA protein prior to translocation (Sargent *et al.* 2001; Alami *et al.* 2003). It should be noted that like TatC, TatB is essential for the transport of endogenous substrates in *E. coli* (Sargent *et al.* 1999).

The TatA protein is the most abundant of the Tat components and is estimated to be present at around 20 times excess relative to TatB and TatC (Jack *et al.* 2001; Sargent *et al.* 2001). As previously mentioned, the TatA protein is predicted to have similar structure to TatB with an N-terminal transmembrane domain followed by a short hinge region leading to an amphipathic helical region and an unstructured soluble C-terminal tail (Figure 1.2). TatA is shorter than TatB by 89 amino acids and contains shorter cytoplasmic helical and soluble domains. The transmembrane domain of TatA is critical for export and is important for interactions with TatB (Lee *et al.* 2002; Dabney-Smith *et al.* 2003; Barrett and Robinson 2005). Upon removal of the transmembrane domain TatA

becomes soluble (De Leeuw *et al.* 2001). TatA is thought to function at a late stage in translocation and likely forms the major constituent of the Tat pore itself (Mori and Cline 2002; Alami *et al.* 2003; Gohlke *et al.* 2005).

The stoichiometry of the TatABC components is critical for the export function. Overexpression of *tatB* results in a complete loss of Tat transport (Sargent *et al.* 1999). The overexpression of *tatA* has a less severe but nonetheless significant effect on translocation (Sargent *et al.* 1999). In contrast, high expression of TatC can relieve saturation of the Tat pathway caused by the high level expression of non-physiological substrates such as GFP fusions (DeLisa *et al.* 2004). A general increase in the protein flux through the Tat pathway is only seen when *tatABC* are overexpressed in an operon together (Sargent *et al.* 2001; Yahr and Wickner 2001; Alami *et al.* 2002).

1.3.3 The Tat translocon

Biochemical studies suggest a highly dynamic picture with the Tat proteins participating in various multimeric membrane complexes that are thought to interact transiently during the translocation cycle. However, the precise steps in this process remain a mystery, including the mechanism by which the pmf provides the energy for vectorial protein transfer across the inner membrane (Lee *et al.* 2006).

Several lines of evidence suggest that TatB and TatC interact with a 1:1 stoichiometry to form a functional complex. TatB and TatC from *E. coli* purify predominantly as a 350-600 kDa membrane complex containing multiple copies of the two proteins in equimolar amounts (Bolhuis *et al.* 2001; de Leeuw *et al.* 2002; Oates *et al.* 2003). Furthermore, TatC is unstable in the absence of TatB (Sargent *et al.* 1999)

suggesting a role for TatB in stabilizing the TatBC complex. A second membrane complex of ~600 kDa containing TatA and TatB has also been purified (Bolhuis *et al.* 2000; Sargent *et al.* 2001; de Leeuw *et al.* 2002) and the two proteins can be co-immunoprecipitated (Bolhuis *et al.* 2000). Chemical cross-linking has revealed that TatB forms homodimers (De Leeuw *et al.* 2001; Hicks *et al.* 2003). TatA also forms at least homotetramers in the *E. coli* inner membrane (De Leeuw *et al.* 2001; Porcelli *et al.* 2002; Hicks *et al.* 2003). When expressed in the absence of TatBC, TatA forms multiple complexes ranging in size from 100 to 600 kDa (De Leeuw *et al.* 2001; de Leeuw *et al.* 2002; Porcelli *et al.* 2002; Oates *et al.* 2005). In *E. coli*, small amounts of a third Tat complex have been purified. This complex contained TatA, TatB, and TatC but with a large molar excess of TatA (de Leeuw *et al.* 2002). The ratio of the Tat proteins in this complex more closely resembles that in the whole cell, as calculated from translational fusions or quantitative Western blotting (Jack *et al.* 2001; Sargent *et al.* 2001).

The isolation of different Tat protein complexes, together with experimental evidence from both the bacterial and cpTat pathways, has led to the idea of a modular and highly dynamic system whereby the Tat proteins exist as separate complexes in the resting state and come together to form a complete translocation pore upon substrate binding (Mori and Cline 2002; Alami *et al.* 2003). Such a mechanism is depicted in Figure 1.3. Complexes of varying sizes may be able to form pores that match the different sizes of folded Tat substrate proteins and thus minimize the space not occupied by the translocated protein, in turn restricting the formation of water-filled regions that can serve as conduits for ion leakage.

A high resolution three-dimensional image of a Tat complex was recently obtained using random canonical tilt electron microscopy. A complex consisting almost entirely of TatA was shown to form pore-like structures of variable sizes which appear gated/blocked on the cytoplasmic side of the inner membrane (Gohlke *et al.* 2005).

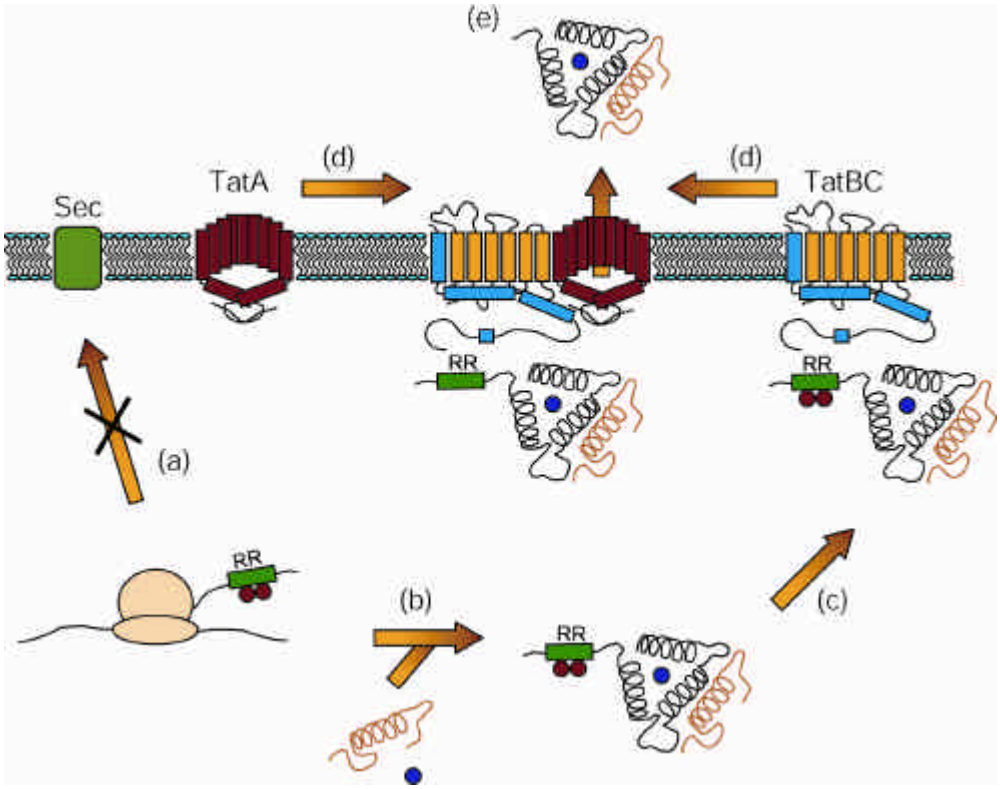

Figure 1.3. Model of Tat targeting and transport. In the membrane TatA is depicted in maroon, TatB in blue and TatC in gold. Upon emerging from the ribosome the pre-protein must avoid targeting to other pathways (a) such as Sec which is aided by the characteristics of the signal peptide and mature protein and/or the binding of Tat-specific chaperones (maroon circles). After folding, any cofactors and/or additional subunits are added (b) prior to targeting to the TatBC receptor complex (c). The proton motive force drives the formation of an active translocase (d) and the substrate is transported through a pore mainly consisting of TatA. Upon removal of the signal peptide the mature protein is released (e) on the periplasmic side of the membrane.

1.3.4 The nature of proteins exported via Tat

Bioinformatic analysis of signal peptides has been used extensively for the prediction of proteins that utilize the Tat apparatus (Rose *et al.* 2002; Dilks *et al.* 2003; Bendtsen *et al.* 2005). However, as with any computational method, predictions have to be interpreted with caution (Jongbloed *et al.* 2002).

Initially the Tat pathway was thought to have evolved for the translocation of complex redox proteins (Berks 1996). For example, in *Thermoplasma acidophilum* and *Thermoplasma volcanicum* all predicted Tat substrates are believed to be cofactor-containing proteins (Hutcheon and Bolhuis 2003). Similarly, in *E. coli* about 6% of all secreted proteins are predicted to be Tat-dependent and the majority are cofactor-containing redox proteins (Stanley *et al.* 2001; Rose *et al.* 2002; Palmer *et al.* 2005). They include hydrogenases, formate dehydrogenases, nitrate reductases, trimethylamine N-oxide (TMAO) reductases, and dimethyl sulfoxide (DMSO) reductases, which all function in anaerobic respiration (Gennis and Stewart 1996). For many of these enzymes, the synthesis and maturation of their cofactors occurs in the cytoplasm and they must be incorporated in a manner concomitant with folding (Berks *et al.* 2003; Palmer *et al.* 2005), rendering them incompatible with the Sec pathway.

Some cofactor-binding sites, such as those for flavin adenine dinucleotide or copper, are also found in proteins that are exported through the Sec pathway. Periplasmic copper proteins are intriguing because those with a single copper binding domain have thus far all been shown to be Sec-dependent. Preference seems to shift toward Tat export for substrates with additional and more complex copper binding sites, such as the *E. coli*

multi-copper oxidase CueO (Stanley *et al.* 2000; Berks *et al.* 2003; Ize *et al.* 2004). Unlike the many Tat substrates that are involved in anaerobic growth, CueO is required for copper homeostasis during aerobic growth (Ize *et al.* 2004; Singh *et al.* 2004).

At least eight proteins in *E. coli* seem to require export via the Tat pathway yet do not have a signal peptide of their own. Instead, they form a multimeric complex with another protein that contains a Tat signal peptide ("hitchhiker mechanism", see step b in Figure 1.3 above and ref. Rodrigue *et al.* 1999). Thus far, all of the proteins predicted to be exported in this way are also redox proteins and include: hydrogenases -1 and -2, formate dehydrogenases, and DMSO reductases. The prototypical example of this class of proteins is the hydrogenase-2 subunit HybC which is transported with its partner HybO in a process mediated by the signal peptide of the latter polypeptide (Rodrigue *et al.* 1999).

In *E. coli*, SufI, a prototypical Tat substrate of unknown function, and the periplasmic amidases AmiA and AmiC are neither exported as multimers nor contain cofactors. The genes *amiA* and *amiC* are particularly important for cell physiology. Inactivation of either gene or mislocalization of the corresponding protein due to the impairment of Tat export results in the formation of long chains of cells that are sensitive to the anionic detergent sodium dodecyl sulfate (SDS) (Bernhardt and de Boer 2003; Ize *et al.* 2003). Interestingly, microarray analysis revealed that the change in expression of 50% of the *E. coli* genes whose expression is affected by a deletion of *tatC* can be attributed to this cell envelope defect (Ize *et al.* 2004). Amidases are predicted to be Tat substrates in a variety of other bacteria, though knockouts do not always result in such

severe pleiotropic membrane defects (Bronstein, P. *et al.* 2004; Schaerlaekens *et al.* 2004; Caldelari *et al.* 2006).

Why have amidases and other proteins evolved to utilize the Tat pathway? After all, the transit times for export are considerably longer relative to export through Sec and the energetic cost may be greater (Stanley *et al.* 2000; Alder and Theg 2003a; Alder and Theg 2003b). The most widely accepted hypothesis is that these proteins may fold too quickly within the cytoplasm to be compatible with the Sec pathway. However, it was recently demonstrated that the co-translational signal recognition particle (SRP) pathway is quite capable of exporting soluble proteins that exhibit very rapid *in vitro* folding kinetics, such as the disulfide bond-forming enzyme DsbA (Huber *et al.* 2005; Shimohata *et al.* 2005). It remains to be determined whether proteins such as AmiA, AmiC and SufI can be targeted to the periplasm via an SRP-dependent process.

An alternative hypothesis is that certain Tat proteins may have to fold prior to export because they cannot reach their biologically active state within the periplasmic milieu. For example, GFP has been shown to be translocated via both the Sec and Tat pathways but is only fluorescent in the periplasm when exported by the Tat pathway (Feilmeier *et al.* 2000; Santini *et al.* 2001; Thomas *et al.* 2001). This finding suggests that some Tat substrates may be incompatible with folding in the periplasmic space. Along these lines it is also conceivable that the Tat pathway is in fact used for the export of *slowly* folding proteins. The periplasm presents a relatively harsh environment and proteins that fail to reach their native state quickly are targeted for degradation by DegP (Ruiz and Silhavy 2005). By allowing folding to proceed in the cytoplasm, slowly

folding proteins may be able to attain a protease-resistant conformation prior to being exposed to the periplasmic milieu. It should also be noted that protein folding in the cytoplasm offers access to chaperones such as GroEL that require ATP hydrolysis for their action. Interactions with GroEL/ES, DnaK/J/GrpE and other cytoplasmic chaperone networks may be required to prevent aggregation or assist in the assembly of certain Tat substrates. Rose *et al.* (2002) postulated that for *Haloferax volcanii* and other halophilic archaea, the high salt conditions of the extracellular milieu could lead to extensive aggregation and for this reason proteins of all types are folded first in the cytoplasm and then exported via the Tat pathway. However, the halophile *Salinibacter ruber* survives in high salt conditions yet this bacterium does not show such extreme dependence on the Tat pathway (Dilks *et al.* 2005). In any event, engagement of the Tat pathway by protein substrates is presumed to be dictated by folding considerations, but whether it is due to slow or rapid folding rates or the availability of appropriate chaperones is an open issue.

1.3.5 Energetics of Tat translocation

Originally, the chloroplast cpTat pathway was termed the ΔpH pathway, since protein translocation was shown to have no requirement for ATP as an energy source, but instead relied solely on the pmf (Cline *et al.* 1992). Similarly, transport of Tat substrates into *E. coli* inner membrane vesicles was shown to have no requirement for ATP and translocation was blocked upon dissipation of the membrane potential (Yahr and Wickner 2001; Alami *et al.* 2002). Alder and Theg (2003a) calculated the energy use of Tat-dependent translocation into isolated thylakoids and showed that on average, 7.9×10^4 protons (equivalent to ~10,000 ATP molecules) were released from the gradient per

protein translocated. For comparison, Sec-dependent translocation is less energetically costly and has been estimated to require one ATP molecule per 20 amino acid residues of the substrate translocated (Uchida *et al.* 1995). A reduction in Tat-dependent transport was observed in the presence of D_2O relative to H_2O, suggesting that proton transfer is involved at some stage, perhaps via a H+/protein counterflux principle (Musser and Theg 2000).

The requirement for the pmf under physiological conditions was challenged by recent *in vivo* data. Finazzi *et al.* (2003) concluded that Tat translocation becomes pH sensitive only when probed *in vitro*, possibly due to the loss of some critical intracellular factors. Di Cola *et al.* (2005) examined cpTat translocation *in vivo* in tobacco protoplasts in the presence of nigericin, valinomycin, or a combination of the two and concluded that the initial stages of Tat transport are not dependent on either component of the pmf. However, the authors point out that it is not possible to conclude from this data whether or not the pmf is required at other points in the translocation process.

1.3.6 Folding quality control

As will be discussed in great detail in Chapter 3, the Tat pathway exports proteins that are folded prior to transport, and unfolded proteins are not compatible with transport (Sanders *et al.* 2001; DeLisa *et al.* 2003) implying that somewhere between protein synthesis and export there must be steps where some sort of quality control is exerted on the folding state of the substrate. This is particularly important for a number of Tat substrates that require cofactor insertion or oligomerization with partner proteins in the cytoplasm prior to export.

Cytoplasmic chaperone-like proteins that function in the assembly of a number of cofactor and/or multimeric Tat substrates prevent targeting to the translocase until assembly is complete. The genes encoding these proteins are often found in the same operon as their corresponding Tat substrate proteins. These types of chaperones have recently been termed redox enzyme maturation proteins, or REMPs (Turner *et al.* 2004). It is not clear whether all Tat-dependent redox proteins have a corresponding REMP. However, homologues of the *E. coli* REMPs are found in many bacteria (Turner *et al.* 2004). Expression of the apparently REMP-less Tat substrate PhsABC from *S. enterica* LT2 in *E. coli* led to the production of active thiosulfate reductase (Hinsley and Berks 2002). This finding suggests that either a REMP is not required for the export of PhsABC or that *E. coli* REMPs are able to compensate for the lack of the native chaperone.

One well-studied group of REMPs is involved in the maturation of Tat substrates that are transported as multimeric complexes with only one of the subunits bearing a Tat signal peptide. For example, HyaE and HybE are chaperones that are required for the assembly of the hydrogenase components HyaA-HyaB and HybO-HybC, respectively. Only HyaA and HybO contain a Tat signal peptide and therefore the respective dimers have to form in the cytoplasm, prior to export (Rodrigue *et al.* 1999). HybE blocks the export of HybO without its hitchhiker partner HybC; without HybE the chaperone-mediated quality control mechanism is lost and transport of HybO is allowed prior to formation of the HybOC dimer (Jack *et al.* 2004).

1.3.7 The role of the Tat pathway in pathogenesis

Mutations in the *tat* genes have pleiotropic effects in *E. coli* including: (i) filamentation and changes in outer membrane permeability due to the mislocalization of AmiA and AmiC, leading to the impairment of biofilm formation (Stanley *et al.* 2001; Ize *et al.* 2004); (ii) inability to use various electron acceptors for growth under anaerobic conditions; (iii) defects in iron acquisition and copper homeostasis (Ize *et al.* 2004). Such growth defects would be expected to compromise the ability of pathogenic bacteria to colonize their host organisms. Indeed, in the food-borne pathogen *E. coli* 0157:H7, the Tat system has been shown to be indirectly important for virulence (Pradel *et al.* 2003). Similarly, a *tat* deletion strain of *Agrobacterium tumefaciens* was unable to induce tumor formation in plant tissue (Ding and Christie 2003).

The Tat pathway is specifically involved in the export of a variety of proteins that are directly involved in virulence. In the opportunistic human pathogen *Pseudomonas aeruginosa* two secreted virulence factors (both phospholipases) are mislocalized in *tat* mutants (Voulhoux *et al.* 2001). Phospholipases are involved in cleavage of phospholipids found in mammalian erythrocytes (Ostroff *et al.* 1990). The two *Pseudomonas* enzymes were shown to be transported across the outer membrane by the type II secretion system (Voulhoux *et al.* 2001), previously thought to be used only by proteins that use the Sec pathway (Cianciotto 2005). The Tat pathway of *P. aeruginosa* was also shown to be involved in the transport of several other virulence factors (Ochsner *et al.* 2002). Disruption of the Tat pathway abolishes virulence of pathogenic pseudomonae including *P. aeruginosa* and the plant pathogens *Pseudomonas syringae*

pathovars pv *tomato* DC3000 and pv *maculicola* ES4326 (Ochsner *et al.* 2002; Bronstein *et al.* 2005; Caldelari *et al.* 2006).

Recently, attention has been focused on the Tat system of *Legionella pneumophila*, a water-borne pathogen which is the causative agent of Legionnaires' disease. Although the Tat pathway has been shown to be important for biofilm formation, replication within the host, growth under iron-limiting conditions and secretion of phospholipase C activity, no Tat substrate directly involved in virulence of *L. pneumophila* has been identified so far (De Buck *et al.* 2005; Rossier and Cianciotto 2005). The Tat pathway is also thought to be required for *Helicobacter pylori* and other gut bacteria to become established in their host organism (Olson and Maier 2002).

The Tat pathway is directly involved in antibiotic resistance mechanisms of Mycobacteria. *Mycobacterium tuberculosis* and *Mycobacterium smegmatis tat* mutants were unable to transport active ß-lactamases (McDonough *et al.* 2005). In fact, *tatC* may be an essential gene for optimal growth in these bacteria (Sassetti *et al.* 2003; McDonough *et al.* 2005). Recently, using ß-lactamase as a reporter for Tat transport, McDonough *et al.* (2005) demonstrated that the phospholipase C proteins (PlcA and PlcB), important for the virulence of *M. tuberculosis* (Raynaud *et al.* 2002; McDonough *et al.* 2005), are exported via the Tat pathway.

1.4 Applications of Tat export in biotechnology

1.4.1 Protein expression

Expression in bacterial hosts is typically the first choice for production of recombinant proteins because bacteria grow quickly to high cell densities, do not require expensive media, and are relatively easy to work with. *E. coli*, which is by far the most widely used bacterial host, grows rapidly, can be easily manipulated genetically, and its cultivation is easily amenable to scale up. Engineering the export of recombinant proteins from the cytoplasm of *E. coli* offers many advantages in terms of bioprocessing and is generally preferred.

For many exported proteins, folding is thermodynamically coupled to disulfide bond formation. The export of disulfide-containing proteins via the Tat pathway can occur only in mutant strains that render the cytoplasm oxidizing (DeLisa *et al.* 2003). Kim *et al.* (2005) reported that a truncated form of tissue plasminogen activator containing nine disulfide bonds fused to a Tat signal peptide and expressed in an oxidizing cytoplasm *E. coli* strain resulted in a higher yield of active enzyme in the periplasm compared with a construct containing a Sec signal peptide. While this result is encouraging, it is not yet clear whether Tat export can result in high yields of periplasmic proteins, as is desired for preparative expression purposes. In *E. coli*, even native Tat substrate proteins saturate the export process when overexpressed from strong promoters (Yahr and Wickner 2001; DeLisa *et al.* 2004; Ize *et al.* 2004). Optimization of the growth conditions or co-expression of other proteins have been shown to partially alleviate the saturation of the export apparatus (Barrett *et al.* 2003; DeLisa *et al.* 2004; Li

et al. 2006). Still, protein yields for Tat-exported proteins are substantially lower than those obtained via Sec export, which often exceed 1 g/l. Therefore, additional fine tuning of cell physiology and fermentation conditions will be required before Tat export can be utilized for protein manufacturing.

1.4.2 Protein engineering

It is often desired to alter or enhance a particular property of a protein – for instance, one may want to increase the ability of an antibody to bind a specific antigen while decreasing its affinity for other structurally similar molecules, or alter the specificity of an enzyme toward various substrates. Protein engineering has not yet advanced to the point of being able to determine exactly which amino acids need to be changed to achieve a particular outcome. Instead, a strategy known as directed evolution is often employed (Cirino *et al.* 2003). The first step of this strategy involves the generation of a diverse pool, or library, of polypeptides that contain amino acid substitutions scattered throughout the protein of interest (Arnold and Georgiou 2003). For example, random mutations can be introduced into the gene encoding the protein of interest via polymerase chain reaction (PCR), either throughout the gene or in a key area known or predicted to affect the property to be altered (You and Arnold 1996; Cirino *et al.* 2003). Further diversity may be obtained by "DNA shuffling", an *in vitro* process that mimics natural evolutionary processes at an accelerated rate by recombining mutated genes with each other (Stemmer 1994). The result is a combinatorial library – a pool of genes encoding proteins containing the various possible combinations of changes.

The second major step in any directed evolution strategy is the screening of each member of this library of mutants for the desired function, a difficult task given the size of a typical combinatorial library (Arnold and Georgiou 2003). It is not uncommon to have a library made up of more than 10^9 different versions of a gene (Barbas *et al.* 1991). Analysis of each of the resultant proteins via traditional methods to find the one with the desired or enhanced function would be time, labor, and cost prohibitive. Therefore, efforts have been focused on developing display platforms such that a library of mutant proteins can be investigated more efficiently.

A display platform is a technology in which a library of proteins is displayed on the surface of cells or viruses to make each protein more accessible for further study in a high throughput manner. Protein libraries can be presented on the outer membrane, inner membrane, or in the periplasm of cells using various display technologies (Arndt *et al.* 2000; Chen and Georgiou 2002, Harvey *et al.* 2004). Once a protein reaches the desired location, it can be further analyzed for an improved property. For example, affinity can be tested via methods that detect ligand-protein binding interactions (Chen and Georgiou 2002). Maintaining a link between the protein and the cell – which contains the DNA sequence encoding the protein – is necessary in order to keep the link between genotype and phenotype. If this link is lost, an improved protein may be found but there is no way to determine what change at the amino acid or nucleotide level resulted in this improvement, and thus no way to reproduce it.

Several display platforms have been developed which maintain a genotype/phenotype link without keeping the protein of interest in or on the bacterial cell.

Filamentous phage display is one such platform, permitting accessible screening of libraries with over 10^8 members (Barbas *et al.* 1991). Filamentous phage have a cylindrical coating surrounding the single-stranded viral genomic DNA and consist of a major coat protein (p8) and a total of four minor coat proteins (p3, p6, p7, and p9) (Figure 1.4A; ref. Russel *et al.* 1997). In filamentous phage display, the gene for the protein of interest is cloned into a vector known as a phagemid (an extra-chromosomal, circular piece of double stranded DNA) and it is fused to the gene for a suitable phage protein. Fusions to gene III (which encodes the p3 coat protein mentioned above) are typically used to display desired polypeptides on the surface of filamentous phage (Winter *et al.* 1994; Hoogenboom *et al.* 1998; Benhar 2001). The resulting fusion proteins are exported via the Sec pathway and are then incorporated into the phage particle during the bacteriophage assembly (Figure 1.4B; ref. Benhar 2001). The phagemid also becomes preferentially packed into the phage particles, in place of the single stranded DNA encoding the phage genome, during the assembly process (Benhar 2001). This occurs because mutations have been engineered into the region of the genomic helper phage DNA such that the assembly machinery does not recognize it. At the same time, the phagemid contains the exact DNA sequence recognized by the assembly machinery. Thus, the bacteriophage particle is assembled with the fusion protein attached to its outer coat, making it an easily accessible target for detection, while the DNA encoding the protein fusion is inside that same particle. Phage panning procedures (an example is shown in Figure 1.4C) are typically used to screen a protein library displayed on phage. In principle a single phage particle may be isolated via the action of the displayed protein

during phage panning, and then its DNA can be amplified and sequenced to reveal the amino acid sequence of the displayed protein.

Both phage and bacterial display require translocation of the polypeptide from the cytoplasm and this is typically accomplished using a Sec signal peptide. However, for the reasons discussed above, certain proteins cannot be exported via the Sec pathway and therefore cannot be displayed on filamentous phage particles. Even if a protein is compatible with Sec export, mutant forms of the protein, such as those that arise by random mutagenesis, may exhibit altered folding characteristics that might prevent their display. The Tat pathway may be useful for the export and display of such proteins and thus expand the diversity of sequences within the library.

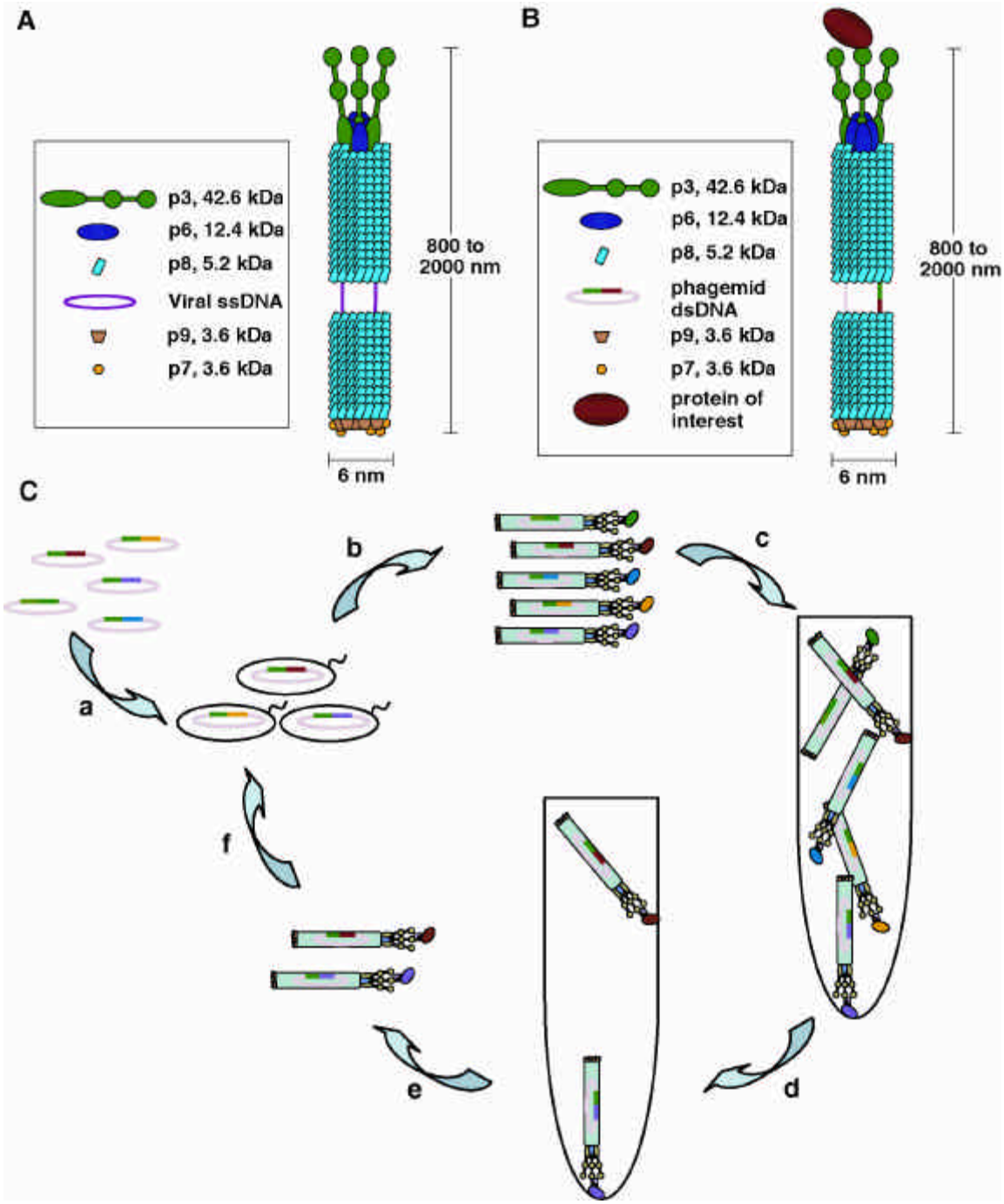

Figure 1.4. Filamentous phage, phage display, and phage panning. **(**A) Diagram of a filamentous phage. (B) Diagram of filamentous phage display. Note the phagemid DNA is now substituted for the viral DNA, and the protein of interest is now fused to p3. (C) Example of a phage panning process. A phagemid library is transformed into *E. coli* (a) and helper phage is added to assemble phage particles displaying the resulting library of proteins (b). The phage is incubated in a tube coated with a desired ligand (c) and phage is washed away if the displayed protein does not bind to the ligand coated on the tube (d). The bound phage is eluted (e) and the process is repeated until the eluted phage are sufficiently enriched with those phage displaying proteins that have the ability to bind the desired ligand.

1.5 Research Objectives

Due to its unique ability to export folded proteins across the bacterial inner membrane, the Tat pathway shows great promise for use in biotechnological applications. However, much remains to be learned about the mechanism of Tat export at a basic science level before it can be manipulated for practical purposes.

At the time this work was begun, only a few Tat substrate proteins had been confirmed, though many had been predicted via bioinformatics searches across different species. Since the utilization of the Tat pathway for both protein expression and protein engineering purposes hinges on uncovering an optimal and well-studied Tat signal peptide, the first goal of this research became to experimentally verify the predicted Tat signal peptides encoded by the *E. coli* genome. To carry this out, reporter proteins were fused to each putative Tat signal peptide and appropriate assays were performed (Chapter 2). Surprisingly, a number of the signal peptides tested were able to direct export to both the Sec and Tat translocons, and therefore it also became necessary to ascertain which features of the signal peptides determine promiscuity.

The ability of the Tat pathway to export fully folded proteins had not been directly proven prior to the work described in Chapter 3. Briefly, fusions to the reporter protein alkaline phosphatase or to a recombinant single chain antibody fragment were constructed. Both proteins contain cysteines that must form disulfide bonds for proper folding to occur. Therefore, the constructs were expressed in wild type and oxidizing cytoplasm strains, and assays were carried out to determine if proteins were exported to the periplasm. This work confirmed that the Tat pathway exports folded proteins, but

more importantly, it demonstrated that the pathway has an intrinsic quality control mechanism to prevent the export of unfolded or misfolded polypeptides.

One goal of this research was to develop and employ Tat-based systems for protein engineering purposes. A system for the phage display of proteins that are exported via the Tat pathway is described in Chapter 4. Furthermore, the use of this technology to develop molecules with the ligand-binding affinity of an antibody combined with the intrinsic fluorescence of GFP is described.

Chapter 2

Export Pathway Selectivity of *Escherichia coli* Tat Signal Peptides

2.1 Introduction

As mentioned in the first chapter, bacterial cells use signal peptides to direct proteins to the periplasm of the cell via the proper export pathway. Signal peptides are amino acid extensions at the N-terminus of the polypeptide that function as molecular "zip codes", telling the cellular machinery what cellular compartment the proteins need to be directed to. Following export, the signal peptide is cleaved from the rest of the protein at a recognition site at the C-terminus by a leader peptidase, giving rise to the mature protein (von Heijne 1985; Izard and Kendall 1994). To utilize the bacterial Tat pathway for biotechnological applications, it would be useful to identify an optimal bacterial Tat signal peptide. In addition, it is important to fully understand the properties of Tat signal peptides, and to understand how targeting to other export pathways is avoided.

Bacterial Tat signal peptides are longer than signal peptides that target proteins to the Sec or to the signal recognition particle (SRP) pathways (Cristobal *et al.* 1999). The amino-terminal (n-) region contains the Tat consensus motif (S/T)-R-R-X-F-L-K, which includes the signature twin-arginine dipeptide. The hydrophobic (h-) region of Tat signal peptides is generally less hydrophobic than the respective region of Sec (and even less so

than SRP-specific) signal peptides (Cristobal *et al.* 1999). Finally, the carboxyl-terminal (c-) region often contains basic residues, in contrast to Sec signal peptides, which are almost never charged in this region. The mean charge of the c-region of Tat signal peptides is +0.5 versus +0.03 for Sec signal peptides (Berks *et al.* 2003).

Bioinformatic analyses of protein sequences for the presence of putative Tat signal peptides suggest that the utilization of the Tat pathway varies widely between different organisms of the same phylum (Robinson and Bolhuis 2001; Rose *et al.* 2002; Dilks *et al.* 2003; Bendtsen *et al.* 2005). In bacteria, polypeptides exported via the Tat pathway include redox enzymes, virulence factors, periplasmic ligand-binding proteins and enzymes involved in cell envelope biogenesis (see Section 1.3.4). The genome of *E. coli* K-12 is predicted to encode between 22-34 Tat signal peptides (Robinson and Bolhuis 2001; Dilks, K. *et al.* 2003; Bendtsen *et al.* 2005). However, only 15 have been confirmed experimentally (Santini *et al.* 1998; Sargent *et al.* 1998; Weiner *et al.* 1998; Rodrigue *et al.* 1999; Gon *et al.* 2000; Sambasivarao *et al.* 2000; Stanley *et al.* 2000; Stanley *et al.* 2002; Bernhardt and de Boer 2003; Ize *et al.* 2003; Ize *et al.* 2004; Lequette *et al.* 2004; Sturm *et al.* 2006). In general, identifying Tat substrates has proven difficult. Jongbloed *et al.* (2002) have shown that existing bioinformatics methods greatly overestimate the number of Tat proteins in *Bacillus subtilis.* Proteomic studies based on comparison of secreted proteins in wild type (wt) and *tat* deletion strains are complicated because many putative Tat proteins are expressed only under specialized conditions (Andersson and von Heijne 1991; Jongbloed *et al.* 2002; Ize *et al.* 2003; Caldelari *et al.* 2006). In addition, data from the localization of reporter protein fusions to Tat signal

peptides have to be interpreted with caution because the nature of the reporter protein can misdirect the fusion to the Sec pathway (Blaudeck *et al.* 2001; Stanley *et al.* 2002; Blaudeck *et al.* 2003).

In this study, we evaluated the export pathway preference of the putative Tat signal peptides of *E. coli* by examining the localization of fusions to three complementary reporter proteins. This analysis, and additional localization studies of the respective epitope-tagged full-length proteins, revealed that the *E. coli* genome encodes at least 27 Tat-targeting signal peptides. Surprisingly, the majority of these signal peptides can direct export via either Tat or Sec depending on the export competence of the reporter polypeptide. We show that increasing the positive charge in the N-terminus of the mature protein prevents export via Sec, without affecting Tat export. Thus, it appears that in *E. coli* the Tat pathway is utilized for the translocation of proteins containing positively-charged N-terminal regions which, as had been established earlier, serve as "stop transfer" signals (Coleman *et al.* 1985; Yamane and Mizushima 1988) for Sec export.

This study was carried out in collaboration with Dr. Yasuaki Kawarasaki, Pooya Iranpour, and Brian Ribnicky (Tullman-Ercek Submitted). Dr. Kawarasaki grew cells and performed cell fractionations and Western blots on FLAG-tagged proteins, Pooya Iranpour assisted with the flow cytometric analysis and the cloning of the PhoA constructs, and Brian Ribnicky assisted in the cloning of the GFP constructs.

2.2 Materials and Methods

2.2.1 Bioinformatics

Putative Tat signal peptides were identified by conducting a BLAST search using the strings: SRRRFLK, SRRXFLX, TRRXFLX, SRRXXLK, SRRXXLA, TRRXXLK, TRRXXLA, SRRXXLT, SRRXXIK, SRRXXIA, SRRXFIX, SRRXFMK, SRRXFVK, SRRXFVA, SRRQFLK, RRXFLA, and RRXFLK within the first 50 residues. Sequences were then analyzed for signal peptide characteristics using SignalP 1.1 (http://www.cbs.dtu.dk/services/SignalP/) (Nielsen *et al.* 1997). DmsA and YaeI, which had apparently been assigned incorrect start codons in the NCBI database (http://www.ncbi.nlm.nih.gov/), were also included in the list, as was YcdO, which did not match our criteria but had been predicted to be a Tat substrate (Bendtsen *et al.* 2005).

2.2.2 Strains and plasmids

The bacterial strains and plasmids used in this study are listed in Table 2.1.

Signal peptides were amplified from *E. coli* XL1-Blue genomic DNA by PCR using the primers 1-56 shown in Appendix Table A.1 and cloned into pKKGS (DeLisa *et al.* 2002), a derivative of plasmid pBAD33 (Guzman *et al.* 1995). While SignalP predicts the likely signal peptidase cleavage site, there is evidence that such predictions are not always accurate for Tat signal peptides (Thomas et al. 1999), and therefore reverse primers were designed to include the sequence encoding the first six to eight aa following

the predicted cleavage site and ensure the actual cleavage site would be present in the fusions.

Table 2.1 Strains and plasmids used in this study

E. coli Strain or Plasmid	Relevant genotype or features	Source
XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI*^{q} *ZΔM15* Tn*10* (Tet^{r})]	Lab collection
MC4100	F- Δ*lacU169 araD139 rpsL150 relA1 ptsF rbs flbB5301*	Lab collection
B1LK0	MC4100 *tatC*	Bogsch *et al.* 1997
MC4100-P	MC4100 *pcnB:Kan*	Buchanan *et al.* 2002
B1LK0-P	B1LK0 *pcnB:Kan*	Buchanan *et al.* 2002
HS3018	MC4100 *malT*(Con)*-1* Δ*malE444*	Davidson *et al.* 1992
HS3018 *tatC*	HS3018 *tatC:Spec*	This study
HS3018 *tatB*	HS3018 *tatB:Kan*	This study
HS3018 *tatA*	HS3018 *tatA*	This study
HS3018 *tatE*	HS3018 *tatE*	This study
DHB4	MC1000 ?*phoAPvuII phoR* ?*malF3* [F' *lac-pro lacI*^{q}]	J. Beckwith
pTrc99A	*trc* promoter, ColE1 *ori*, amp^{r}	Amersham Biosciences
pBAD33	*araBAD* promoter, pACYC184 *ori*, cm^{r}	Guzman *et al.* 1995
pBAD18Cm	*araBAD* promoter, pBR322 *ori*, cm^{r}	Guzman *et al.* 1995
pCueO-FLAG	CueO-FLAG in pBAD33	This study
pNrfC-FLAG	NrfC-FLAG in pBAD33	This study
pMdoD-FLAG	MdoD-FLAG in pBAD33	This study
pSufI-FLAG	SufI-FLAG in pBAD33	This study
pYdhX-FLAG	YdhX-FLAG in pBAD33	This study
pYfhG-FLAG	YfhG-FLAG in pBAD33	This study
pssTat-GS	ssTat-GFP-SsrA in pBAD18Cm	This study
pssTat-MBP	ssTat-Δ(1-26)-MalE in pBAD18Cm	This study
pssTat-AP	ssTat-Δ(1-22)-AP in pBAD18Cm	This study

The gene fragments encoding *ssTat-GFP-SsrA* fusions were sub-cloned into pBAD18-Cm (Guzman *et al.* 1995). All primers can be found in Appendix Table A.1. To construct the *ssYaeI-GFP-SsrA* fusion, *GFP-SsrA* was first sub-cloned into pBAD18-Cm and then the *ssYaeI* fragment was inserted using *Xma*I and *Xba*I restriction sites. The set of MBP fusions was constructed by replacing the *GFP-SsrA* gene in pBAD18-Cm with the *E. coli malE* gene

amplified with primers 57-58 flanked by in-frame *Xba*I-*Hin*dIII sites. To construct the *ssHybO-malE* fusion, the *GFP-SsrA* fusion in pBAD18-Cm was excised with *Xba*I and *Sph*I and replaced with an *Xba*I-*Sph*I *malE* gene fragment amplified by primers 58-59. Subsequently, the *ssTat-malE* genes were amplified with primers 60-87 and sub-cloned into the *Nco*I and *Hin*dIII (or *Eco*RI and *Sph*I in the case of *ssHybO-malE* and *Eco*RI and *Hin*dIII in the case of *ssYnfF-malE*) sites of pTrc99a (Amersham) so that induction of gene expression can be performed with isopropyl-β-D-thiogalactopyranoside (IPTG), rather than arabinose, to enable selection on maltose plates. Construction of the respective PhoA fusions was similarly carried out by removing an *Xba*I- *Hin*dIII fragment encoding the *malE* fusion in pTrc99a and substituting it with an in-frame *Xba*I-*Hin*dIII *phoA* gene (or an *Xba*I-*Sph*I *phoA* gene in the case of the *ssHybO* fusion) amplified from *E. coli* via primers 88-90. To make wtPhoA, *ssPhoA* was amplified with primers 91-92 and cloned into the *Nco*I and *Xba*I sites of empty pTrc99a, followed by sub-cloning the *Xba*I-*Hin*dIII *phoA* fragment into the corresponding pTrc99a sites. Similarly, the altered-charge versions of the *ssMdoD* fusions were constructed using primers 70 and 93-94 by cloning the altered *ssMdoD* fragment into the *Nco*I and *Xba*I sites of empty pTrc99a, followed by sub-cloning of the *Xba*I-*Hin*dIII *malE* or *phoA* fragment into the appropriate sites.

Full-length proteins were tagged with a C-terminal FLAG (DYKDDDDK) epitope tag (Sigma) by PCR amplification (Table A.1) and cloning into the *Sac*I and *Xba*I sites of pBAD33 (Guzman *et al.* 1995).

To construct HS3018 *tatE*, the tatE deletion allele from pFAT44 (Sargent *et al.* 1998) was moved onto the chromosome of HS3018 by the method of Hamilton *et al.* (1989). Likewise, to construct HS3018 *tatA*, the tatA deletion allele described in Sargent *et al.* (1999) was moved onto the chromosome of HS3018 by the method of Hamilton *et*

al. (1989). To construct HS3018 *tatC*, the ΔtatC::Spec allele from strain BUDDY (Stanley *et al.* 2000) was moved by P1 transduction into HS3018. Similarly, to construct HS3018 *tatB*, the ΔtatB::Kan allele from strain MCMTA (Chanal *et al.* 1998) was moved by P1 transduction into HS3018.

2.2.3 Cell fractionations

Cell fractionations were performed by the cold osmotic shock procedure. Briefly, cells were harvested by centrifugation and resuspended in 7.5 ml fractionation buffer (30 mM Tris-HCl pH 8.0, 20% w/v sucrose, 1 mM ethylenediaminetetraacetic acid). After 10 minutes at 25°C, the cells were centrifuged at 6000 g for 15 minutes at 4°C. The pellet was then resuspended in 1 ml of 5 mM $MgSO_4$, ice-cold, and left on ice for 10 minutes. After centrifuging at 13,000 g for 15 minutes at 4°C, the supernatant was kept as the periplasmic fraction. The soluble cytoplasmic fraction was obtained by resuspending the cell pellet obtained after fractionation into 1 ml of phosphate buffered saline (PBS), sonicating for 30 seconds, and recentrifuging to precipitate and remove insoluble cell debris.

2.2.4 Growth conditions, expression, and reporter protein assays

Unless otherwise noted, cells were grown at 37°C on LB media with 50 μg/ml chloramphenicol, 25 μg/ml kanamycin or 100 μg/ml ampicillin, as appropriate. To test growth on maltose, *E. coli* HS3018 or its derivatives containing plasmids encoding MBP

fusions were grown overnight, diluted, plated on M9 minimal media containing 0.4% maltose and incubated at 37°C for two to three days.

PhoA assays were performed by subculturing in triplicate into M9 minimal media containing 0.1 mM IPTG, 0.2% casein, and 0.2% glucose and growing for 4 hours at 30°C. Each culture was then mixed for 30 minutes in a 2:3 ratio (v:v) with a 2:1 (v:v) mixture of B-PER solution (Pierce) and 0.4 M iodoacetamide, to release the cellular contents without additional oxidation of free cysteines. Normalized amounts of this lysates solution were then added to 200 µl of para-nitrophenyl phosphate (pNPP) solution (5.0 mg tablet of pNPP from Sigma in 25 ml 0.2 M Tris-HCl pH 8.0). Absorbance at 405 nm was measured at regular time intervals for up to 1 hour.

Western blotting was performed as described earlier with slight changes (Levy *et al.* 2001). Briefly, fractionation or lysate samples were mixed with equal volumes of 2xSDS loading buffer containing 2-mercaptoethanol. The samples were boiled for 5-10 minutes and loaded and run on 10% Tris-glycine precast gels (Cambrex). Proteins were transferred to a polyvinylidine fluoride membrane using a semidry transfer graphite electroblotting system (Millipore). The membranes were blocked in 5% nonfat dry milk/TBST buffer (15 mM Tris-HCl, 4 mM Tris, 137 mM NaCl, 0.1% Tween-20, to pH 7.6) for 1 hour at room temperature or overnight at 4°C. Antibodies diluted in 1% milk/TBST were incubated at room temperature for 1 hour. The following primary antibodies were used: monoclonal mouse anti-FLAG-M2 (Sigma), monoclonal rabbit anti-DsbA (Stressgen), monoclonal rabbit anti-GroEL (Sigma), polyclonal rabbit anti-PhoA (Rockland), and monoclonal mouse anti-MalE (Sigma).

For flow cytometric analysis, *E. coli* MC4100-P and B1LK0-P containing plasmids encoding GFP-SsrA fusions were grown overnight in LB media as described above and 500 μl of each overnight culture were used to inoculate 10 ml of fresh media. After 1 h shaking at 37°C, gene expression was induced with arabinose to a final concentration of 0.01%, the cells were incubated for an additional 4 hours and 1 ml samples were harvested by centrifugation, diluted in PBS containing 5 μg/ml propidium iodide and analyzed with a Becton-Dickinson FACSort flow cytometer.

2.3 Results

2.3.1 Bioinformatic identification of putative Tat signal peptides in *E. coli*

Putative Tat signal peptides were identified by first using BLAST to search for *E. coli* open reading frames encoding proteins that contained, in the first 50 aa, a twin-arginine dipeptide and, in addition, any two of the additional four aa that comprise the conserved Tat motif. Hits were then analyzed for signal peptide characteristics using SignalP (Nielsen *et al.* 1997). A total of 26 sequences meeting =25% of the SignalP 1.1 threshold values were found. Two more were included that have misannotated start codons in the database but otherwise fit the above criteria (Table 2.2). Furthermore, YcdO was also included because it is in the same operon as another putative Tat protein and was previously predicted to be a Tat substrate itself (Bendtsen *et al.* 2005). The final set contained all 15 confirmed Tat substrates (AmiA, AmiC, CueO, DmsA, FdnG, FhuD, HyaA, HybO, MdoD, NapA, NrfC, SufI, TorA, TorZ, and YcdB) and is in good agreement with the prediction from program TATFIND (Dilks *et al.* 2003). However, TATFIND does not predict AmiC, DmsA, YaeI, YcdO, and YfhG to be Tat substrates and in addition it predicts plasmid-encoded and membrane proteins that were not considered in the present study. Another program used to identify Tat signal peptides, TatP, is also in good agreement with our list but the latter missed FhuD, YagT, YcdO, and YfhG and predicted b3000, which was not considered in this study (Bendtsen *et al.* 2005). While SignalP predicts signal peptidase cleavage sites, there is evidence that such predictions are not always accurate for Tat signal peptides (Thomas *et al.* 1999). To

eliminate ambiguities, all signal peptide fusions in this study were designed to include the first six to eight aa following the SignalP predicted cleavage site (Table 2.2).

Table 2.2. Characteristics, reporter protein fusion assay results, and assigned export pathway for each putative Tat signal peptide

Signal Peptide	Sequence*	Hydro-phobicity, h-region†	Charge, c-region	Charge, mature N-term	Charge, net (cleavage area)	GFP-SsrA FACS Mean Fluorescence‡ wt	GFP-SsrA FACS Mean Fluorescence‡ *tatC*	PhoA Activity (%)§	MBP Plate Growth¶ wt	MBP Plate Growth¶ *tatC*	Export Pathway
FdnG	MDVSRRQFFKICAGGMAGTTVAALGFA**PKQALA***QARNYKL*	2.25	+1	+2	+3	2	3	n.e.	+	-	Tat
FdoG	MQVSRRQFFKICAGGMAGTTAAALGFA**PSVALA***ETRQYKL*	1.91	0	+1	+1	2	4	n.e.	+	-	Tat
NapG	MSRSAKPQNGRRRFLRDVVRTAGGLAAVGVAL**GLQQQTARA***SGVRLR*	2.00	+1	+2	+3	2	2	n.e.	+	-	Tat
NrfC	MTWSRRQFLTGVGVLAAV**SGTAGRVVA***KTLNIN*	2.28	+1	+1	+2	9	3	n.e.	+	-	Tat
HyaA	MNNEETFYQAMRRQGVTRRSFLKYCSLAATSLGLGAGMA**PKIAWA***LENKPR*	1.36	+1	+1	+2	27	3	0	+	-	Tat
YnfE	MSKNERMVGISRRTLVKSTAIGSLALAA**GGFSLPFTLRNAAA***AVQQAREK*	2.01	+1	+1	+2	9	3	0	+	-	Tat
WcaM	MPFKKLSRRTFLTASSALAFL**HTPFARA***LPARQS*	1.82	+1	+1	+2	13	3	1	+	-	Tat
TorA	MNNNDLFQASRRRFLAQLGGLTVAGMLGPSLL**TPRRATAAQA***ATDA*	1.42	+2	-1	+1	90	2	2	+	-	Tat
NapA	MKLSRRSFMKANAVAAAAAAA**GLSVPGVARA***VVGQ*	1.53	+1	0	+1	2	3	3	+	-	Tat
YagT	MSNQGEYPEDNRVGKHEPHDLSLTRRDLIKVSAATAATAVV**YPHSTLAA***SVPAA*	1.63	0	0	0	57	5	3	+	-	Tat
YcbK	MDKFDANRRKLLALGGVALGAAIL**PTPAFA***TLSTPR*	2.40	0	+1	+1	82	4	3	+	-	Tat
DmsA	MKTKIPDAVLAAEVSRRGLVKTTAIGGLAMASSAL**TLPFSRIAHA***VDSAIP*	1.56	+1	-1	0	90	6	2	+	+	Tat+Sec
YdhX	MSWIGWTVAATALGDNQMSFTRRKFVLGMGTVIFF**TGSASSLLA***NTRQEK*	2.31	0	+1	+1	7	5	2	+	+	Tat+Sec‖
YahJ	MKESNSRREFLSQSGKMVTAAALF**GTSVPLAHA***AVAGTL*	2.17	0	0	0	100	6	3	+	+	Tat+Sec
YedY	MKKNQFLKESDVTAESVFFMKRRQVLKALGISATAL**SLPHAAHA***DLLSWF*	1.73	0	-1	-1	47	6	4	+	+	Tat+Sec
CueO	MQRRDFLKYSVALGVASALPLW**SRAVFA***AERPTL*	1.79	+1	0	+1	21	4	5	+	+	Tat+Sec
SufI	MSLSRRQFIQASGIALCAGAV**PLKASA***AGQQQP*	1.74	+1	0	+1	29	2	5	+	+	Tat+Sec
YcdB	MQYKDENGVNEPSRRRLLKVIGALALA**GSCPVAHA***QKTQSA*	2.66	0	+1	+1	123	5	6	+	+	Tat+Sec
TorZ	MIREEVMTLTRREFIKHSGIAAGALVVTSAA**PLPAWA***EEKGGK*	1.98	0	0	0	56	4	6	+	+	Tat+Sec
HybA	MNRRNFIKAASCGALLTGAL**PSVSHA***AAENRPP*	1.98	0	0	0	33	3	11	+	+	Tat+Sec
YnfF	MMKIHTTEALMKAEISRRSLMKTSALGSLALASSAF**TLPFSQMVRA***AEAPVE*	1.55	+1	-2	-1	49	5	15	+	+	Tat+Sec
HybO	MTGDNTLIHSHGINRRDFMKLCAALAA**TMGLSSKAAA***EMAESV*	2.47	+1	-2	-1	28	3	17	+	+	Tat+Sec
AmiC	MSGSNTAISRRRLLQGAGAMWLL**SVSQVSLAA***VSQVV*	1.40	0	0	0	36	9	18	+	+	Tat+Sec
AmiA	MSTFKPLKTLTSRRQVLKAGLAAL**TLSGMSQAIA***KDELLK*	2.10	0	0	0	22	2	79	+	+	Tat+Sec
YfhG	MRHIFQRLLPRRLWLAGLPCLA**LLG***CVQNHNKPAIDT*	2.16	0	0	0	21	5	93	+	+	Tat+Sec
MdoD	MDRRRFIKGSMAMAAVCGTSGIASLF**SQAAFA***ADSDIA*	2.42	0	-2	-2	15	5	165	+	+	Tat+Sec
FhuD	MSGLPLISRRRLLTAMALSPLLW**QMNTAHA***AAIDPN*	1.70	0	-1	-1	6	2	256	+	+	Tat+Sec‖
YaeI	MISRRRFLQATAATIATSSGFG**YMHY***CEPGWFELIRHRLAFFKDNAAPFKIL*	2.80	0	+1	+1	2	2	20	+	+	Sec
YcdO	MTINFRRNALQLSVAALFSSAFM**ANA***ADVPQVK*	1.63	0	0	0	5	4	107	+	+	Sec

*Shown in bold is the signal peptide c-region and in italics is the predicted n-terminal region of the mature protein.

†Average hydrophobicity using Kyte Doolittle values is calculated for the h-region, defined as described by Cristóbal *et al.* (1999).

‡Signal peptide-GFP-SsrA fusions were expressed both in wild type and *tatC* strains and whole cell fluorescence values were measured by flow cytometry (FACS). The average mean fluorescence values of at least three replicate experiments are shown.

§Activity of whole cell lysates of a *phoAR* strain expressing each signal peptide-PhoA fusion as percentage of activity of PhoA with its own signal peptide; n.e. = no expression detected.

¶Colonies present (+) or absent (-) after two days at 37°C on maltose minimal media plates in *malE*(wt) and *malE tatC*(ΔtatC) strains. For ssDmsA and ssSufI colonies present after three days in a ΔtatC strain.

‖Tat portion of assignment based on epitope tagging for ssYdhX and on the results of Ize *et al.* (2004) for ssFhuD; Sec portion of assignment based on reporter fusion results for both ssYdhX and ssFhuD.

2.3.2 Export pathway analysis using reporter fusions

We sought to experimentally verify the export pathway preference of the putative 29 signal peptides above by systematic analysis using reporter fusions (Figure 2.1). DeLisa *et al*. (2002) have shown that a short-lived version of GFP containing an SsrA tag confers cell fluorescence only when it is exported to the periplasm by virtue of a Tat signal peptide. In the absence of export, GFP-SsrA is rapidly degraded by ClpXP and ClpAP in the cytoplasm, whereas targeting to the Sec apparatus also results in non-fluorescent cells. For these reasons, the fluorescence of cells expressing fusions between Tat signal peptides and GFP-SsrA is proportional to the efficiency of export into the periplasm (DeLisa *et al.* 2004).

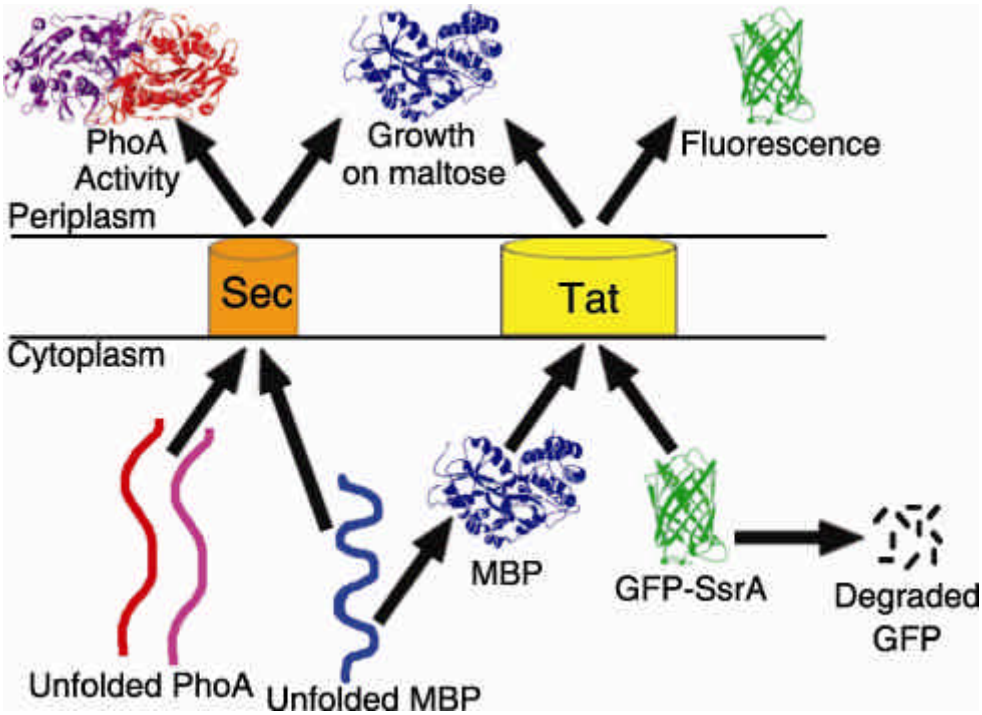

Figure 2.1. Routing of PhoA, MBP, and GFP-SsrA via the Sec and Tat pathways. PhoA and MBP are exported via Sec in an unfolded conformation while MBP and GFP-SsrA are exported via Tat in a folded conformation. The SsrA tag targets GFP for degradation by ClpXP in the cytoplasm.

GFP-SsrA fusions to the set of 29 putative signal peptides were expressed under the control of the pBAD promoter. Since high-level expression of SsrA-tagged proteins can saturate the ClpXP and ClpAP machinery, the fusions were expressed in *E. coli* MC4100 carrying a *pcnB* deletion that lowers the copy number of pBR322 and its derivatives (Lopilato *et al.* 1986). Cell fluorescence was analyzed by flow cytometry (Appendix Figure A.1) and the mean fluorescence values conferred by the 29 fusions are reported in Table 2.2. Signal peptides were classified as capable of mediating Tat export if they conferred a mean cell fluorescence =10 absorbance units and the cell fluorescence decreased by more than 3-fold in an isogenic *tatC* strain. These criteria were derived from genetic studies in which we isolated TatC suppressor mutations that allow the export of twin-lysine signal peptides (E. Strauch and G. Georgiou, unpublished data). Nineteen signal peptides satisfied these criteria (Table 2.2). However, GFP-SsrA fusions to several experimentally confirmed Tat signal peptides (ssFdnG [signal sequence FdnG], ssFhuD, ssNapA, and ssNrfC), resulted in low cell fluorescence. Low whole cell fluorescence can result from poor expression and does not necessarily imply absence of Tat export.

Maltose binding protein (MBP) is naturally a Sec substrate in *E. coli* but is also competent for Tat export (Figure 2.1; ref. Blaudeck *et al.* 2003). A low level of periplasmic MBP is sufficient to allow growth of *malE* cells on minimal media with maltose as the sole carbon source. We reasoned that growth on maltose minimal media might be a more sensitive reporter system for Tat export of fusions that are expressed

poorly. However, Tat export of signal peptide-MBP fusions can only be inferred if a *tatC* mutation abolishes growth on maltose, thus ruling out the possibility that the fusion protein is instead exported via Sec. Eleven of the 29 fusions conferred growth on maltose only in wt cells (Table 2.2). These included ssFdnG, ssFdoG, ssNapA, ssNapG, ssNrfC, and ssYnfE, whose Tat export could not be assigned with GFP-SsrA fusions due to low fluorescence. To ensure that lack of growth in a *tatC* mutant strain was not due to lower protein expression levels, fusions to the six signal peptides above were further tested using maltose minimal media plates containing inducer. Nonetheless, Tat-dependent growth was still observed only in wt cells and not in the *tatC* mutant strain. Western blotting revealed that expression levels are nearly equal in wt and *tatC* mutant cells (with the exception of ssTorA, for which expression was higher in the *tatC* mutant relative to the wt cells). Representative results are shown in Figure 2.2C. The ssNrfC-MBP fusion could not be detected by Western blotting in either the wt or *tatC* strain, despite its ability to confer Tat-dependent growth on maltose plates with and without inducer. Collectively, these data indicate that the set of 11 signal peptides is capable of directing export via Tat but not via the Sec pathway.

We tested whether the *mal*$^{+}$ phenotype conferred by the set of 11 signal peptide fusions above was dependent on other Tat proteins. As expected, colony formation was abolished or severely retarded in a *tatA* strain but was unaffected in a *tatE* strain (data not shown). Interestingly, not all fusions were dependent on TatB for export. A *tatB malE* strain expressing ssHyaA-MBP or ssYcbK-MBP grew normally on minimal media with maltose even though growth was abolished in the *tatC* strain. These observations suggest

that interactions with TatB may not be required for the initiation of export by some *E. coli* Tat signal peptides.

The remaining 18 MBP fusions conferred growth in both wt and *tatC* cells. Since Tat translocation is completely inactivated in a *tatC* strain, the export of MBP into the periplasm of these cells must have occurred through the Sec translocon. Export through Sec does not rule out the ability to target the Tat translocon as well, but these two possibilities cannot be distinguished from the phenotype of MBP fusions alone. However, 14 of these 18 signal peptides satisfied the criteria for Tat export of GFP-SsrA. Since they allow translocation of the MBP reporter in a *tatC* strain, these 14 signal peptides also possess the ability to direct export via Sec, and therefore are assigned as Tat+Sec in Table 2.2. The ssFhuD-GFP-SsrA fusion gave 3-fold higher fluorescence in wt cells relative to the *tatC* strain but was lower compared to other fusions. Nonetheless, independent evidence has demonstrated that FhuD utilizes the Tat pathway (Ize *et al.* 2004) and therefore this signal peptide was also assigned as Tat+Sec.

Sec export was also evaluated quantitatively based on the alkaline phosphatase (PhoA) activity conferred by PhoA fusions. The two disulfide bonds that are essential for the folding of PhoA into its native conformation cannot form within the reducing cytoplasm of *E. coli* (Sone *et al.* 1997). Thus, PhoA is incompatible with Tat transport under normal cellular conditions. Unfolded PhoA is exported only via the Sec pathway and attains its active conformation within the periplasm (Figure 2.1). Consequently, whole cell PhoA activity represents a measure of the efficiency of Sec export. The activity values for the fusions were normalized relative to the PhoA activity obtained

when alkaline phosphatase was expressed with its own signal peptide (ssPhoA) from the same vector. Expression levels of each signal peptide:PhoA fusion were checked via Western blotting. The ssFdnG, ssFdoG, ssNapG, and ssNrfC fusions could not be detected but the other 25 proteins were expressed at comparable levels (data not shown). All of the signal peptides that failed to confer growth on maltose when fused to MBP and expressed in a *tatC* mutant strain also gave background activity when fused to PhoA, further highlighting their specific targeting to the Tat pathway.

Eight signal peptides that conferred growth on maltose in a *tatC* strain when fused to MBP gave only background activity when fused to PhoA, underscoring the differences in sensitivity between the two reporter proteins. The other eight signal peptides that already had been assigned as Tat+Sec, namely ssHybA, ssYnfF, ssHybO, ssAmiC, ssAmiA, ssYfhG, ssMdoD, and ssFhuD, gave PhoA activities >10% of those obtained when PhoA was exported with its own signal peptide. Notably, PhoA fusions to ssYfhG, ssMdoD, and ssFhuD gave 90%, 160% and 260% activity relative to the control, suggesting that these signal peptides are as efficient as ssPhoA, a *bona fide* Sec signal peptide.

The localization of fusions that display Tat-specific or Tat+Sec targeting was also examined by cell fractionation and Western blotting and representative results are shown in Figure 2.2. DsbA and GroEL were employed as periplasmic and cytoplasmic markers, respectively. For the TorA signal peptide, which our analysis showed to be Tat-specific, the PhoA fusion was localized exclusively in the spheroplast fraction whereas the MBP fusion was found mostly in the periplasm. In the *tatC* strain ssTorA-MBP accumulated

exclusively in the cytoplasm. In contrast, with ssMdoD, a signal peptide that exhibits secretion pathway promiscuity, both the PhoA and the MBP fusions were localized mainly in the periplasm and localization was unaffected in a *tatC* strain.

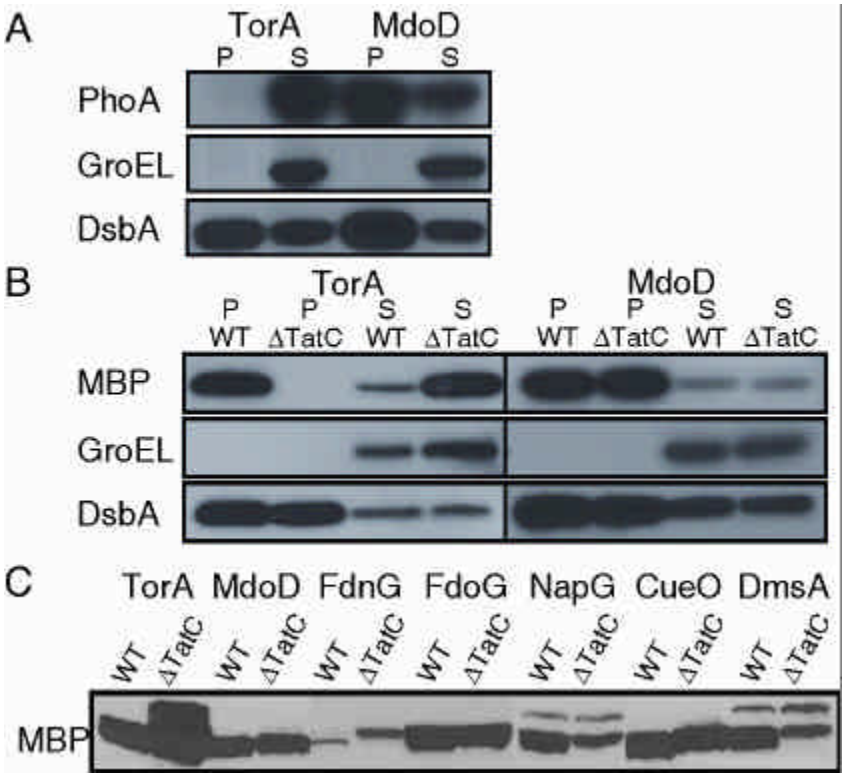

Figure 2.2. Subcellular localization of reporter protein fusions to ssTorA (Tat-specific) and ssMdoD (Tat+Sec) signal peptides. Western blots of periplasmic and spheroplast fractions from cells expressing signal peptide fusions to PhoA (A) and MBP (B). P: periplasmic fraction; and S: spheroplast fraction. (C) Western blots of whole cell lysates from cells expressing selected signal peptide fusions to MBP. PhoA fusions were expressed in *E. coli* strain DHB4 and MBP fusions were expressed in HS3018 and HS3018 *tatC*. GroEL and DsbA were used as fractionation markers by probing with anti-GroEL and anti-DsbA serums, respectively.

2.3.4 Localization of epitope-tagged fusions

To test whether the secretion pathway assignments using reporter fusions conform to those of the authentic proteins, we constructed an additional set of Tat proteins C-terminally-tagged with the FLAG epitope. Even though the FLAG fusions were

transcribed from the pBAD arabinose-inducible promoter on a low-copy plasmid, for several proteins we failed to detect expression by Western blotting in cells grown in minimal or rich media, at various growth temperatures, with or without oxygen and with different concentrations of inducer (Y. Kawarasaki and G. Georgiou, unpublished data). Further complications arose from the propensity of some of the fusions to accumulate in an aggregated state and from the saturation of Tat export, a common phenomenon when Tat proteins are expressed from multicopy plasmids (Chanal *et al.* 2003; DeLisa *et al.* 2003). For all of the FLAG-tagged proteins that could be successfully expressed, the results were consistent with the reporter fusion analysis (Figure 2.3). Expression and export of NrfC-FLAG could be detected only in anaerobically grown cells (Figure 2.3). In addition, while the ssYdhX-GFP fusion gave low fluorescence, the full-length protein was shown to be exported in a TatC-dependent fashion. Thus, YdhX must also be routed through the Tat pathway, bringing the total number of Tat-targeting signal peptides to 27.

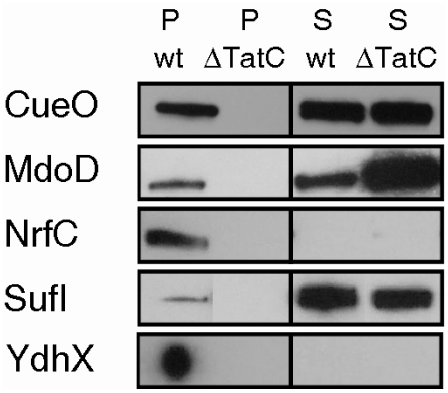

Figure 2.3. Subcellular localization of FLAG-tagged full-length proteins. Samples were normalized on the basis of OD_{600}, resolved on SDS-12% polyacrylamide gels, and probed with anti-FLAG serum. All fractionations were confirmed with fractionation markers by probing with anti-GroEL and anti-DsbA serums (data not shown).

2.3.5 Determinants of targeting specificity

As discussed above, 11 out of the 29 signal peptides tested in this study supported export only via the Tat pathway, eight signal peptides directed a small amount of Sec export (revealed by growth on maltose accompanied by low PhoA activity) and eight signal peptides appeared to be routed efficiently through both pathways. The eight signal peptides that displayed efficient Sec targeting had all the features of typical Tat signal peptides, i.e. a longer overall length, a twin-arginine consensus motif, and a less hydrophobic h-region. Earlier, von Heijne and coworkers (Cristobal *et al.* 1999) had proposed that a positive charge in the c-region of the signal peptide serves as a "Sec-avoidance motif" similar to that found in substrates of the plant chloroplast ΔpH-dependent/cpTat pathway (Bogsch *et al.* 1997). However, on a genomic scale, the data shown in Table 2.2 indicate that the overall charge of the c-region does not correlate well with avoidance of Sec export. As explained above, the reporter fusions we constructed also contained the first six to eight predicted aa of the respective mature proteins. When the total charge of the predicted signal peptide c-region (shown in bold in Table 2.2) together with the first few predicted aa of the mature protein (shown in italics in Table 2.2) was taken into consideration, the correlation between charge and Sec avoidance improved significantly (Table 2.2). In particular, all fusions with a =+2 charge in this region showed exclusive Tat export. Fusions with a +1 charge in this area gave inefficient export via Sec as indicated by a low PhoA activity. Only signal peptide fusions that have a net neutral or negative charge in the entire region displayed efficient export of PhoA via Sec.

To further examine the idea that the N-terminus of the mature protein might affect targeting specificity, we replaced the Asp residues at positions +2 and +4 (relative to the predicted signal peptidase cleavage site) in the ssMdoD signal peptide with Lys and Arg, imparting a net +2 charge. Additionally, we replaced the Asp residue at +2 with Lys and truncated the remaining portion of the mature protein to confer a net +1 charge. The ssMdoD signal peptide was selected because it does not encode a redox enzyme and there is no evidence of specific chaperones involved in its export. The resulting signal peptides, ssMdoD+2 and ssMdoD+1, were fused to PhoA and MBP, and the subcellular distribution of the respective proteins was determined by Western blotting. PhoA activities and growth on maltose were also determined and the results are summarized in Figure 2.4. As expected, an overall +2 charge completely abolished Sec targeting without affecting the efficiency of Tat translocation. A +1 charge allowed growth of the ssMdoD+1-MBP fusion on maltose in a *tatC* strain only after three days and gave very low PhoA activity. Consequently, these results support the notion that the charge of the first few aa of the mature protein together with the c-region of the signal peptide plays a role in avoiding routing to the Sec translocon.

Interestingly, even though ssMdoD can target the Sec translocon efficiently, the FLAG-tagged full-length MdoD protein was localized in the periplasm only in strains with a functional Tat apparatus. Furthermore, attempts to export MdoD with the Sec signal peptide ssPhoA were unsuccessful, providing more evidence that the mature protein also prevents mislocalization to the Sec pathway under physiological conditions (Bogsch *et al.* 1997; Stanley *et al.* 2000).

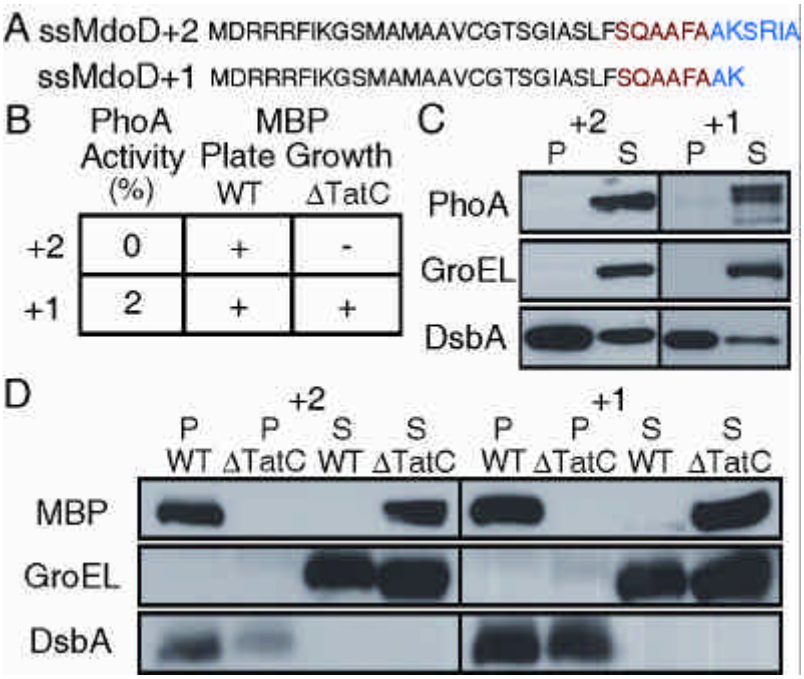

Figure 2.4. A positive charge at the N-terminus of the mature protein acts as a Sec avoidance signal. The charge of the first few amino acids of the mature MdoD was changed from -2 to +1 or +2 by site specific mutagenesis (ssMdoD+1 and ssMdoD+2). (A) Protein sequences of ssMdoD+1 and ssMdoD+2 up to the fusion site. (B) PhoA activity and growth on maltose minimal media. (C) Western blots of periplasmic *(P)* and spheroplast *(S)* fractions from cells expressing fusions to PhoA and (D) MBP.

2.4 Discussion

In this work we employed reporter and epitope-tagged fusions to investigate the export routing for 29 putative Tat signal peptides identified by bioinformatics means. We analyzed the localization of a total of 116 fusions and found that at least 27 can direct export via the Tat pathway. Thus, there are at least 27 Tat substrate proteins in *E. coli* in addition to "hitchhiker" substrates (DmsB, FdnH, HyaB, and HybC, ref. Berks *et al.* 2005) which are exported through Tat but do not contain a signal peptide. A Tat-targeting capability could not be confirmed nor ruled out for two signal peptides, ssYcdO and ssYaeI.

Fusions to the Tat-specific reporter GFP-SsrA, the Sec-only reporter PhoA, and MBP, which is competent for export via both Sec and Tat, were used in order to obtain quantitative information on the efficiency with which the various signal peptides can initiate export through the two translocons (Figure 2.1). It should be noted that no single reporter protein is sufficient for this analysis since there are three experimental outcomes (i.e. Tat-only, Sec-only, or Tat+Sec). For five signal peptides, Tat export of the respective FLAG-tagged full-length proteins was confirmed by Western blotting of periplasmic fractions from wt and *tatC* strains. However, most epitope-tagged full-length proteins showed poor expression under a variety of conditions or were prone to aggregation. Moreover, epitope-tagged full-length proteins do not provide information on the propensity of the signal peptide to target the Sec translocon since the mature polypeptides can exhibit folding features that are not compatible with Sec export.

As mentioned above, earlier studies had led to the experimental verification of 15 *E. coli* Tat signal peptides. The GFP-SsrA and MBP fusions correctly predicted the Tat export of 14 of these signal peptides. ssFhuD is the only known Tat signal peptide for which Tat export could not be confirmed from reporter fusion analysis and therefore the Tat portion of the assignment shown in Table 2.2 for that signal peptide had to be based on earlier experimental data (Ize *et al.* 2004). Our analysis provided definitive experimental evidence for 12 additional *E. coli* Tat substrates.

We addressed the targeting specificity of Tat signal peptides in a comprehensive manner. Approximately 60% of the signal peptides exhibited at least some level of promiscuous targeting to the Sec translocon, as manifest by the growth of MBP fusions on maltose in a *tatC* strain. Consistent with previous studies (Kebir and Kendall 2002; DeLisa *et al.* 2003), some of these signal peptides conferred a low level of PhoA activity. However, ssHybA, ssHybO, ssYnfF, ssAmiC, ssAmiA, ssYfhG, ssMdoD, and ssFhuD gave PhoA activities between 10-260% of the level obtained when PhoA is exported by its own signal peptide. Notably, ssFhuD, which gave the highest PhoA activity, has a less hydrophobic h-region typical of a Tat signal peptide, yet evidently is efficient in targeting the Sec translocon and activating SecA.

Earlier it had been suggested that signal peptides containing an h-region sufficiently hydrophobic for Sec-mediated export require additional signals for Sec avoidance (Cristobal *et al.* 1999; Kebir and Kendall 2002). Clearly, such signals are not found in many Tat signal peptides and therefore other mechanisms, including the

characteristics of the mature polypeptide and possibly interactions with specific chaperones, may be at play in order to prevent misrouting through Sec.

For 11 of the 27 signal peptides, we did not detect any Sec export of PhoA or MBP indicating that these signal peptides are Tat specific. While earlier data had suggested that the presence of a positive charge in the c-region serves as a Sec-avoidance signal (Bogsch *et al.* 1997; Cristobal *et al.* 1999) the results shown in Table 2.2 indicate that additional factors must be important for Tat specificity. For example, ssFdoG fusions, which have an uncharged c-region, were routed only via Tat whereas ssHybO and ssYnfF fusions with positively charged termini showed appreciable Sec activity. Our data suggest that the charge of the first few aa of the mature protein also plays a role in Sec avoidance (Table 2.2). Indeed, site-specific mutagenesis of the first six aa of the MdoD mature protein, resulting in a change in the charge from -2 to +1, severely reduced the Sec export of MBP and PhoA fusions while a further increase in the charge completely abolished it (Figure 2.4). Thus, alteration of the N-terminus of the mature protein alone was sufficient to completely switch the export pathway.

It has long been known that a positive charge in the first few aa of the mature protein severely retards or even abolishes Sec export, and that such a region serves as a stop transfer signal for Sec-dependent membrane proteins (Li *et al.* 1988; Yamane and Mizushima 1988; Andersson and von Heijne 1991; Geller *et al.* 1993; Kajava *et al.* 2000). The critical window has been proposed to comprise the first five aa (Li *et al.* 1988), and more recently, up to 16 aa (Kajava *et al.* 2000). Only 4% of 191 Sec-exported proteins contain a net positive charge in the first 16 aa of the mature protein (Kajava *et al.*

2000). In contrast, 37% (10/27) of N-terminal regions of proteins with a Tat signal peptide display a charge =+1 within either the five aa or the 16 aa windows. This observation raises the possibility that some proteins, particularly those that do not contain cofactors or form multimers, may have evolved to use the Tat pathway because they require positively charged N-termini for function and thus are not compatible with Sec export.

We note, however, that one signal peptide, ssYagT, has no charge in the critical region yet exhibits Tat specificity, indicating that apart from charge, an additional mechanism for specific Tat targeting must also exist. Signal peptide-binding chaperones (Oresnik *et al.* 2001; Jack *et al.* 2004) could potentially fulfill this role by sequestering the signal peptide from recognition by the Sec machinery.

Chapter 3

Folding Quality Control Mechanism in the Export of Proteins via the Bacterial Tat Pathway

3.1 Introduction

The bacterial Tat export machinery is structurally and mechanistically similar to the cpTat-driven protein import pathway found in plant thylakoids (Settles and Martienssen 1998; Dalbey and Robinson 1999; Keegstra and Cline 1999). In general, the Tat pathway is responsible for the export of polypeptides that are incompatible with the Sec machinery. As discussed in Chapter 1, these include proteins that require the incorporation of cofactors (e.g. glucose-fructose oxidoreductase, ref. Halbig *et al.* 1999), proteins that assemble different subunits in the cytoplasm (e.g. HybO, ref. Rodrigue *et al.* 1999) or proteins that cannot be exported by the Sec pathway in a functional form, for unknown reasons (e.g. GFP, ref. Thomas, J. D. *et al.* 2001). Processes such as cofactor incorporation or the assembly of protein subunits hinge on the formation of a substantial degree of secondary or tertiary structure in the polypeptide. *In vitro* experiments in the plant thylakoid system have shown that: (i) bovine pancreatic trypsin inhibitor (BPTI) fused to a cpTat-dependent pathway precursor is imported into thylakoids, even when BPTI was permanently folded using irreversible disulfide cross-linkers which prevented any unfolding during the translocation process (Clark and Theg 1997); and (ii) dihydrofolate reductase (DHFR) bound to methotrexate, a ligand that greatly stabilizes

the folded state of the protein, is specifically imported into the thylakoid lumen via the cpTat-dependent pathway (Hynds *et al.* 1998). However, BPTI and DHFR are significantly smaller (56 aa and 187 aa, respectively) than typical cpTat-dependent polypeptide substrates.

In vivo the export of cofactor-containing proteins by the bacterial Tat pathway is abolished when the incorporation of the cofactor in the cytoplasm is blocked (Santini *et al.* 1998; Jack *et al.* 2004). The attachment of cofactors is a complex process that involves interactions with other cellular components including specific chaperones. Whether the protein assumes a fully native conformation upon cofactor attachment but before export, and if so, whether partial unfolding occurs during translocation were far from clear at the time this study was undertaken. Moreover, the folding state of Tat-competent proteins that do not contain cofactors had not been established. An equally intriguing question was whether the Tat pathway is able to accept both folded and unfolded polypeptides as substrates, or if only proteins that have reached a substantially native state in the cytoplasm are competent for translocation. Earlier *in vitro* data regarding the import of DHFR by the cpTat system provide some support for the former model (Hynds *et al.* 1998). The alternative model, whereby the Tat pathway can export only native-like proteins, implies the existence of a folding quality-control mechanism inherent to this pathway. Because the Tat translocon is necessary and sufficient for protein export, the ability to assess the folding competence of exported proteins would then have to reside with the membrane TatABCE components.

In this chapter, the relationship between protein folding and export competence was examined by analyzing the subcellular localization of proteins with structural disulfide bonds in strains that either allow or disfavor oxidative protein folding in the cytoplasm. Tat-dependent accumulation of PhoA, as well as other multidisulfide proteins including single-chain and heterodimeric F_{AB} antibody fragments, into the periplasm was observed only in strains that enable the formation of disulfide bonds in the cytoplasm. Thus only proteins with substantially native-like conformation can be exported by the Tat translocon *in vivo*, demonstrating the presence of a folding quality-control mechanism that is intrinsic to the export process and is unrelated to cofactor association.

The work in this chapter was carried out in collaboration with Dr. Matthew DeLisa (DeLisa *et al.* 2003). Dr. DeLisa constructed the F_{AB} expression plasmids described later, grew cells and fractionated samples, and performed the flow cytometric analysis and Western blotting.

3.2 Materials and Methods

3.2.1 Bacterial strains and plasmids

The bacterial strains and plasmids used in this study are described in Table 3.1. Strains DHBA and DRA were obtained by P1 transduction of the *dsbA*::*kan1* allele from JCB571 into *E. coli* strains DHB4 and DR473, respectively. Strain DRB was obtained by P1 transduction of *tatB*::*kan* allele from MCMTA (MC4100 *tatB*::*kan*) into *E. coli* strain DR473. Strains DRC and FA113 were obtained by P1 transduction of the *tatC*::*spec* allele from BUDDY (MC4100 *tatC*::*spec*) into *E. coli* strains DR473 and FA113, respectively. *E. coli* strain XL1-Blue was used for cloning and plasmid propagation.

Table 3.1. Bacterial strains and plasmids used in this study.

E. coli strain or plasmid	Relevant phenotype or features	Source
DHB4	MC1000 *phoR* Δ(*phoA*) *Pvu*II Δ(*malF*)3 F'[*lacI*^{q}*ZYA pro*]	Laboratory stock
DR473	DHB4 Δ*trxB gor552* Tn*10*Tet *ahpC** Tn*10*Cm (*araC* P_{ara}-*trxB*)	J. Beckwith
FA113	DHB4 *trxB gor552* Tn*10*tet^r *ahpC**	J. Beckwith
DHA, DRA	DHB4 *dsbA*::*kan*, DR473 *dsbA*::*kan*	This work
DRB	DR473 *tatB*::*kan*	This work
DRC, FUDDY	DR473 *tatC*::*spec*, FA113 *tatC*::*spec*	This work
pTrc99A	*trc* promoter, ColE1 *ori*, amp^r	Amersham-Pharmacia
pAID135	Δ(2-22)PhoA	Derman *et al.* 1993
pTorA-PhoA	ssTorA-Δ(1-22)PhoA cloned in pTrc99A	This work
pKK-PhoA	as pTorA-PhoA with R11K;R12K mutation in ssTorA	This work
pFdnG-PhoA	ssFdnG-Δ(1-22)-PhoA in pTrc99A	This work
pTrc99-F_{AB}	V_H-C_{H1} and V_L-C_L in pTrc99a	Levy *et al.* 2001
pssTorA- F_{AB}	ssTorA-V_H-C_{H1} and V_L-C_L dicistronic operon in pTrc99a	This work
pKK- F_{AB}	as above with ssTorA(R11K;R12K)	This work
pBAD-ΔssdsbC	signal sequenceless dsbC in pBAD33	Levy *et al.* 2001

For monitoring the export of PhoA, cells grown in LB media overnight were subcultured into M9 salts supplemented with 0.2% glucose, 1 µg/ml thiamine, 1 mM $MgSO_4$, and 50 µg/ml 18 amino acids (excluding methionine and cysteine) at a 100-fold dilution, and then incubated at 30°C. For the expression of scFv and F_{AB} antibody fragments, cells were subcultured from overnight cultures into fresh LB medium (5% v/v) and then incubated at 30°C. Protein synthesis was induced by adding isopropyl β-D-thiogalactoside to a final concentration of 0.1 mM when the cells reached an OD_{600} of ~0.5. Where appropriate, the co-expression of DsbC was induced using 0.2% arabinose. Antibiotic selection was maintained for all markers on plasmids at the following concentrations: ampicillin, 100 µg/ml; chloramphenicol, 25 µg/ml; kanamycin, 50 µg/ml; and spectinomycin, 100 µg/ml.

3.2.2 Enzyme activity assays

Cells expressing PhoA were harvested three hours after induction, treated with 100 mM iodoacetamide, pelleted by centrifugation and fractionated by the cold osmotic shock procedure described in Chapter 2. Soluble protein was quantified by the Bio-Rad protein assay, using bovine serum albumin (BSA) as standard. PhoA activity (Chapter 2) and β-galactosidase activity assays (Derman *et al.* 1993) were performed as described previously. Only data from fractionation experiments in which =95% of the β-galactosidase activity was in the cytoplasmic fraction are reported. The trypsin resistance of PhoA was assessed as described previously (Sone *et al.* 1997) with the iodoacetamide step included.

ELISA assays were performed according to Levy *et al*. (2001). Western immunoblotting was performed as described in Chapter 2. The following primary antibodies were used: rabbit anti-PhoA (Rockland) diluted 1:5,000, rabbit anti-mouse IgG (specific for $(F_{AB})_2$ light chains, Pierce) diluted 1:5,000, monoclonal rabbit anti-DsbC (gift from John Joly, Genentech) diluted 1:10,000 and monoclonal rabbit anti-GroEL (Sigma) diluted 1:10,000. The secondary antibody was 1:10,000 diluted goat anti-rabbit-HRP.

3.3 Results

3.3.1 Alkaline phosphatase can be exported by Tat only in strains with an oxidizing cytoplasm

In bacteria, the thioredoxin and glutaredoxin pathways maintain the cytoplasm in a highly reducing state which strongly disfavors the oxidation of protein thiols (reviewed by Ritz and Beckwith (2001)). For this reason, proteins requiring disulfide bonds have to be exported into a more oxidizing environment. PhoA is normally exported via a Sec-specific signal peptide into the periplasmic space where it is rapidly oxidized by DsbA to form its two disulfide bonds that are critical for the stability and catalytic activity of the protein (Sone *et al.* 1997). In the previous chapter as well as in earlier studies by other groups, it has been demonstrated that fusions of PhoA to Tat-specific signal peptides are not exported via the Tat pathway (Berks 1996; Kebir and Kendall 2002; Stanley *et al.* 2002). The inability of PhoA fusions to be exported by the Tat pathway might be due to the fact that such proteins are normally unfolded in the cytoplasm, which is maintained in a strongly reducing state that precludes the formation of disulfide bonds. To investigate the above hypothesis and to examine the relationship between folding and Tat export competence, mutant strains of *E. coli* were employed that allow the formation of disulfide bonds in the cytoplasm or, conversely, disable protein oxidation in the periplasm.

Bessette *et al.* (1999) had demonstrated that disulfide bonds can form readily when PhoA is expressed without its signal peptide in the cytoplasm of oxidizing mutant

strains such as *E. coli* DR473 (*trxB gor ahpC**). For convenience, this strain background is designated C:ox/P:ox because both the cytoplasm and the periplasm are oxidizing (Figure 3.1). In a *dsbA* derivative of DR473 (DRA), protein oxidation in the periplasm is impaired (Bardwell *et al.* 1991) and therefore such a strain is designated C:ox/P:red. In DRA cells, the formation of disulfide bonds in PhoA can only occur within the cytoplasm.

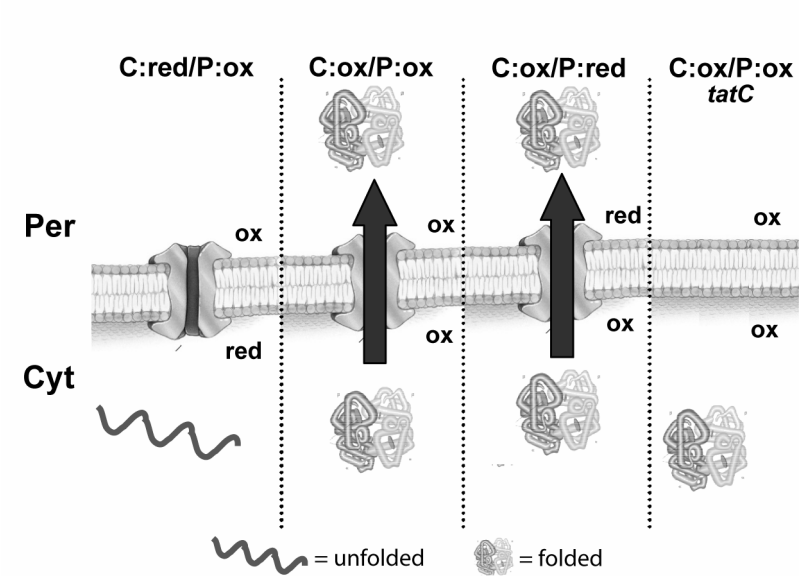

Figure 3.1. PhoA folding and export in different strain backgrounds. Ox and red refer to oxidizing and reducing redox potentials in the specified subcellular compartment. In C:ox (DR473) cells, PhoA is able to fold in the cytoplasm and thus can serve as a substrate for the Tat pathway. The localization of PhoA is not impaired in the C:ox/P:red strain DRA (DR473 *dsbA*). Deletion of *tatC* (or *tatB*) in the C:ox/P:ox strain blocks PhoA export.

A set of seven signal peptides (Table 3.2) that were shown to direct export via the Tat pathway in the previous chapter were fused to PhoA and expressed in DHB4, DR473, DHB4 *dsbA:kan* (DHA), DRA and finally in DR473 *tatB::kan* (DRB) and in DR473 *tatC::spec* (DRC). In the latter two strain backgrounds, the mutational

inactivation of *tatB* or *tatC* impairs export through the Tat pathway. Cells were grown in minimal media and the subcellular distribution of PhoA was determined. Following cell harvesting, cell samples were treated with 100 mM iodoacetamide to prevent the formation of disulfide bonds during fractionation by osmotic shock and the PhoA enzymatic activities in the cytoplasmic and periplasmic fractions were then measured. The degree of leakage of cytoplasmic components during fractionation was <5% as determined by the subcellular distribution of β–galactosidase activity and of GroEL protein (detected by Western blotting).

Table 3.2. Sequences of the Tat signal peptides used in this study

Signal Peptide	Sequence
FdnG	MDVSRRQFFKICAGGMAGTTVAALGFAPKQALA
FdoG	MQVSRRQFFKICAGGMAGTTAAALGFAPSVALA
HyaA	MNNEETFYQAMRRQGVTRRSFLKYCSLAATSLGLGAGMAPKIAWA
TorA	MNNNDLFQASRRRFLAQLGGLTVAGMLGPSLLTPRRATAAQA
DmsA	MKTKIPDAVLAAEVSRRGLVKTTAIGGLAMASSALTLPFSRIAHA
SufI	MSLSRRQFIQASGIALCAGAVPLKASA
CueO	MQRRDFLKYSVALGVASALPLWSRAVFA

The fusion proteins were classified into two classes on the basis of whether the formation of active PhoA was strictly dependent on an oxidizing cytoplasm (Table 3.3). The first class of Tat signal peptides (class I; ssFdnG, ssFdoG, ssHyaA, ssTorA) was found to export PhoA to the periplasm in a C:ox-dependent and Tat-dependent manner. In other words, accumulation of PhoA activity in the periplasm only occurred in strains having both an oxidizing cytoplasm and an intact Tat apparatus. Specifically, when class I signal peptides were expressed in the parental strain DHB4, which has a reducing cytoplasm (i.e. it is C:red/P:ox) only background PhoA activity was observed.

However, in a C:ox strain, a significant fraction of the PhoA activity (25-50% of total) was found in the periplasm. Importantly, the periplasmic PhoA activity was not dependent on DsbA and was thus virtually the same in C:ox/P:ox cells relative to C:ox/P:red cells (i.e., in DRA). This finding indicates that the oxidation of PhoA occurs only in the cytoplasm prior to export. The accumulation of active protein in the periplasm was abolished when the signature RR motif in the signal peptide was changed to a KK sequence (data not shown), a substitution known to completely block export (Cristobal *et al.* 1999). Likewise, inactivation of either *tatB* or *tatC* in a C:ox/P:ox strain resulted in near complete loss of PhoA enzymatic activity in the periplasm while the cytoplasmic activity remained virtually unchanged. For one of the four class I Tat-signal peptides (ssHyaA), export was only partially blocked in a *tatB* strain, as expected based on maltose binding protein fusion results discussed in the previous chapter.

Table 3.3. Periplasmic alkaline phosphatase activity in cells expressing Tat signal peptide-PhoA fusions*

Signal Peptide	DHA C:red/P:red	DHB4 C:red/P:ox	DR473 C:ox/P:ox	DRA C:ox/P:red	DRB (DR473 *tatB::kan*) C:ox/P:ox	DRC (DR473 *tatC::spec*) C:ox/P:ox
Δ2-20[a]	43	56 (52)	69 (7)	57 (8)	56 (8)	47 (9)
FdnG	32	45 (48)	420 (42)	371 (38)	30 (5)	6 (2)
FdoG	55	62 (54)	446 (53)	385 (50)	17 (2)	11 (2)
HyaA	17	28 (46)	187 (61)	185 (59)	65 (23)	20 (4)
TorA	42	51 (51)	437 (26)	403 (28)	38 (3)	17 (1)
DmsA	35	123 (>90)	280 (23)	175 (14)	112 (12)	74 (9)
SufI	75	323 (>90)	459 (30)	381 (19)	263 (15)	308 (18)
CueO	25	102 (>90)	153 (18)	135 (12)	185 (19)	80 (9)

All signal sequences were fused in frame to *E. coli* PhoA(Δ1-22) and the fusions were inserted into vector pTrc99A (see Experimental Procedures).
*Cells were treated with 100 mM iodoacetamide immediately after harvesting to prevent formation of disulfide bonds during subsequent sampling processing. PhoA activity was calculated as the amount of *p*-nitrophenyl phosphate hydrolyzed (in μmol) per min at 25°C and pH 8.0. Reported values for PhoA activity are the average of 3 separate measurements from 2 independent experiments (n=6). Standard error is less than 10% for all reported data. Values in parenthesis indicate the percentage of the total enzymatic activity in the periplasmic fraction.
[a]PhoA with a Δ2-20 in the native signal peptide

Figure 3.2A-C shows the subcellular distribution of one class I signal peptide fusion, ssFdnG-PhoA, in the different strain backgrounds. Inactivation of *tatC* in the C:ox strain background resulted in a lower level of cytoplasmic fusion protein. Similarly, a smaller amount of ssFdnG-PhoA fusion was observed in the cytoplasm of C:red cells. A higher molecular weight species corresponding to the ssFdnG-PhoA precursor could be resolved in 4-20% acrylamide gradient gels (Figure 3.2D). This higher molecular weight species could be processed to a band with an electrophoretic

mobility identical to the mature protein by incubation with trypsin, due to the presence of trypsin sensitive sites in the signal peptide (Kebir and Kendall 2002). The mature protein was completely resistant to trypsin under conditions that result in the disappearance of full length OmpA (Figure 3.2D). Further, no loss of PhoA enzymatic activity was detected in these experiments consistent with the observation of Sone *et al*. (Sone *et al.* 1997) that only the native, fully oxidized form of PhoA is trypsin resistant.

Three out of the seven Tat signal peptides (class II; ssDmsA, ssSufI, and ssCueO) exhibited high PhoA activity in the wt strain DHB4 (C:red/P:ox). Previous analysis (Chapter 2) as well as further study on the PhoA fusions to these signal peptides indicated that they can engage both the Tat and the Sec pathways. Specifically, the ability to engage the Sec pathway was evident by: (i) significant reduction of periplasmic PhoA activity in *dsbA* mutants relative to *dsbA*+ cells (compare DHB4 (C:red/P:ox) to DHA (C:red/P:red), and also DR473 (C:ox/P:ox) to DRA (C:ox/P:red)). For all three class II signal peptides, the level of periplasmic PhoA activity in the *dsbA* mutant strains was approximately 60% of what was obtained in the isogenic *dsbA*+ *E. coli*. (ii) The appearance of significant periplasmic PhoA activity in *tatB* or *tatC* mutants. (iii) Loss of PhoA activity in a *secA* conditional mutant (strain MM52 *secA51*(temperature-sensitive), C:red/P:ox) following upshift to the non-permissive temperature of 42^{o}C relative to wt cells also shifted to 42^{o}C (data not shown). The ability of class II signal peptides to also engage the Tat pathway is revealed by the presence of high PhoA activity in the periplasmic fraction of DRA (C:ox/P:red) cells, in addition to evidence provided in Chapter 2.

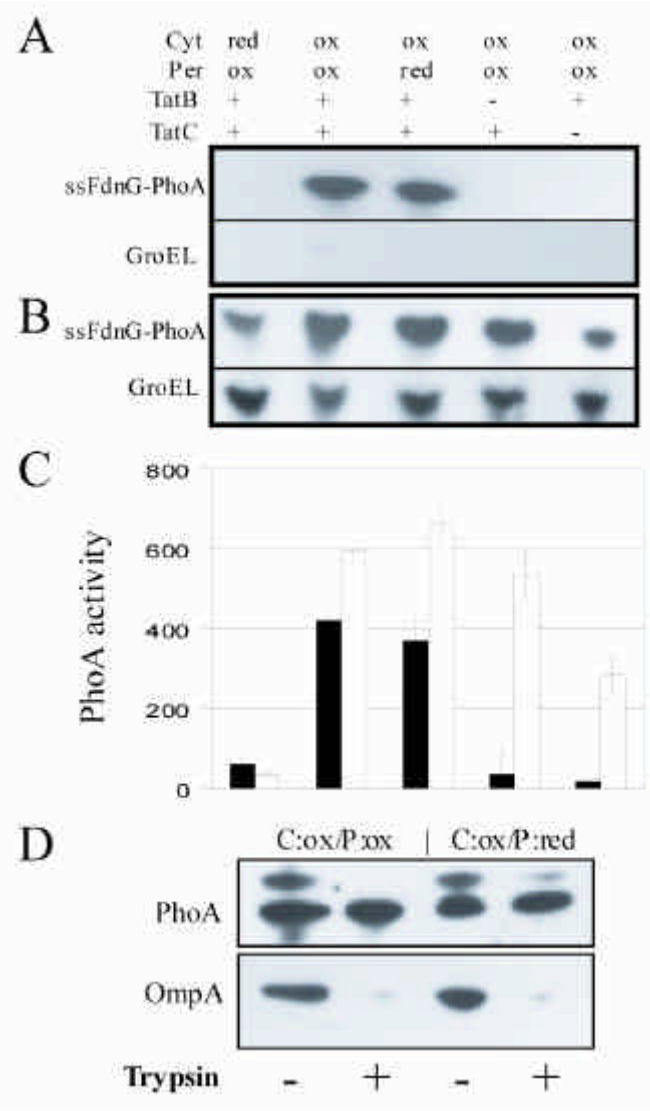

Figure 3.2. Subcellular localization of ssFdnG-PhoA. Immunoblotting of (A) periplasmic and (B) cytoplasmic fractions from cells expressing ssFdnG-PhoA. Samples were normalized based on the amount of total cell protein and resolved on 12% SDS-PAGE gels. GroEL was used as a fractionation marker by probing with anti-GroEL serum. (C) PhoA activity for the same periplasmic (solid) and cytoplasmic (empty) fractions as in A and B. (D) Trypsin sensitivity analysis of periplasmic fractions collected from C:ox/P:ox and C:ox/P:red cells. Samples were separated on 4-20% SDS-PAGE gels and probed with anti-AP and anti-OmpA serum.

3.3.2 Tat export of antibody fragments

In addition to PhoA, class I signal peptides could mediate the export of other proteins with intra- or intermolecular disulfide bonds from the cytoplasm of C:ox strains. Single chain antibodies (scFv) contain two intramolecular disulfide bonds, one in the variable heavy (V_H) and one in the variable light (V_L) chain. The 26.10 anti-digoxin scFv (Huston *et al.* 1988) was fused to the class I Tat signal peptide ssTorA and the accumulation of antigen binding protein in the periplasmic space of C:ox/P:ox cells was monitored by Enzyme-Linked ImmunoSorbent Assay (ELISA, Figure 3.3). Nearly 45% of the digoxin binding activity was localized in the periplasmic space. The accumulation of active anti-digoxin scFv antibody in the periplasm was not affected in DRA (C:ox/P:red) but was reduced to background levels in a *tatC* mutant.

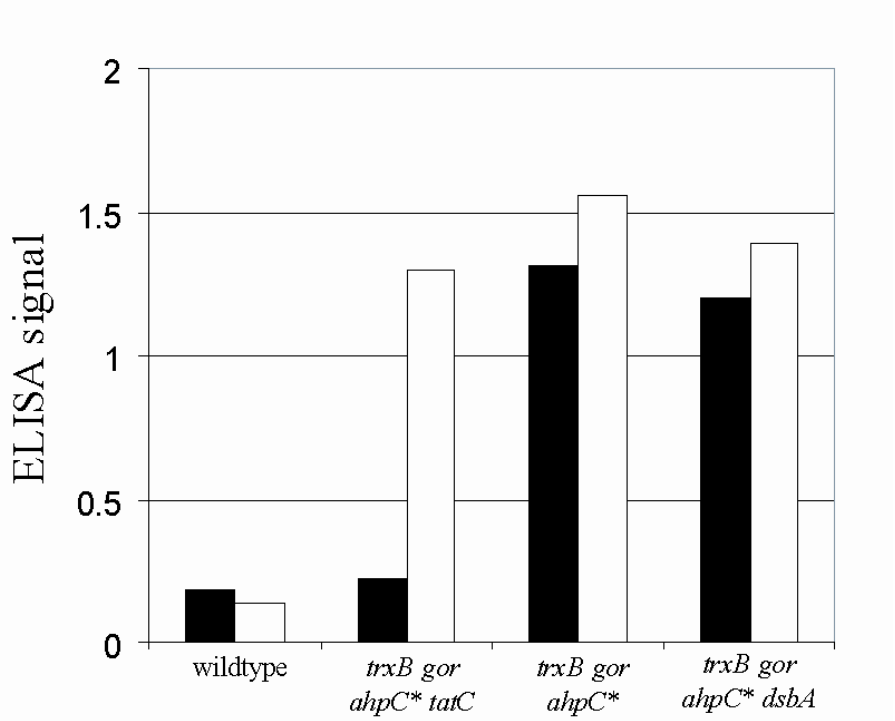

Figure 3.3. Tat export of an scFv antibody. Detection of scFv by ELISA in the periplasmic and cytoplasmic fractions. Periplasmic (solid) and cytoplasmic (empty) samples were serially diluted and started from the same amount of total protein. Data reported is from a 4-fold dilution and is the average of two independent experiments.

F_{AB} antibodies are heterodimeric proteins in which the heavy chain (V_H-C_{H1}) and light chain (V_L-C_L) are linked by an intermolecular disulfide bond. In addition, F_{AB} proteins contain four more intrachain disulfides, two within each of the heavy and light chains. A dicistronic operon consisting of a gene encoding an ssTorA- V_H-C_{H1} fusion followed by the light chain (V_L-C_L) was constructed (Figure 3.4A; ref. Levy *et al.* 2001). In this construct the light chain does not contain a signal peptide and therefore cannot be exported from the cytoplasm by itself. Induction of protein synthesis in DRA (i.e. C:ox/P:red) was found to result in the accumulation of a small but significant amount of light chain polypeptide in the osmotic shock fraction. Western blot analysis

using an anti-mouse IgG $(F_{AB})_2$ antibody specific for the F_{AB} light chain indicated that the intensity of the V_L-C_L band in the osmotic shock fraction was approximately 15-20% of that in the cytoplasm (data not shown). Neither light chain nor F_{AB} protein could be detected in the osmotic shock fraction of a *tatC* mutant strain or from DHA (C:red/P:ox) cells.

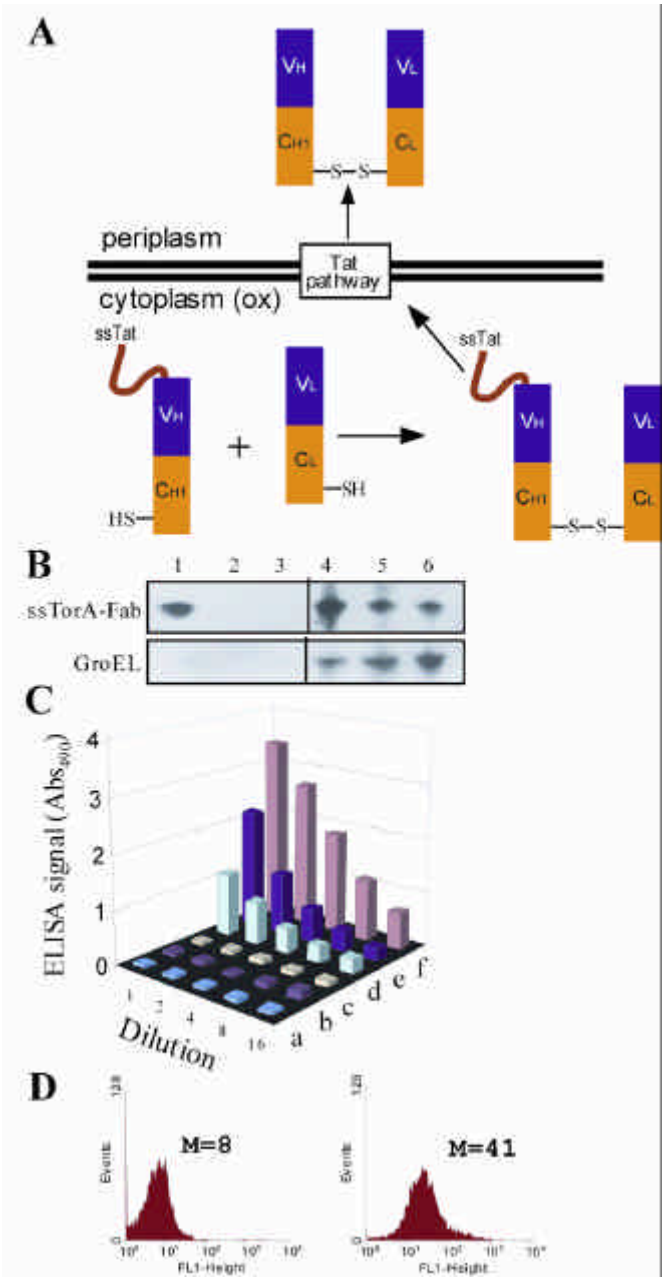

Figure 3.4. Tat export of an anti-digoxin antibody fragment (F_{AB}). (A) Tat export of F_{AB} antibodies. (B) Western blotting of periplasmic (lanes 1-3) and cytoplasmic (lanes 4-6) fractions collected from cells co-expressing ssTorA-F_{AB} fusion and ΔssDsbC. Strains used were: (*1*,*4*) C:ox/P:ox; (*2*,*5*) C:ox/P:ox/*tatC* and (*3*,*6*) C:red/P:ox. All lanes were loaded with the same amount of total protein and anti-mouse IgG $(F_{AB})_2$ antibody was used to detect the F_{AB} light chain. GroEL was used as a fractionation marker for the spheroplast fractions. (C) ELISA of periplasmic (*a*,*c*,*e*) and cytoplasmic (*b*,*d*,*f*) fractions collected from (*a*,*b*) C:red/P:ox; (*c*,*d*) C:ox/P:ox/*tatC*; and (*e*,*f*) C:ox/P:ox cells. (D) Flow cytometric analysis of C:red/P:ox cells (left) and C:ox/P:red cells (right) expressing ssTorA:F_{AB} and ΔssDsbC and labeled with FITC-digoxin.

If pre-association of the heavy and light chains is required for export via the Tat pathway, then it follows that conditions that enhance the yield of correctly folded F_{AB} may increase the export efficiency. The yield of the anti-digoxin F_{AB} expressed in the cytoplasm of a *trxB gor ahpC** mutant (the C:ox strain FA113 was used for these experiments to enable co-expression of DsbC from the *ara* promoter) is increased markedly upon co-expression of a signal-sequenceless version of the periplasmic disulfide isomerase DsbC (ΔssDsbC; ref. Levy *et al.* 2001). Consistent with this hypothesis, in cells co-expressing ΔssDsbC the intensity of the light chain band in the osmotic shock fraction was 70% of that in the cytoplasm (compared to 15-20% without ΔssDsbC, as noted above). A similar partitioning of the F_{AB} digoxin binding activity was detected by ELISA (Figure 3.4C). In the FA113 *tatC::spec* strain, no F_{AB} could be detected in the periplasmic fraction and a significant reduction in the amount of protein remaining in the cytoplasm was also observed. It has been observed that depletion of the *tat* genes results in inactivation or degradation of Tat substrate proteins in the cytoplasm (Angelini *et al.* 2001). Similarly, no light chain protein or F_{AB} binding activity could be detected in *E. coli* DHB4 (Figure 3.4B and C) or when the RR dipeptide in ssTorA was substituted by KK (data not shown).

Finally, an experiment was devised to test whether a functional F_{AB} protein capable of binding to the hapten digoxin is formed *in vivo.* Incubation of *E. coli* in a hypertonic solution (5x PBS) increases the outer membrane permeability such that fluorescent ligands can equilibrate within the periplasm while periplasmic proteins are unable to leak out of the cell (Chen *et al.* 2001). Cells binding the fluorescent ligand can

thus be detected by flow cytometry. C:ox/:P:ox cells expressing ssTorA-V_H-C_{H1} and V_L-C_L and incubated with a digoxin-fluorescein complex exhibited a marked increase in cell fluorescence relative to control cells (C:red/P:ox, Figure 3.4D), indicating that the F_{AB} protein is indeed functional within the periplasm. The detection of digoxin-binding activity in intact cells *in situ* further rules out the possibility that formation of disulfide bonds in the F_{AB} protein may have occurred during cell fractionation.

3.4 Discussion

The hallmark of the Tat pathway that sets it apart form all other modes of protein translocation across lipid bilayer membranes is the ability to export polypeptides that have already assumed a degree of stable secondary or perhaps even tertiary structure. However, the relationship between the folding state of a protein and Tat export competence was poorly understood prior to this study. In fact, export of folded proteins had only been demonstrated experimentally *in vitro* for the plant equivalent of the Tat pathway, the cpTat-dependent import system of thylakoids. Clark and Theg (1997) showed that prefolded and chemically crosslinked BPTI could be imported into thylakoids when linked to the 17-kDa subunit of the oxygen-evolving complex of photosystem II (prOE17). However, BPTI is a very small protein (56 residues, 6.5 kDa) and therefore not typical of native Tat and cpTat substrates. Furthermore, the crosslinked BPTI could also be translocated across the chloroplast envelope membrane, a process that normally transports polypeptides that are largely unfolded (Schnell and Blobel 1993). Hynds *et al*. (1998) found that DHFR, bound to methotrexate, could be transported in a cpTat-dependent manner. The incorporation of amino acid analogues, which destabilize DHFR and render it protease-sensitive, did not inhibit transport. Such "malfolded" DHFR may: (i) be able to fold into a non-native compact conformation that is accepted by the cpTat system, (ii) remain unfolded and able to engage the thylakoid Sec pathway for import, in analogy with the Tat class II signal peptides described in the present study or (iii) be transported in an unfolded state by the Tat pathway. While Hynds *et al*. (1998)

favored the third scenario the other two interpretations could not be ruled out based on their findings.

The ability to genetically promote oxidative protein folding selectively in the cytoplasm or in the periplasmic space enabled us to examine the requirements for export competence of proteins that: (i) do not contain cofactors and (ii) exhibit well characterized folding kinetics. Comparison of PhoA accumulation in the periplasmic fraction of wt *E. coli*, mutant strains that enable the formation of disulfide bonds in the cytoplasm and their *dsbA* derivatives (Table 3.2) revealed that for all seven signal peptides, oxidative folding in the cytoplasm is a prerequisite for Tat export. Sone *et al.* demonstrated that partially oxidized PhoA lacking the Cys-168-Cys-178 disulfide is proteolytically unstable and is subject to rapid degradation by DegP *in vivo* (Sone *et al.* 1997). Thus, the stable periplasmic accumulation of the PhoA fusions in our studies (in a *degP*+ strain background) as well as the trypsin stability of the PhoA domain of ssFdnG-PhoA in the osmotic shock fraction (Figure 3.2) indicate that the exported form of the PhoA fusion must be fully oxidized. Based on these considerations, it can be concluded that the Tat pore is capable of translocating PhoA that is fully oxidized and contains the native disulfide bonds. At present it is not possible to ascertain whether the protein is exported as a dimer or whether the association of the oxidized, folded monomers occurs following translocation into the periplasmic space. However, because *in vitro* dimerization represents a slow step in the folding of PhoA (Walker and Gilbert 1994) the latter mechanism appears more plausible.

For 4/7 signal peptides (class I: ssFdnG, ssFdoG, ssHyaA and ssTorA) export occurs exclusively via the Tat pathway as evidenced by: (i) the fact that PhoA enzymatic activity in the osmotic shock fraction is unaffected by the presence or absence of DsbA; and (ii) export was completely abolished by the substitution of the invariant RR in the signal peptide by KK, in *tatC* mutants and, with the exception of ssHyaA, also in *tatB* mutants. The role of TatB in the export of hydrogenases has been the subject of some controversy (Chanal *et al.* 1998; Sargent *et al.* 1999). However, TatB is dispensable in the export of the non-physiological substrate colicin V (Ize *et al.* 2002).

In cells expressing PhoA fusions to class I signal peptides approximately 26% (for ssTorA) to almost about 60% (for ssHyaA) of the total enzymatic activity in cell lysates was found in the periplasmic fraction. The observed efficiency of PhoA export is typical of native Tat substrates and of Tat fusions to heterologous proteins (Angelini *et al.* 2001; Thomas *et al.* 2001; Yahr and Wickner 2001; DeLisa *et al.* 2002; Stanley *et al.* 2002). If indeed the Tat translocon exports preferentially oxidized, monomeric PhoA, as discussed above, then the enzymatically active protein remaining in the cytoplasm could represent dimerized protein that is incompatible with export. In addition, various amounts of inactive, export incompetent PhoA accumulated in the cytoplasm. The inactive cytoplasmic PhoA in C:ox cells was trypsin sensitive and its abundance was elevated at higher growth temperatures and upon expression of the fusions from high copy number plasmids, and was also dependent on the signal peptide used (data not shown). In contrast, there was no evidence of inactive PhoA fusions in the periplasmic fraction for any of the constructs tested. In some cases (e.g. for ssFdnG-PhoA in Figure 3.2D) a

portion of the PhoA fusion protein in the periplasm migrated as a higher molecular weight band having the expected electrophoretic mobility of the precursor. This higher molecular weight species was processed by trypsin to the mature band, presumably due to the presence of sensitive sites within Tat signal peptides (Kebir and Kendall 2002).

Class II Tat signal peptides (ssDmsA, ssSufI, and ssCueO) afforded some PhoA export from cells with a reducing cytoplasm. This PhoA activity was not abolished in C:red *tatB* or *tatC* mutants indicating that export of the protein to the periplasm did not involve the Tat machinery. When class II signal peptide fusions were expressed in C:ox/P:ox cells, the amount of active PhoA in the periplasm increased by between 42% (for ssSufI-PhoA) and >100% (for ssDmsA-PhoA). However, inactivation of *dsbA* in the C:ox strain DR473 resulted in a significant reduction in periplasmic PhoA. The two possible explanations for these results are either that translocation of misfolded or unfolded PhoA occurs via the Tat apparatus or that class II signal peptides can engage both the Sec and Tat translocons. The finding that the export of PhoA activity in C:red cells is eliminated in a *secA51*(temperature sensitive) mutant strain grown at the non-permissive temperature (*E. coli* MM52 at 42^{o}C) supports the latter hypothesis. The ability of class II signal peptides to engage both the Sec and the Tat export apparatus is in agreement with studies by Sanders *et al.* (2001) who found that the export of heterologous cytochrome c fused to the HyaA signal peptide could proceed via the Sec pathway in the absence of the heme cofactor, whereas when the heme was enzymatically ligated in the cytoplasm, the Tat pathway became the predominant export route.

Several earlier efforts to export PhoA via the Tat pathway had been unsuccessful (Sambasivarao *et al.* 1990; Berks 1996; Stanley *et al.* 2002). In some cases PhoA activity could be detected in the periplasmic space but export was independent of the Tat pathway (e.g. Stanley *et al.* 2002). The analysis presented here explains these previous observations: while the export of PhoA via the Tat pathway is dependent on oxidative folding in the cytoplasm, certain signal peptides are able to engage both the Sec and Tat pathways. The export flux among these two pathways presumably depends on the folding kinetics of the polypeptide, with folded molecules exported through the Tat system and proteins that have not yet reached a critical state in folding able to access the Sec pathway.

In addition to PhoA, fusions to other multidisulfide proteins, such as scFv and F_{AB} antibody fragments, become competent for Tat export only when expressed in strains where the cytoplasm promotes oxidative protein folding. The accumulation of F_{AB} antibody in the periplasm could only be observed in C:ox strains and was dependent on an intact Tat apparatus and on the presence of a functional signal peptide on the heavy chain. The simplest explanation for these findings is that the F_{AB} first assembles in the cytoplasm where the two chains become linked by an intermolecular disulfide followed by Tat export of the folded protein. Increasing the folding efficiency of F_{AB} in the cytoplasm through the co-expression of ΔssDsbC greatly enhanced the export efficiency of F_{AB}. The export of fully-folded F_{AB} molecules into the periplasm even though only one of the two chains contains a signal peptide is representative of the “hitchhiker” mode

of export whereby a signal-sequenceless polypeptide is exported via its association with a second polypeptide that can engage the export apparatus (Rodrigue *et al.* 1999).

The results outlined above provide conclusive evidence that: (i) in the absence of folding in the cytoplasm there is no protein export via the Tat pathway; and (ii) folded proteins of at least ~43 kDa (the size of monomeric PhoA), as well as proteins that must assemble subunits in the cytoplasm, are exported in a folded conformation via the bacterial Tat pathway *in vivo*. The inability to export PhoA, scFv and F_{AB} proteins in the absence of disulfide bonds indicates the existence of a quality control mechanism that is an integral part of Tat translocation machinery. Since only the Tat proteins are required for translocation, the folding quality control mechanism must be inherent to the TatABCE proteins in the membrane. For example, a cytoplasmic domain of a Tat component may be able to function as a chaperone by binding to exposed hydrophobic regions in unfolded protein intermediates. Alternatively, interactions with aberrantly folded proteins could inhibit either the assembly of the Tat translocation pore or the gating of the pore once formed. The isolation of Tat mutations that enable the export of unfolded PhoA fused to class I signal peptides in C:red strains could help distinguish between these two mechanisms.

Finally, it should be noted that the results of the present study have significant ramifications for biotechnology. First, this work demonstrates that proteins with structural disulfide bonds can be efficiently exported by the Tat pathway. Many industrially important exported proteins including enzymes contain multiple disulfide bonds that are required for proper folding. Such proteins may be exported via the Tat

pathway in C:ox strains. The advantage of utilizing the Tat pathway is that it may alleviate problems with cell toxicity often encountered when proteins exported by the Sec pathway become 'stuck' in the translocon. Second, the folding quality control feature of the Tat pathway may be exploited in combinatorial library screening experiments to "weed out" mutant polypeptides that cannot fold properly.

Chapter 4

Engineering of a Tat-Dependent Phage Display System for Use in Evolving Streptavidin-Binding GFP Variants

4.1 Introduction

The screening of combinatorial protein libraries is critically dependent on technologies for the display of proteins on the surface of viruses or cells. As discussed in Chapter 1, protein translocation on a biological surface in most instances requires export across a lipid bilayer membrane (the only exception is display on viruses that assemble within the cytoplasm, such as lambda or T7). So far, display technologies have relied only on the Sec apparatus for protein export. However, biological limitations associated with Sec export have precluded the display – and thus the engineering – of many proteins. Since protein export via the Tat pathway is mechanistically completely distinct from the Sec pathway, adaptation of the Tat pathway for use in combinatorial library screening can have a significant impact on protein engineering.

The display of proteins on filamentous phage is widely used for the isolation of ligand-binding proteins, such as antibodies, from very large libraries (Rader and Barbas 1997; Hoogenboom *et al.* 1998). Filamentous phage have an outer shell, known as the protein coat, comprising a major coat protein (p8) and four minor coat proteins (p3, p6, p7 and p9). This coat is assembled in the periplasm of *E. coli*, and each of the five coat proteins must reach the inner membrane for the assembly process (Ohkawa and Webster

1981; Endemann and Model 1995). Minor coat protein p3 is the only coat protein known to require an export pathway (Sec) to arrive at its membrane location (Peters *et al.* 1994); the other coat proteins are smaller in size and are likely spontaneously inserted into the membrane (Kuhn 1995; Gao *et al.* 1999; Hufton *et al.* 1999). Typically, p3 is used for the display of protein libraries. When other coat proteins are utilized, Sec-dependent signal peptides are usually employed to ensure that the larger fusions can still be integrated into the membrane for assembly (Kuhn 1995; Gao *et al.* 1999). To date, there has not been any evidence to indicate that p3 can be exported via the Tat pathway (Paschke and Hohne 2005; Droge *et al.* 2006) or that the number of phage particles produced would be unaffected if export were to take place via the Tat pathway. Two critical events must occur for Tat-compatibility: p3 must be able to fold in the cytoplasm in order to be accepted by the Tat translocon, and the Tat machinery must be able to correctly target p3 to the inner membrane for phage assembly to occur (Figure 4.1 steps A and C).

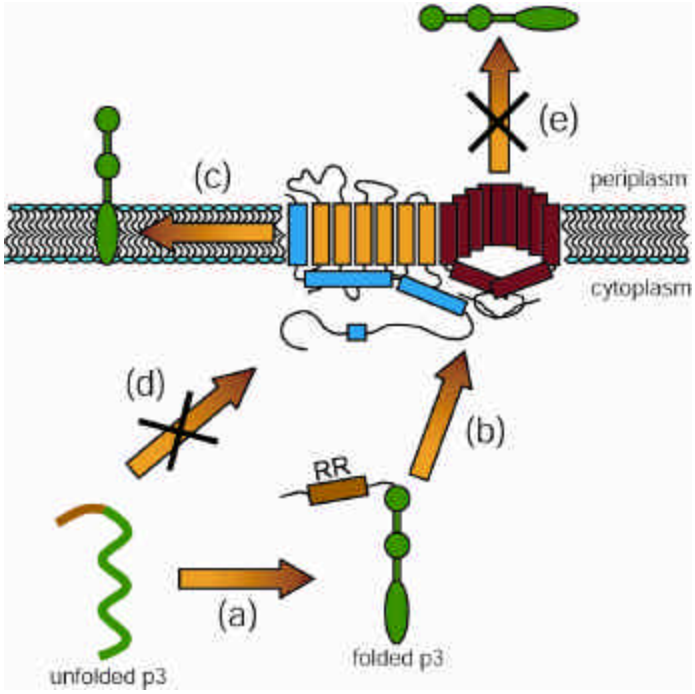

Figure 4.1. Requirements for Tat-compatibility of p3. Unfolded p3 must be able to fold in the cytoplasm (a). The Tat signal peptide indicated by the RR directs folded p3 to the Tat translocon (b) which must then properly insert p3 into the inner membrane (c). Any p3 that is not properly folded cannot be exported via Tat (d) and phage particles cannot be assembled if the Tat translocon releases p3 to the periplasm (e).

An intriguing potential target of a Tat-based display system is GFP, which as discussed in Chapters 1 and 2 is incompatible with Sec export. Therefore it cannot be displayed as a Sec-exported fusion to p3 because the resulting protein is localized in the periplasm in non-fluorescent form. In contrast, fluorescent GFP is obtained in the periplasm when first folded in the cytoplasm and then exported via the Tat pathway; it was successfully used as a Tat reporter protein in the studies described in Chapter 2. GFP consists of a β-barrel structure made up of 11 beta sheets, with both unstructured and alpha-helical loop regions at each end of the barrel connecting the beta strands. The barrel forms a protective shell around the fluorophore, a p-

hydroxybenzylideneimidazolidone moiety. In the past, there have been attempts to evolve or design the loops on either end of its barrel structure to bind various ligands (Abedi *et al.* 1998; Baird *et al.* 1999; Zeytun *et al.* 2003). The goal of such efforts is to engineer a single fluorescent, ligand-binding molecule for diagnostic and research applications. Such GFP variants would be extremely useful in a variety of assays. Currently, such assays rely on a ligand-binding protein or peptide and a separate detection molecule. The detection molecule must either be conjugated or be non-covalently bound to the ligand-binding protein. Ligand-binding GFP variants would eliminate the need for such two-molecule systems, and could be used to provide real-time information on the presence of a ligand (Zeytun *et al.* 2003). Furthermore, the fluorescence theoretically would be directly correlated to ligand concentration (Zeytun *et al.* 2003).

In this study, it was found that either p3 is not compatible with the Tat export process, or p3 cannot reach the inner membrane-bound state necessary for incorporation into the phage virion when fused to a Tat signal peptide. To circumvent this issue, we made use of two proteins, Jun and Fos, with domains that bind tightly to form noncovalent, stable, coiled coil interactions known as leucine zippers (Kouzarides and Ziff 1988; Ransone *et al.* 1989). In the Tat-based filamentous phage system we developed, p3 was still routed via the Sec pathway, but fused to the leucine zipper domain of the protein Fos. The protein of interest was fused to the leucine zipper domain of the protein Jun and exported via the Tat pathway. Thus the phage virion was assembled with p3 fused to one half of the leucine zipper pair, which formed a complex

with the other half of the leucine zipper pair that was fused to the protein of interest. This approach was tested with GFP as a model protein of interest. Furthermore, evidence is provided that GFP can be evolved to bind ligands via this system.

4.2 Materials and Methods

4.2.1 Strains, plasmids, and growth conditions

Strains and plasmids used in this study are described in Table 4.1. Unless otherwise noted, cells were grown at 37°C on LB media with 50 μg/ml chloramphenicol, 30 μg/ml kanamycin, or 100 μg/ml ampicillin as appropriate.

Table 4.1. Strains and plasmids used in this study

E. coli strain or plasmid	Relevant phenotype or features	Source
DHB4	MC1000 *phoR* Δ(*phoA*) *Pvu*II Δ(*malF*)3 F'[*lacI*^{q}*ZYA pro*]	Laboratory stock
DR473	DHB4 Δ*trxB gor552* Tn*10*Tet *ahpC** Tn*10*Cm (*araC* P_{ara}-*trxB*)	J. Beckwith
TG-1	*supE thi-1* ? (*lac-proAB*) ? (*mcrB-hsdSM*)*5* [F´ *traD36 proAB lacI*q*Z? M15*]	J. Beckwith
pHEN2	*Lac* promoter, ColE1 and M13 *ori*, amp^r	Amersham-Pharmacia
pBAD18	*araBAD* promoter, pBR322 and M13 *ori*, amp^r	(Guzman *et al.* 1995)
pBAD33	*araBAD* promoter, pACYC184 and M13 *ori*, cm^r	(Guzman *et al.* 1995)
pssTorA-scFv-p3	ssTorA-anti-digoxin 26.10 scFv cloned in pHEN2	This work
pssPelB-scFv-p3	ssPelB-anti-digoxin 26.10 scFv cloned in pHEN2	M. Fitza
pssTorA-scFv	ssTorA-anti-digoxin 26.10 scFv-stop codon cloned in pHEN2	This work
pssPelB-scFv	ssPelB-anti-digoxin 26.10 scFv-stop codon cloned in pHEN2	(Levy *et al.* 2001)
pssTorA-p3	ssTorA cloned in pHEN2	This work
pTJG-PFP	Dicistronic: ssTorA-Jun-GFP and ssPelB-Fos-p3 in pBAD18	This work
pTJG-n-mer	ssTorA-Jun-GFP with an n-mer inserted at aa157 of GFP	This work
pPFP18	ssPelB-Fos-p3 in pBAD18	This work

4.2.2 Biochemical assays

Western blotting was performed as described in Chapter 2. The following primary antibodies were used: monoclonal mouse anti-GFP (Molecular Probes) and mouse-anti gene III protein (Exalpha Biologicals).

ELISA assays were performed according to Levy *et al*. (2001). To coat the plates, either digoxin-BSA (2 μg/ml) or streptavidin (100 μg/ml) was used. The primary antibody mouse anti-His-HRP was used for protein ELISAs at a dilution of 1:30000. The primary antibody mouse anti-M13(p8)-HRP (Amersham Pharmacia) was used at a dilution of 1:5000 for phage ELISAs. Chemiluminescent detection ELISA assays were performed the same way except that 100 μl of Pico West chemiluminescent substrate (Pierce) was added to each well in place of the o-phenylenediamine solution and no sulfuric acid was necessary to stop the reaction. Finally, for chemiluminescent detection ELISAs, detection was carried out using a luminescence plate reader and readings are reported as relative light units (RLUs).

For flow cytometric analysis, *E. coli* TG-1 containing plasmids encoding GFP fusions were grown overnight in LB media as described above and 250 μl of each overnight culture were used to inoculate 5 ml of fresh media. After 1 h shaking at 25°C, gene expression was induced with arabinose to a final concentration of 0.2%, the cells were incubated for an additional 4 hours at 25°C and 1 ml samples were harvested by centrifugation, diluted in PBS and analyzed with a Becton-Dickinson FACSort flow cytometer.

4.2.3 Phage production, titering and assays

Helper phage was produced by infecting 5 ml of 2x yeast tryptone media (2xYT, 16 g tryptone, 10 g yeast extract, and 5 g NaCl per liter) containing *E. coli* TG-1 at exponential phase with M13K07 helper phage (Pharmacia) in the ratio of 20 phage particles per bacterial cell. The infected culture was grown for 2 hours with shaking at

37°C and then used to inoculate 500 ml of 2xYT media. Following infection, the cells were grown for 1 hour with shaking at 37°C before adding 30 μg/ml kanamycin and continuing growth for an additional 16 hours. The bacteria were centrifuged at 10,800 g for 15 minutes at 4°C and the supernatant (containing the phage) was collected. One fifth volume of 20% polyethylene glycol (PEG)/2.5 M NaCl was added to the supernatant and left overnight with mixing at 4°C. The supernatant mixture was then spun at 10,800 g for 15 minutes at 4°C. The pellet was resuspended in 2 ml PBS and filter sterilized through a 0.45 micron filter. The stock was titered (see below) and diluted to $5x10^{12}$ phage forming units per ml in 30% glycerol before storing at -80°C.

Phage displaying the p3 fusions was produced as follows: An overnight culture of *E. coli* TG-1 with the appropriate phagemid construct was used to inoculate 50 ml of 2xYT media containing 0.2% glucose. This culture was grown with shaking at 37°C to an OD_{600} of approximately 0.5. Helper phage was then added in the ratio of 20 phage particles per bacterial cell, and the culture was incubated without shaking at 37°C for 30 minutes. The infected cells were then harvested by centrifugation and resuspended in 30 ml of 2xYT media containing 100 μg/ml ampicillin and 0.2% arabinose. The culture was grown with shaking at 25°C for two hours before 30 μg/ml kanamycin was added, and was then left to continue growing with shaking at 25°C overnight. Subsequently, the supernatant containing the produced phage particles was obtained by centrifuging the cell suspension at 10,800 g for 45 minutes.

PEG precipitation of the phage was carried out as follows: To the supernatant collected from the phage production protocol, 1/5 volume PEG/NaCl was added and

mixed overnight at 4°C. The mixture was spun at 10,800 g for 30 minutes at 4°C. The pellet was resuspended in 1.6 ml water and 0.4 ml PEG/NaCl and kept for 2-4 hours at 4°C with mixing. The suspension was then centrifuged at 10,800 g for 10 minutes at 4°C. The supernatant was aspirated off and the pellet resuspended in 1 ml of 1xPBS. The suspension was spun at full speed in a microcentrifuge at 4°C for 10 minutes to remove bacterial cell debris, and the supernatant was transferred to a fresh tube. Phage was stored at 4°C for 2 weeks or less.

To titer the produced phage, 1 µl of phage was diluted into 1 ml of PBS and 1 µl of that was then used to infect 1 ml of *E. coli* TG-1 at an OD_{600} of approximately 0.5. The infected cells were then placed in a 37°C water bath without shaking for 30 minutes. Dilutions were plated on TYE (15 g bacto-agar, 10 g tryptone, 5 g yeast extract, and 8 g NaCl per liter) plates containing 1% glucose and grown overnight at 37°C.

Fluorescent measurements of phage were carried out in 96-well plates using a Tecan Safire fluorescent plate reader, with wells containing only 1x PBS used as a negative control.

4.3 Results

4.3.1 Phage protein p3 is incompatible with Tat export

To investigate the potential of using the Tat pathway to export p3, the TorA signal peptide, which specifically targets proteins only to the Tat pathway (Chapter 2), was fused to the model protein scFv 26.10, an antibody fragment against digoxin, and to p3 (Figure 4.2A). The fusion was expressed from the pHEN2 phagemid that is widely used for phage display (Griffiths *et al.* 1994). A 6x histidine tag, recognizable by specific, commercially available antibodies, was cloned between the scFv sequence and gene III. Thus, any scFv-6xHis-p3 proteins exported via the Tat pathway and incorporated into assembled phage can be readily detected by antibodies that recognize the peptide tag which is not present in the wild type p3. As a control, a similar construct was made using the Sec-dependent PelB signal peptide in place of ssTorA.

Phage protein p3 contains multiple cysteines, which are oxidized to form four disulfide bonds in the native structure. As shown in Chapter 3, proteins with disulfide bonds have to be expressed in cells with an oxidized cytoplasm in order for folding to occur, a requirement for export via the Tat pathway. Therefore, the ssTorA-scFv-p3 fusion protein was expressed in the *trxB gor* mutant strain of *E. coli* that maintains an oxidizing cytoplasm (C:ox, see also Chapter 3). Helper phage was added to begin the phage assembly process. The produced phage was harvested, purified via PEG-precipitation, and titered to determine the phage concentration. An ELISA was then carried out to compare the display of the scFv antibody fragment on phage produced in *E. coli* expressing p3 fusion proteins exported either via the Tat or Sec pathways (Figure

4.2B). The Tat-exported p3 fusion resulted in a higher yield of phage, possibly because of higher expression. However, it gave an ELISA signal that was three and five fold lower compared to the Sec-based phage in the wt and C:ox strain, even though a much higher concentration of the former phage was used in the ELISA assay.

We tested the possibility that the scFv was incompatible with Tat export, and therefore inhibited the export of the ssTorA-scFv-p3 fusion. To test this hypothesis, a stop codon was introduced between the scFv and the p3 gene. Cell fractionation following by detection of the scFv in periplasmic and cytoplasmic fractions by ELISA was carried out in either wt cells or C:ox cells. An ssPelB-scFv fusion was used as a positive control (Figure 4.2C). The Tat-exported scFv gave a signal comparable to the Sec-exported protein in the C:ox strain, and no signal in the wt strain. This result is also consistent with nearly identical experiments performed on ssTorA-scFv and ssTorA-F_{AB} fusions in Chapter 3 (Figure 3.3). Therefore, the scFv is compatible with export via Tat in a C:ox strain.

Since the Tat pathway could export the scFv when expressed alone in a C:ox strain, the next step was to examine the compatibility of p3 with the Tat pathway. To this end, ssTorA was fused to p3 in phagemid pHEN2 without the scFv. Membrane fractionations were performed on both wt and C:ox strains expressing the construct, and each fraction was run on a polyacrylamide gel followed by Western blotting (data not shown). No phage protein could be detected in the periplasmic or membrane fractions of either strain. This indicates that the phage produced in the earlier experiments consisted entirely of the p3 produced from the helper phage DNA, rather than the ssTorA-scFv-p3

fusion from the phagemid. The absence of p3 from both the periplasmic space and the inner membrane implies that either p3 does not properly fold even in the C:ox strain, or it is incompatible with Tat export for some other, unknown reason.

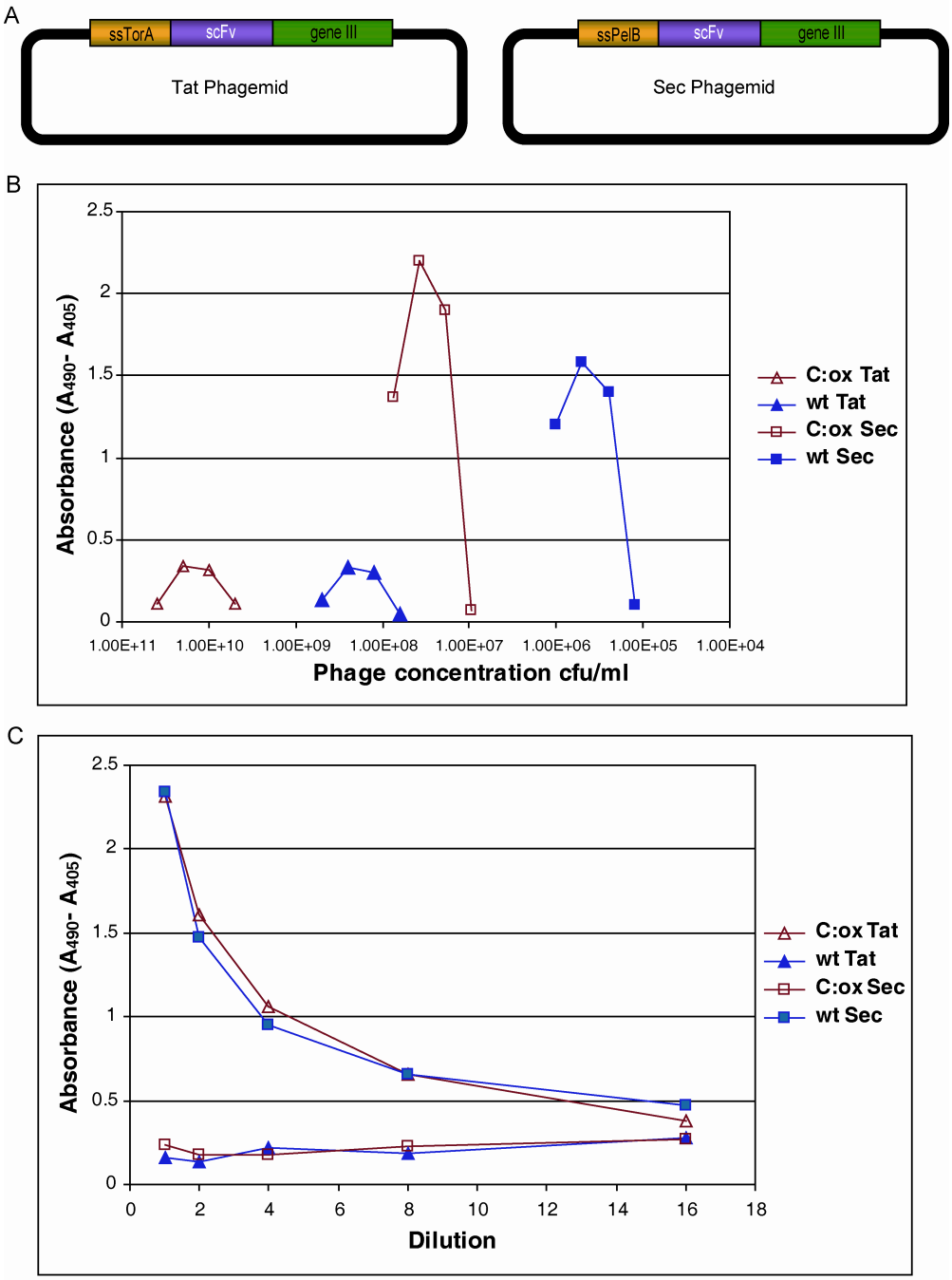

Figure 4.2. Tat versus Sec export of an anti-digoxin scFv-p3 fusion and of an scFv alone. (A) The phagemid constructs used for the experiment in *B*. (B) ELISA on phage particles expressing either ssPelB-scFv-p3 (Sec) or ssTorA-scFv-p3 (Tat). (C) ELISA on protein obtained via osmotic shock from the periplasmic fraction of cells expressing either ssPelB-scFv (Sec) or ssTorA-scFv (Tat). For the ELISAs, fusions were expressed in DHB4 (wt) and DR473 (C:ox) strains of *E. coli*. The ELISA plates were coated in 2 µg/ml digoxin-BSA. For the protein ELISA, cells were normalized to the same OD_{600} prior to fractionation, and equal volumes of the periplasmic fraction were loaded in each top lane.

4.3.2 Engineering of a Tat-dependent phage display platform

An alternative method was devised that bypasses the need to export p3 via the Tat pathway. This approach capitalized on the use of two proteins with domains that bind tightly to each other to form a noncovalent, stable complex. Jun and Fos, two proteins with leucine zipper domains that interact in this fashion (Kouzarides and Ziff 1988; Ransone *et al.* 1989), were chosen to fulfill this function. The sequence encoding the leucine zipper domain of Fos was inserted between the Sec signal sequence and gene III in a phagemid vector. The sequence encoding the leucine zipper domain of Jun was placed between the Tat signal sequence and the gene encoding the protein of interest in the same phagemid vector (Figure 4.3A). Once both fusions reach the periplasm, the leucine zipper domains are able to interact. As a result, the protein of interest becomes associated with p3 and is displayed on the phage particle without the need for p3 to be exported via the Tat pathway (Figure 4.3B). Importantly, the constructs were designed to contain flanking cysteines around the Jun and Fos leucine zipper domains such that disulfide bonds form upon interaction in the periplasm, preventing disassociation. To test this system, GFP was used as the protein of interest and phage was produced in both TG-1 and an isogenic *tatC* mutant strain. The phage was harvested, concentrated and purified via PEG precipitation, and titered to determine the phage concentration. Equal concentrations of phage were loaded in triplicate in 96 well plates and the fluorescence was measured (Figure 4.3C). The results clearly indicate that there is indeed Tat-dependent display of GFP. Subsequent to this portion of the study, Paschke and Hohne demonstrated the success of a very similar system (Paschke and Hohne 2005).

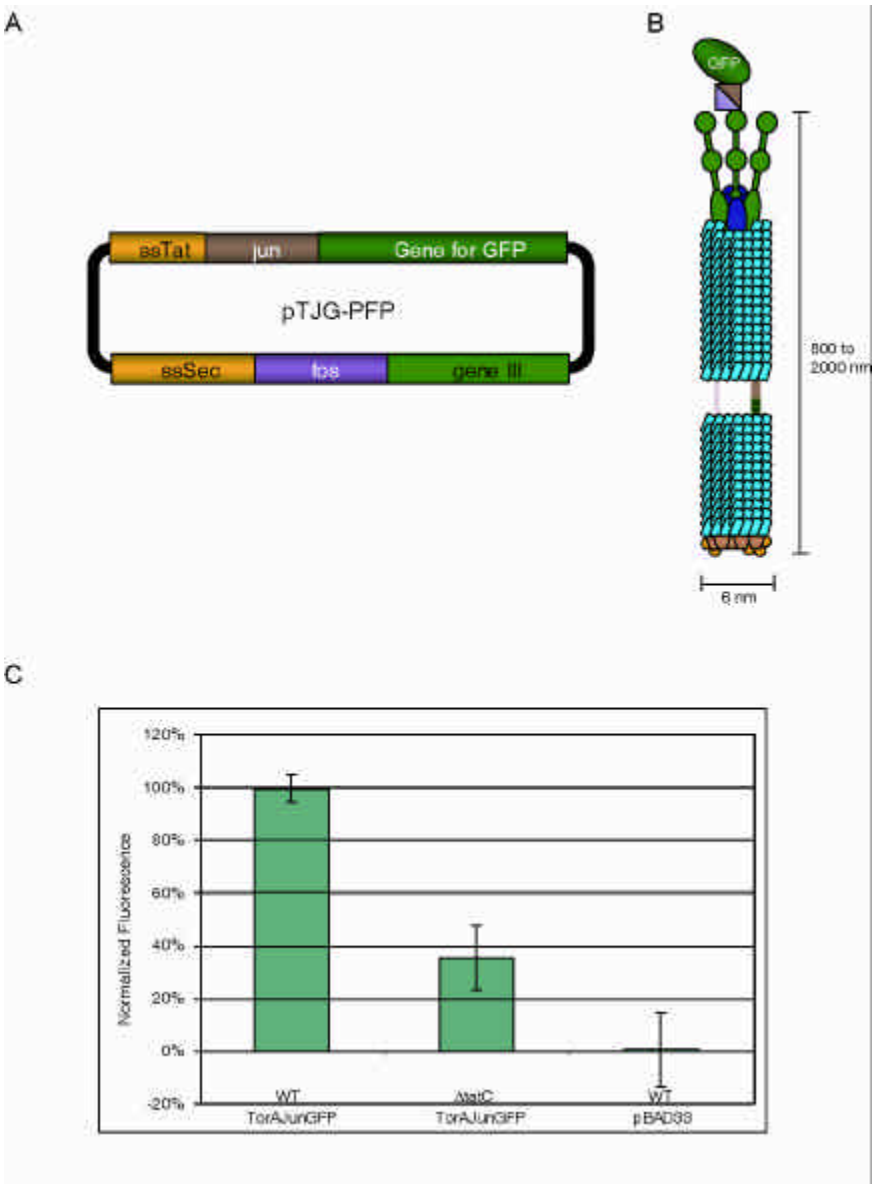

Figure 4.3. A Tat-dependent phage display system. (A) Cartoon of dicistronic phagemid system, with gene for GFP inserted as a model gene of interest. (B) Phage particle produced by cells containing the phagemid shown in A. (C) Fluorescence measurements, normalized to the highest reading on the plate after background was subtracted. WT refers to TG-1 and ΔtatC to TG-1 *tatC*.

4.3.3 Toward the development of a ligand-binding GFP variant

A useful application of the Tat-dependent display platform is the engineering of a ligand-binding variant of GFP. Previous attempts to do so using phage display have been unsuccessful, largely because GFP is incompatible with Sec export (Feilmeier *et al.* 2000; Zeytun *et al.* 2003). Studies have been carried out to determine the most permissive sites in the loop regions for insertion of polypeptides (Abedi *et al.* 1998; Baird *et al.* 1999; Topell *et al.* 1999; Cetinkaya *et al.* 2006). Insertion of peptides within permissive sites does not affect the fluorescence or the stability of the protein whereas insertion into non-permissive sites abrogates fluorescence or reduces the thermodynamic stability of the protein. The degree to which a particular site within GFP tolerates the insertion of peptide sequences is typically measured by the intensity of fluorescence compared to the wt GFP protein. However, the studies cited above report differing optimal insertion sites for polypeptides of assorted lengths, and varying degrees of loss of fluorescence intensity even for insertions into the same loop of GFP.

The first step toward a ligand-binding GFP was to identify a suitable permissive loop and also the ideal length of polypeptide sequence that can be inserted at that position. To accomplish this, 6-mer, 8-mer, and 16-mer peptides that bind to the protein streptavidin were inserted in between amino acids 157 and 158 of GFP, which had been reported as a tolerant site by multiple studies (Figure 4.4; refs. Abedi *et al.* 1998; Topell *et al.* 1999). The Tat-specific TorA signal peptide was used to direct the various GFP constructs to the Tat pathway. The plasmid was constructed such that the leucine zipper domain of Jun was also fused to GFP. Two of the three resulting GFP variants were

fluorescent as measured by flow cytometry when expressed in TG-1, with 85% and 95% of the fluorescence of wt GFP for the 6-mer and 8-mer insertions, respectively (Figure 4.5A). Insertion of the 16-mer peptide into GFP severely reduced its fluorescence and thus the cells were only slightly more fluorescent than the negative control, exhibiting 2% of the wt GFP fluorescence. Importantly, fluorescence decreased significantly for all the GFP mutants in cells grown at temperatures higher than 25°C (data not shown). To ensure that expression levels are unaffected by the peptide insertions, the amount of GFP protein was detected by Western blot in cell samples harvested at the same time as those for the FACS analysis (Figure 4.5B). Unexpectedly, two bands were observed in the lysates of the cells expressing wt GFP or GFP variants with peptide insertions. The lower band exhibits the expected electrophoretic mobility for GFP. The data in Fig 4C reveal that the wt GFP and the 6-mer and 8-mer variants were expressed at similar levels whereas the GFP containing the 16-mer peptide insertion is present at a much lower level. Thus, the lower fluorescence of the GFP with the 16-mer peptide insertion is due to its lower expression level.

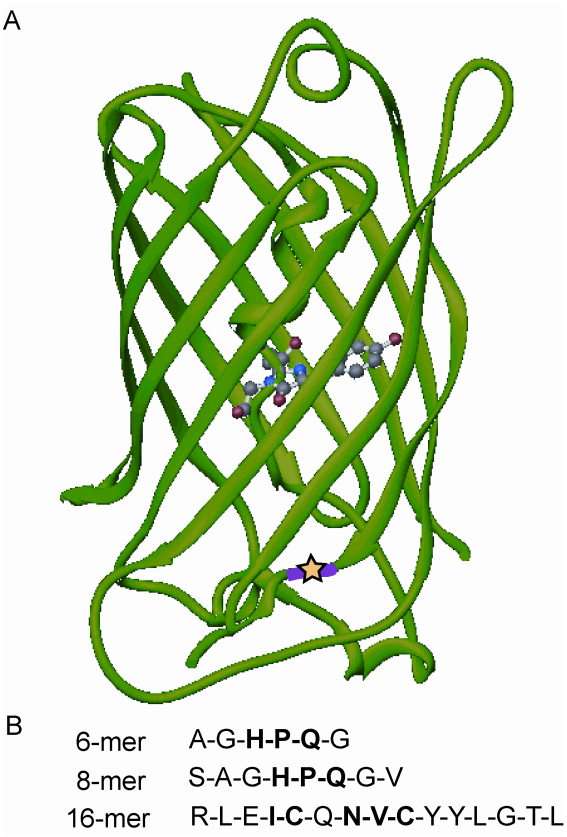

Figure 4.4. Peptide insertions in GFP loop at aa157. (A) Structure of GFP, with aa 157 and 158 marked with a star. (B) Amino acid sequences of the peptides inserted between GFP aa 157 and 158, with streptavidin-binding residues indicated in bold.

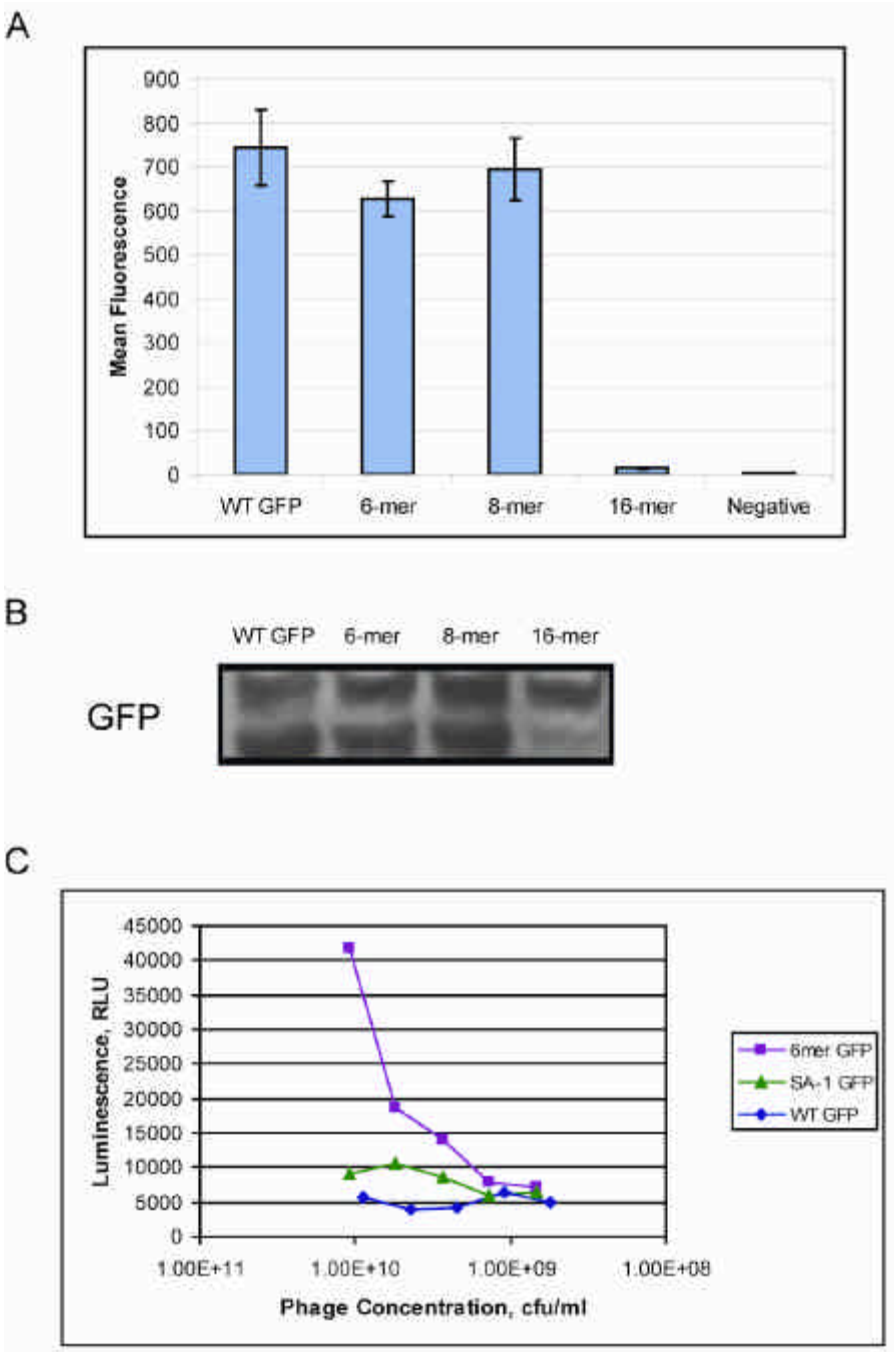

Figure 4.5. Characterization of GFP variants containing 6-mer, 8-mer, and 16-mer insertions at aa157. (A) Fluorescence means as measured by flow cytometry. (B) Expression levels of wt GFP and variants. (C) Chemiluminescent phage ELISA against streptavidin, using only phage displaying the 6-mer and 16-mer GFP variants and wt GFP.

As a final test, phage displaying wt GFP and the peptide-inserted GFP variants were produced in TG-1 cells containing two phagemids, one expressing the Sec-dependent Fos-p3 fusion and the other expressing the Tat-dependent GFP (or a peptide-inserted GFP variant) fused to Jun. The resulting phage was PEG precipitated and titered. Low phage titers (on the order of 10^{10} cfu/ml) were obtained for the wt GFP and each variant, with the exception of the 8-mer variant, which gave a 10-fold lower titer (10^{9} cfu/ml). It remains to be determined if the difference will be observed in the one phagemid system as well. The phage displaying the wt GFP, 6-mer variant, and 16-mer variant were then tested to ensure that the peptides could still bind streptavidin when constrained in the GFP loop. To test for this streptavidin-binding function, a sandwich ELISA was carried out. For more sensitive detection given the lower phage concentrations obtained when displaying GFP, the ELISA signal was detected using chemiluminescence. As seen in Figure 4.5C, the phage displaying the 6-mer GFP variant was indeed able to bind streptavidin but wt GFP was not. Similarly, no signal was detected with the phage expressing the GFP with a 16-mer inserts, which as discussed above is produced at a lower level.

4.4 Discussion

In this work, we set out to engineer a Tat-dependent filamentous phage display system. This was first attempted by using a Tat-specific signal peptide in place of the Sec signal peptide. A model protein that binds digoxin, scFv 26.10, was used. Expression of scFv-p3 fusions in strains having either a wt or oxidizing cytoplasm did not result in phage with a significant level of scFv display. We showed that the scFv moiety of the p3-fusion is compatible with Tat export. Finally, a Western blot confirmed that no p3 protein was found in the periplasm or membrane fraction when fused to the ssTorA signal peptide without the scFv. Taken together, these results reveal that the p3 protein of filamentous phage is incompatible with Tat export. It is unclear whether this is because p3 is unable to properly fold in the cytoplasm, is degraded in the cytoplasm before it can fold, or is unable to be exported via the Tat pathway for reasons unrelated to its folding state.

To circumvent these problems we devised a strategy whereby the p3 and the scFv protein are exported via separate routes and then associate once localized in the periplasmic space. Using this strategy, p3 is exported via the Sec pathway and fused to the leucine zipper domain of Fos, while the Tat pathway exports the protein of interest fused to the leucine zipper domain of Jun. Once in the periplasm, the leucine zipper domains form a complex and the phage particles produced by such a system display a Tat-exported protein of interest. Using GFP as a model protein, phage produced in this manner in a wt strain indeed had significantly higher fluorescence than the phage assembled either in a *tat* deficient strain or phage that did not display GFP (Figure 4.3C).

Importantly, the phage did not have any significant loss of infectivity, giving titers on the order of 10^{11} cfu/ml. It is interesting to note that phage produced in the *tat* strain did give some fluorescence over background (Figure 4.3C). The most likely explanation for this result is that many cells lyse during the 16-hour overnight phage production step, and these lysed cells release the Jun-GFP to the media where the leucine zipper domain of Jun can interact with the Fos domain that is on the already-assembled phage. Because *tat* strains are known to have severe outer membrane defects, it is possible that such background noise would be less in a wt strain, though that was not tested in this study. The successive enrichment rounds of a library screening process would almost certainly gradually eliminate the frequency of such false positives.

Given these encouraging results, we decided to utilize this Tat-dependent display system to engineer GFP to bind ligands. As a proof-of-concept experiment, streptavidin was chosen as the target ligand because it is recognized and is bound tightly by specific short peptides. A site that had proved permissive for peptide insertions in GFP, namely, aa157-158, was chosen and proved to be tolerant of 6-mer and 8-mer, but not 16-mer, peptide insertions with minimal loss of mean fluorescence and little effect on expression levels.

The titers of phage displaying the 8-mer GFP variant were lower than those of other fusions, and so it was not possible to test for binding ability. Phage titers for the other two variants and for wt GFP were two orders of magnitude lower than expected. The lower phage titer may in part be due to the fact that the phage was produced using two phagemids. This made addition of another antibiotic necessary, resulting in an

additional stress on the cells. The phage displaying the 16-mer GFP variant gave rise to negligible signal in the ELISA, but the 6-mer GFP variant phage yielded a strong signal. This result indicates that a peptide inserted into that particular GFP loop can have access to ligands even at lengths as short as 6 aa.

To fully test the utility of the Tat-dependent phage display system with GFP, it will be necessary to screen libraries of random peptides and isolate GFP variants that can bind to desired targets and also retain their fluorescence. To this end, a randomized 6-mer will be inserted at aa 157 of the GFP in a single-phagemid system. The library of phage particles produced will then be subjected to rounds of screening against streptavidin, interspersed with rounds of sorting via flow cytometry to maintain only the fluorescent molecules. It is expected that such a screen will yield peptides that have a known streptavidin-binding consensus sequence. Other ligands can then also be used as screening targets to demonstrate that utility is not limited to streptavidin experiments.

Chapter 5

Conclusions and Recommendations

In this work, different approaches were used to understand and utilize the Tat pathway. All of the putative Tat signal peptides of *E. coli* were characterized in an effort to uncover the natural Tat substrates and find an optimal Tat signal peptide for various applications. The requirement for folding prior to export via the Tat pathway was investigated through the use of proteins containing disulfide bonds. All of this information was utilized in the development of a Tat-dependent system for the screening of combinatorial libraries. These studies opened the way for the discovery of fluorescent protein variants that can bind to desired ligands.

The *E. coli* genome encodes at least 29 putative signal peptides containing a twin-arginine motif characteristic of proteins exported via the Tat pathway. Fusions of these putative signal peptides plus six to eight amino acids of the mature proteins to three reporter proteins (short-lived GFP, MBP and PhoA) and also data from the cell localization of epitope-tagged full-length proteins were employed to determine the ability of the 29 signal peptides to direct export through the Tat pathway, through the Sec pathway, or through both. It was found that 27/29 putative signal peptides could export one or more reporter protein through Tat. Of these, 11 signal peptides displayed Tat specificity in that they could not direct the export of Sec-only reporter proteins. The rest (16/27) were promiscuous in that they were capable of directing export of the appropriate

reporter either via Tat (GFP, MBP) or via Sec (PhoA, MBP). Mutations that conferred a = +1 charge to the N-terminus of the mature protein abolished or drastically reduced routing through the Sec pathway without affecting the ability to export via the Tat pathway. These experiments demonstrated that the charge of the mature protein N-terminus affects export promiscuity, independent of the effect of the folding state of the mature protein.

To examine the relationship between folding and export competence via the Tat pathway, we analyzed the subcellular localization of fusions between a set of eight Tat signal peptides and PhoA in isogenic *E. coli* strains that either allow or disfavor the formation of protein disulfide bonds in the cytoplasm. Export of these PhoA fusions via the Tat translocon was observed only in strains that enabled oxidative protein folding in the cytoplasm. Additionally, other disulfide-containing proteins, namely scFv and heterodimeric F_{AB} antibody fragments, were Tat export-competent only in strains having an oxidizing cytoplasm. Finally, functional, heterodimeric F_{AB} protein was exported from the cytoplasm by means of a Tat signal peptide fused to the heavy chain alone, indicating that the formation of a disulfide bonded dimer precedes export. These results demonstrated that *in vivo* only proteins that have attained the native conformation are exported via the Tat pathway, indicating that a folding quality control mechanism is intrinsic to the export process.

In an effort to develop a Tat-based display platform for the screening of combinatorial libraries, we showed that phage protein p3 is not compatible with Tat export. Thus a phage display system was devised in which p3 could be exported via the

Sec pathway and the protein to be displayed would be exported via the Tat pathway. To accomplish this, two proteins with domains that interact to form a leucine zipper pair (Jun and Fos) were used. In the system, the Tat-exported protein of interest was fused to the leucine zipper domain of Jun while the Sec-exported p3 was fused to the leucine zipper domain of Fos. When the two fusions reached the periplasm by the appropriate pathways, the coiled coils of Jun and Fos would interact. The phage particle would then be assembled with the Tat-exported protein of interest displayed on the p3 coat protein. Cysteines were engineered on either side of the leucine zipper domains to help eliminate dissociation of the non-covalent leucine zipper interaction. The system produced fluorescent phage particles when tested with GFP as a model protein. Furthermore, phage particles displaying a GFP variant containing a streptavidin-binding peptide within one of its loops were able to bind streptavidin with only a slight loss of fluorescence intensity. Thus a first step was taken towards the directed evolution of a ligand-binding GFP, an effort that could not be achieved using Sec-dependent display systems.

5.1 Recommendations for future studies

5.1.1 Examination of cofactorless, non-multimeric proteins and the Tat pathway

Nearly all of the putative signal peptides analyzed in Chapter 2 were found to be capable of Tat export and their corresponding natural substrates were therefore extremely likely to be Tat substrates. However, an explanation for why some of these proteins were Tat-dependent was not found; no apparent need for cytoplasmic cofactors, chaperones, or multimer associations was noted. In the first chapter, it is suggested that proteins that fold too quickly for Sec export may utilize the Tat pathway instead. If this is the case, it is likely that an SRP-dependent signal peptide would successfully be able to target such a Tat substrate to the periplasm. To investigate this idea, I examined the subcellular localization of the amidases AmiA and AmiC when fused to the SRP-dependent DsbA signal peptide, but found no tagged protein in any cellular compartment. Despite this negative result, it would be interesting to see if any other proteins exported via the Tat pathway are capable of being exported in an SRP-dependent manner. Alternatively, some Tat-exported proteins could fold so slowly that they are degraded in the periplasm if not exported in an already-folded state. To test this alternate hypothesis, either an SRP or a Sec signal peptide, such as ssPhoA or ssPelB, could be fused to the Tat substrates in question and localization could be examined in a strain devoid of the periplasmic protease DegP. Understanding why certain Tat substrates have evolved to use that more energetically-costly pathway could provide inspiration for other heterologous targets to be used with Tat-based expression or display systems.

5.1.2 Examination of biosynthetic biases imposed by the Sec and Tat pathways

With a similar goal in mind, the Tat-dependent phage display system could be put to use in determining what biosynthetic biases arise solely due to the export pathway utilized in directed evolution experiments. This could be accomplished by comparing the set of functional proteins obtained from screening a protein library for improved function via both a Sec and Tat-dependent display system. It would be expected that some differences would arise given the different protein folding requirements of each pathway, but such an experiment could also highlight differences that have been overlooked.

5.1.3 Engineering of GFP to bind complex ligands

Finally, while the 16-mer insert rendered GFP nearly non-fluorescent when inserted at aa 157, it is quite possible that another loop may be more permissive for such longer peptides. Therefore future efforts could investigate other flexible sites noted in the literature. The permissive loops confirmed in such a search can then be used simultaneously with aa157 insertions for engineering GFP to bind ligands that require more than a single peptide sequence for a tight interaction. It may also be possible to constrain the peptides on either side with cysteines, and express the variants in an oxidizing cytoplasm strain to ensure disulfide bond formation. This could potentially allow much longer peptides to be inserted, such as the complementarity determining regions of antibodies, without perturbing the structure and therefore fluorescence intensity of the GFP.

Appendix A

Table A.1. Oligonucleotides used in Chapter 2 studies

Number	Name	Sequence (5' to 3')
1	amiA for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGAGCACTTTTAAACCA
2	amiC for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGTCAGGATCCAACACTGC
3	cueO for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGCAACGTCGTGATTTC
4	dmsA for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGAAAACGAAAATCCCTGATG
5	fdnG for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGGACGTCAGTCGCAGA
6	fdoG for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGCAGGTCAGCAGAAGG
7	fhuD for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGAGCGGCTTACCTCTT
8	hyaA for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGAATAACGAGGAAACATTTTACCAG
9	hybA for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCGTGAACAGACGTAATTTTATTAAAGCAGCCTC
10	hybO for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGACTGGAGATAACACC
11	mdoD for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGGATCGTAGACGATTT
12	napA for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGAAACTCAGTCGTCGT
13	napG for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGTCCCGGTCAGCGAAA
14	nrfC for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGACCTGGTCTCGTCGC
15	sufI for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGTCACTCAGTCGGCGT
16	torZ for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGACATTAACAAGACGTGAATTT
17	wcaM for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGCCATTTAAAAAACTCTCCCGA
18	yaeI for	GCGATGCCCGGGTTAAAGAGGAGAAAGGTC ATGATTTC
19	yagT for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGAGCAACCAAGGCGAA
20	yahJ for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGAAAGAAAGCAATAGC
21	ycbK for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGGACAAATTCGACGCT
22	ycdB for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGCAGTATAAAGATGAAAACGG
23	ycdO for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGACCATTAACTTCCGCC
24	ydhX for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGTCATGGATAGGGTGG
25	yedY for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGAAAAAGAATCAATTTTTAAAAGAATC
26	yfhG for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGCGACACATTTTTCAA
27	ynfE for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGTCCAAAAATGAACGAATGGTG
28	ynfF for	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCATGATGAAAATCCATACCACAGAGGCG
29	amiA rev	GCGATGTCTAGATTTTAAAAGTTCGTCTTTGG
30	amiC rev	GCGATGTCTAGACGCCACGACCTGGCTGAC
31	cueO rev	GCGATGTCTAGATAACGTTGGGCGTTCTGC
32	dmsA rev	GCGATGTCTAGATGGAATGGCGCTATCGAC
33	fdnG rev	GCGATGTCTAGATAATTTGTAGTTTCGCGCCTG
34	fdoG rev	GCGATGTCTAGACAGTTTATACTGCCGGGTTTC
35	fhuD rev	GCGATGTCTAGAATTGGGATCAATAGCCGC
36	hyaA rev	GCGATGTCTAGAGCGCGGTTTGTTCTCCAG
37	hybA rev	GCGATGTCTAGATGGCGGGCGGTTTTCAGC
38	hybO rev	GCGATGTCTAGAAACCGATTCGGCCATCTC
39	mdoD rev	GCGATGTCTAGAGGCAATATCAGAATCTGC
40	napA rev	GCGATGTCTAGACTGACCAACAACGGCGCG
41	napG rev	GCGATGTCTAGAGCGCAACCGCACGCCAGAT
42	nrfC rev	GCGATGTCTAGAATTGATATTCAACGTTTTCGCCAC
43	sufI rev	GCGATGTCTAGACGGTTGCTGTTGCCCGGC
44	torZ rev	GCGATGTCTAGATTTACCGCCCTTCTCTTC

45	wcaM rev	GCGATGTCTAGAGCTTTGTCGGGCGGGAAG
46	yaeI rev	GCGATGTCTAGAAAGAATTTTGAATGGTGCTGC
47	yagT rev	GCGATGTCTAGAAGCTGCCGGAACGCTTGC
48	yahJ rev	GCGATGTCTAGATAGGGTGCCAGCTACCGC
49	ycbK rev	GCGATGTCTAGAGCGTGGGGTAGAGAGTGT
50	ycdB rev	GCGATGTCTAGACGCACTTTGCGTTTTTTG
51	ycdO rev	GCGATGTCTAGATTTGACCTGCGGCAC
52	ydhX rev	GCGATGTCTAGACTTTTCTTGCCTCGTGTT
53	yedY rev	GCGATGTCTAGAAAACCAGCTAAGCAGATC
54	yfhG rev	GCGATGTCTAGACGTATCAATGGCTGGCTT
55	ynfE rev	GCGATGTCTAGATTTTTCGCGGGCCTGTTG
56	ynfF rev	GCGATGTCTAGATTCTACCGGAGCCTCTGC
57	malE HindIII rev	GCGGCGAAGCTTTTACTTGGTGATACGAGTCTGCGC
58	malE for	GCGGCGTCTAGAAAAATCGAAGAAGGTAAACTGGTA
59	malE SphI rev	GCGGCGGCATGCGTCGACTTACTTGGTGATACGAGTCTGCGC
60	amiA for-m	GCGGCGTCATGAGCACTTTTAAACCA
61	amiC for-m	GCGGCGTCATGATATCAGGATCCAACACTGC
62	cueO for-m	GCGGCGTCATGATACAACGTCGTGATTTC
63	dmsA for-m	GCGGCGTCATGAAAACGAAAATCCCTGATG
64	fdnG for-m	GCGGCGTCATGATAGACGTCAGTCGCAGA
65	fdoG for-m	GCGGCGTCATGATACAGGTCAGCAGAAGG
66	fhuD for-m	GCGGCGTCATGAGCGGCTTACCTCTT
67	hyaA for-m	GCGGCGTCATGAATAACGAGGAAACATTTTACCAG
68	hybA for-m	GCGGCGTCATGAACAGACGTAATTTTATTAAAGCAGCCTC
69	hybO for-m	GCGGCGGAATTCACTGGAGATAACACC
70	mdoD for-m	GCGGCGTCATGATAGATCGTAGACGATTT
71	napA for-m	GCGGCGTCATGAAACTCAGTCGTCGT
72	napG for-m	GCGGCGTCATGATATCCCGGTCAGCGAAA
73	nrfC for-m	GCGGCGTCATGACCTGGTCTCGTCGC
74	sufI for-m	GCGGCGTCATGATATCACTCAGTCGGCGT
75	torZ for-m	GCGGCGTCATGATCAGGGAGGAAGTT
76	wcaM for-m	GCGGCGTCATGATACCATTTAAAAAACTCTCCCGA
77	yaeI for-m	GCGGCGTCATGATTTCACGCCGCCGA
78	yagT for-m	GCGGCGTCATGAGCAACCAAGGCGAA
79	yahJ for-m	GCGGCGTCATGAAAGAAAGCAATAGC
80	ycbK for-m	GCGGCGTCATGATAGACAAATTCGACGCT
81	ycdB for-m	GCGGCGTCATGATACAGTATAAAGATGAAAACGG
82	ycdO for-m	GCGGCGTCATGACCATTAACTTCCGCC
83	ydhX for-m	GCGGCGTCATGATATCATGGATAGGGTGG
84	yedY for-m	GCGGCGTCATGAAAAAGAATCAATTTTTAAAAGAATC
85	yfhG for-m	GCGGCGTCATGATACGACACATTTTTCAA
86	ynfE for-m	GCGGCGTCATGATATCCAAAAATGAACGAATGGTG
87	ynfF for-m	GCGGCGGAATTCATGATGAAAATCCATACCACAGAGGCG
88	phoA HindIII rev	GCGATGAAGCTTTTATTTCAGCCCCAGAGCGGCTT
89	phoA for	GCGATGTCTAGACGGACACCAGAAATGCCTGT
90	phoA SphI rev	GCGATGAAGCTTGCATGCTTATTTCAGCCCCAGAGCGGCTT
91	ssphoA for	GCGACGTCATGAAACAAAGCACTATTGCAC
92	ssphoA rev	GCGACGTCTAGAGGCTTTTGTCACAGGGG

93	mdoD +2 rev	GCGATGTCTAGAGGCAATACGAGATTTTGCCGC
94	mdoD +1 rev	GCGATGTCTAGATTTTGCCGCGAATGCC
95	nrfC FLAG rev	GCTCTAGATTATCCCTTGTCGTCATCGTCCTTGTAGTCTGCTCCTTGGCTCACCTCCCCGTA
96	cueO FLAG rev	GCTCTAGATTATCCCTTGTCGTCATCGTCCTTGTAGTCTGCTCCTACCGTAAACCCTAACAT
97	mdoD FLAG rev	GCTCTAGATTATCCCTTGTCGTCATCGTCCTTGTAGTCTGCTCCACTCATCACGCGGTC
98	yfhG FLAG rev	GCTCTAGATTATCCCTTGTCGTCATCGTCCTTGTAGTCTGCTCCTGGCTCATCAGGAGT
99	MdoD(no ss) for	GCGATGTCTAGAGATTCTGATATTGCCGACGGG
100	ssphoA for SacI	GCGATGGAGCTCTTAAAGAGGAGAAAGGTCGTGAAAC

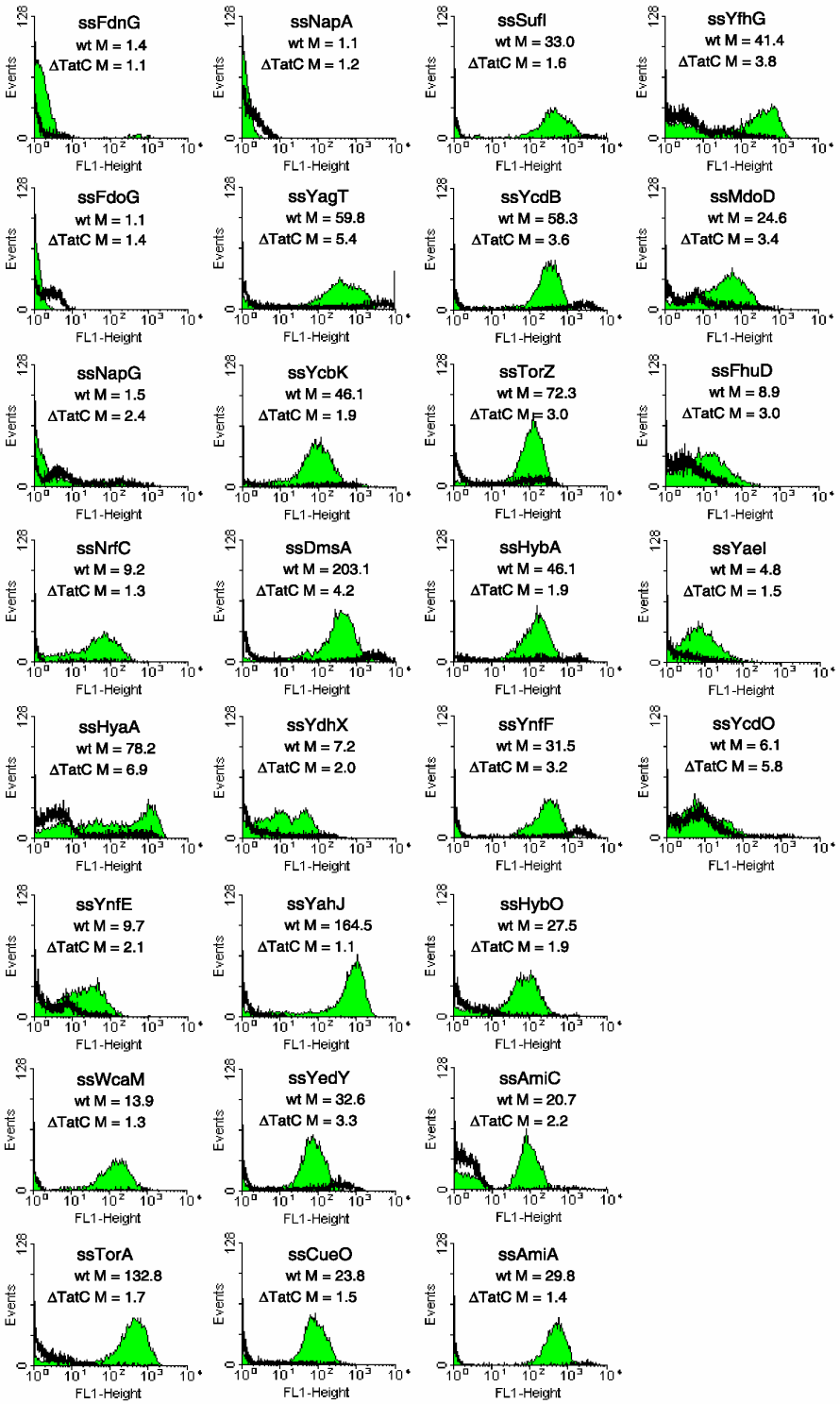

Figure A.1. FACS histograms for each signal peptide-GFP-SsrA fusion. Means are shown for both wt and *tat* strains containing each fusion.

Glossary

BLAST *B*asic *l*ocal *a*lignment *s*earch *t*ool – a program available from NCBI that searches a query peptide or nucleotide sequence against databases of sequences and provides the most statistically significant matches.

ELISA *E*nzyme-*l*inked *i*mmuno*s*orbent *a*ssay - an extremely sensitive assay for detection and quantitation of either an antigen or an antibody via the use of an enzyme covalently attached to an antibody and substrate that forms a colored reaction product in the presence of that enzyme.

F_{AB} – A heterodimeric ligand-binding protein in which the heavy chain (V_H-C_{H1}) and light chain (V_L-C_L) of an antibody are linked by an intermolecular disulfide bond.

FACS *F*luorescence-*a*ctivated *c*ell *s*orting – a powerful method of analyzing cells in which a fluid stream containing the cells is passed through a laser, causing light scattering; also, any fluorescent molecules present in or on the cells will become excited and emit light, and the resulting scattering and emission is measured.

FLAG tag – A particular peptide sequence, D-Y-K-D-D-D-D-K, that is often fused to proteins for ease of purification and analysis via a specific antibody to the sequence.

HRP *H*orse*r*adish *p*eroxidase – the enzyme linked to secondary antibodies used in ELISA and Western blot applications, catalyzing a colorimetric or chemiluminescent reaction.

Phagemid – an extra-chromosomal, circular piece of double stranded DNA containing at least one phage gene

Pmf *P*roton *m*otive *f*orce – the potential energy of a proton gradient; it comprises an electrical part and a chemical part due to differences in the charge and solute concentration, respectively, across the inner membrane. It can be tapped for energy by allowing the protons to flow in a controlled manner across the membrane.

REMP *R*edox *e*nzyme *m*aturation *p*rotein – a cytoplasmic chaperone-like protein that functions in the assembly of cofactor-containing and/or multimeric Tat substrates and prevents targeting to the translocase until assembly is complete.

scFv Single-chain variable fragment – the minimum component of an antibody necessary for ligand-binding, consisting of a variable heavy chain and a variable light chain connected by a small linker.

Sec General *Sec*retory pathway – in gram-negative bacteria, a pathway used to transfer unfolded polypeptides from the cytoplasmic compartment to the periplasmic space, where folding can then take place.

Signal peptide – in bacteria, a peptide sequence fused to the n-terminus of a protein that functions to direct that protein to the periplasm of the cell via the proper export pathway.

SRP *S*ignal *r*ecognition *p*article pathway – in bacteria, a pathway used to transfer unfolded polypeptides to the cytoplasmic (inner) membrane or periplasmic space in a co-translational manner, involving use of the Sec machinery as well as the signal recognition particle.

SsrA tag – A particular polypeptide sequence that, when fused to a protein, targets it for degradation by the ClpAP/ClpXP apparatus.

Tat *T*win-*a*rginine *t*ranslocation pathway – a pathway used to transfer proteins already in a folded state to the periplasmic space.

Author's Publication List

Tullman-Ercek, D., DeLisa, M.P., Kawarasaki, Y., Iranpour, P., Ribnicky, B., Palmer, T., Georgiou, G. (Submitted). "Export pathway selectivity of *Escherichia coli* twin-arginine translocation signal peptides."

Lee, P. A., D. Tullman-Ercek, Georgiou G. (2006). "The bacterial twin-arginine translocation pathway." *Annu. Rev. Microbiol.* **60** (In Press).

DeLisa, M. P., D. Tullman, 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**(10): 6115-6120.

References

Abedi, M. R., Caponigro, G. and Kamb, A. (1998). "Green fluorescent protein as a scaffold for intracellular presentation of peptides." *Nucleic Acids Res.* **26**(2): 623-630.

Alami, M., Luke, I., Deitermann, S., Eisner, G., Koch, H. G., Brunner, J. and Muller, M. (2003). "Differential interactions between a twin-arginine signal peptide and its translocase in Escherichia coli." *Mol. Cell* **12**(4): 937-946.

Alami, M., Trescher, D., Wu, L. F. and Muller, M. (2002). "Separate analysis of twin-arginine translocation (Tat)-specific membrane binding and translocation in Escherichia coli." *J. Biol. Chem.* **277**(23): 20499-20503.

Alder, N. N. and Theg, S. M. (2003a). "Energetics of protein transport across biological membranes: A study of the thylakoid Delta pH-dependent/cpTat pathway." *Cell* **112**(2): 231-242.

Alder, N. N. and Theg, S. M. (2003b). "Energy use by biological protein transport pathways." *Trends Biochem. Sci.* **28**(8): 442-51.

Allen, S. C. H., Barrett, C. M. L., Ray, N. and Robinson, C. (2002). "Essential cytoplasmic domains in the Escherichia coli TatC protein." *J. Biol. Chem.* **277**(12): 10362-10366.

Andersson, H. and von Heijne, G. (1991). "A 30-residue-long "export initiation domain" adjacent to the signal sequence is critical for protein translocation across the inner membrane of Escherichia coli." *Proc. Natl. Acad. Sci. USA* **88**(21): 9751-4.

Angelini, S., Moreno, R., Gouffi, K., Santini, C. L., Yamagishi, A., Berenguer, J. and Wu, L. F. (2001). "Export of Thermus thermophilus alkaline phosphatase via the twin-arginine translocation pathway in Escherichia coli." *FEBS Lett.* **506**(2): 103-107.

Arndt, K. M., Jung, S., Krebber, C. and Pluckthun, A. (2000). "Selectively infective phage technology." *Methods Enzymol.* **328**: 364-388.

Arnold, F. H. and Georgiou, G. (2003). Directed Enzyme Evolution: Screening and Selection Methods. Totowa, New Jersey, Humana Press.

Baird, G. S., Zacharias, D. A. and Tsien, R. Y. (1999). "Circular permutation and receptor insertion within green fluorescent proteins." *Proc. Natl. Acad. Sci. U. S. A.* **96**(20): 11241-11246.

Barbas, C. F., 3rd, Kang, A. S., Lerner, R. A. and Benkovic, S. J. (1991). "Assembly of combinatorial antibody libraries on phage surfaces: the gene III site." *Proc. Natl. Acad. Sci. U. S. A.* **88**(18): 7978-82.

Bardwell, J. C. A., McGovern, K. and Beckwith, J. (1991). "Identification of a protein required for disulfide bond formation in vivo." *Cell* **67**(3): 581-589.

Barrett, C. M. L., Ray, N., Thomas, J. D., Robinson, C. and Bolhuis, A. (2003). "Quantitative export of a reporter protein, GFP, by the twin-arginine translocation pathway in Escherichia coli." *Biochem. Biophys. Res. Commun.* **304**(2): 279-284.

Barrett, C. M. L. and Robinson, C. (2005). "Evidence for interactions between domains of TatA and TatB from mutagenesis of the TatABC subunits of the twin-arginine translocase." *FEBS J.* **272**(9): 2261-2275.

Behrendt, J., Standar, K., Lindenstrauss, U. and Bruser, T. (2004). "Topological studies on the twin-arginine translocase component TatC." *FEMS Microbiol. Lett.* **234**(2): 303-308.

Bendtsen, J. D., Nielsen, H., Widdick, D., Palmer, T. and Brunak, S. (2005). "Prediction of twin-arginine signal peptides." *BMC Bioinformatics* **6**: 167.

Benhar, I. (2001). "Biotechnological applications of phage and cell display." *Biotechnol. Adv.* **19**(1): 1-33.

Berks, B. C. (1996). "A common export pathway for proteins binding complex redox cofactors?" *Mol. Microbiol.* **22**(3): 393-404.

Berks, B. C., Palmer, T. and Sargent, F. (2003). "The Tat protein translocation pathway and its role in microbial physiology." *Advances in Microbial Physiology, Vol 47* **47**: 187-254.

Berks, B. C., Palmer, T. and Sargent, F. (2005). "Protein targeting by the bacterial twin-arginine translocation (Tat) pathway." *Curr. Opin. Microbiol.* **8**(2): 174-181.

Bernhardt, T. G. and de Boer, P. A. J. (2003). "The Escherichia coli amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway." *Mol. Microbiol.* **48**(5): 1171-1182.

Bessette, P. H., Aslund, F., Beckwith, J. and Georgiou, G. (1999). "Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm." *Proc. Natl. Acad. Sci. U. S. A.* **96**(24): 13703-13708.

Blaudeck, N., Kreutzenbeck, P., Freudl, R. and Sprenger, G. A. (2003). "Genetic analysis of pathway specificity during posttranslational protein translocation across the Escherichia coli plasma membrane." *J. Bacteriol.* **185**(9): 2811-2819.

Blaudeck, N., Sprenger, G. A., Freudl, R. and Wiegert, T. (2001). "Specificity of signal peptide recognition in Tat-dependent bacterial protein translocation." *J. Bacteriol.* **183**(2): 604-610.

Blazek, D. and Celer, V. (2003). "The production and application of single-chain antibody fragments." *Folia Microbiol. (Praha).* **48**(5): 687-698.

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

Bogsch, E. G., Sargent, F., Stanley, N. R., Berks, B. C., Robinson, C. and Palmer, T. (1998). "An essential component of a novel bacterial protein export system with homologues in plastids and mitochondria." *J. Biol. Chem.* **273**(29): 18003-18006.

Bolhuis, A., Bogsch, E. G. and Robinson, C. (2000). "Subunit interactions in the twin-arginine translocase complex of Escherichia coli." *FEBS Lett.* **472**(1): 88-92.

Bolhuis, A., Mathers, J. E., Thomas, J. D., Barrett, C. M. L. and Robinson, C. (2001). "TatB and TatC form a functional and structural unit of the twin-arginine translocase from Escherichia coli." *J. Biol. Chem.* **276**(23): 20213-20219.

Bronstein, P., Marrichi, M. and DeLisa, M. P. (2004). "Dissecting the twin-arginine translocation pathway using genome-wide analysis." *Res. Microbiol.* **155**(10): 803-810.

Bronstein, P. A., Marrichi, M., Cartinhour, S., Schneider, D. J. and DeLisa, M. P. (2005). "Identification of a twin-arginine translocation system in Pseudomonas syringae pv. tomato DC3000 and its contribution to pathogenicity and fitness." *J. Bacteriol.* **187**(24): 8450-8461.

Buchanan, G., de Leeuw, E., Stanley, N. R., Wexler, M., Berks, B. C., Sargent, F. and Palmer, T. (2002). "Functional complexity of the twin-arginine translocase TatC component revealed by site-directed mutagenesis." *Mol. Microbiol.* **43**(6): 1457-1470.

Buchanan, G., Sargent, F., Berks, B. C. and Palmer, T. (2001). "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." *Arch. Microbiol.* **177**(1): 107-112.

Caldelari, I., Mann, S., Crooks, C. and Palmer, T. (2006). "The tat pathway of the plant pathogen Pseudomonas syringae is required for optimal virulence." *Mol. Plant. Microbe Interact.* **19**(2): 200-212.

Campbell, N. A. a. R., J.B. (2001). Membrane Structure and Function. Biology, Benjamin Cummings. **1:** 138-155.

Cetinkaya, M., Zeytun, A., Sofo, J. and Demirel, M. C. (2006). "How do insertions affect green fluorescent protein?" *Chem. Phys. Lett.* **419**(1-3): 48-54.

Chanal, A., Santini, C. L. and Wu, L. F. (1998). "Potential receptor function of three homologous components, TatA, TatB and TatE, of the twin-arginine signal sequence-dependent metalloenzyme translocation pathway in Escherichia coli." *Mol. Microbiol.* **30**(3): 674-676.

Chanal, A., Santini, C. L. and Wu, L. F. (2003). "Specific inhibition of the translocation of a subset of Escherichia coli TAT substrates by the TorA signal peptide." *J. Mol. Biol.* **327**(3): 563-570.

Chen, G., Hayhurst, A., Thomas, J. G., Harvey, B. R., Iverson, B. L. and Georgiou, G. (2001). "Isolation of high-affinity ligand-binding proteins by periplasmic expression with cytometric screening (PECS)." *Nat. Biotechnol.* **19**(6): 537-542.

Chen, W. and Georgiou, G. (2002). "Cell-surface display of heterologous proteins: From high-throughput screening to environmental applications." *Biotechnol. Bioeng.* **79**(5): 496-503.

Cianciotto, N. P. (2005). "Type II secretion: a protein secretion system for all seasons." *Trends Microbiol.* **13**(12): 581-588.

Cirino, P. C., Mayer, K. M. and Umeno, D. (2003). "Generating mutant libraries using error-prone PCR." *Methods Mol. Biol.* **231**: 3-9.

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**(5): 923-934.

Cline, K., Ettinger, W. F. and Theg, S. M. (1992). "Protein-specific energy requirements for protein transport across or into thylakoid membranes. Two lumenal proteins are transported in the absence of ATP." *J. Biol. Chem.* **267**(4): 2688-96.

Cline, K. and Mori, H. (2001). "Thylakoid Delta pH-dependent precursor proteins bind to a cpTatC-Hcf106 complex before Tha4-dependent transport." *J. Cell Biol.* **154**(4): 719-729.

Coleman, J., Inukai, M. and Inouye, M. (1985). "Dual functions of the signal peptide in protein transfer across the membrane." *Cell* **43**(1): 351-360.

Creighton, A. M., Hulford, A., Mant, A., Robinson, D. and Robinson, C. (1995). "A monomeric, tightly folded stromal intermediate on the delta pH-dependent thylakoidal protein transport pathway." *J. Biol. Chem.* **270**(4): 1663-9.

Cristobal, S., de Gier, J. W., Nielsen, H. and von Heijne, G. (1999). "Competition between Sec- and Tat-dependent protein translocation in Escherichia coli." *EMBO J.* **18**(11): 2982-2990.

Dabney-Smith, C., Mori, H. and Cline, K. (2003). "Requirement of a Tha4-conserved transmembrane glutamate in thylakoid Tat translocase assembly revealed by biochemical complementation." *J. Biol. Chem.* **278**(44): 43027-33.

Dalbey, R. E. and Robinson, C. (1999). "Protein translocation into and across the bacterial plasma membrane and the plant thylakoid membrane." *Trends Biochem. Sci.* **24**(1): 17-22.

Davidson, A. L., Shuman, H. A. and Nikaido, H. (1992). "Mechanism of maltose transport in Escherichia coli: transmembrane signaling by periplasmic binding proteins." *Proc. Natl. Acad. Sci. U. S. A.* **89**(6): 2360-2364.

De Buck, E., Maes, L., Meyen, E., Van Mellaert, L., Geukens, N., Anne, J. and Lammertyn, E. (2005). "Legionella pneumophila Philadelphia-1 tatB and tatC

affect intracellular replication and biofilm formation." *Biochem. Biophys. Res. Commun.* **331**(4): 1413-1420.

de Leeuw, E., Granjon, T., Porcelli, I., Alami, M., Carr, S. B., Muller, M., Sargent, F., Palmer, T. and Berks, B. C. (2002). "Oligomeric properties and signal peptide binding by Escherichia coli Tat protein transport complexes." *J. Mol. Biol.* **322**(5): 1135-1146.

De Leeuw, E., Porcelli, I., Sargent, F., Palmer, T. and Berks, B. C. (2001). "Membrane interactions and self-association of the TatA and TatB components of the twin-arginine translocation pathway." *FEBS Lett.* **506**(2): 143-148.

DeLisa, M. P., Lee, P., Palmer, T. and Georgiou, G. (2004). "Phage shock protein PspA of Escherichia coli relieves saturation of protein export via the Tat pathway." *J. Bacteriol.* **186**(2): 366-373.

DeLisa, M. P., Samuelson, P., Palmer, T. and Georgiou, G. (2002). "Genetic analysis of the twin arginine translocator secretion pathway in bacteria." *J. Biol. Chem.* **277**(33): 29825-29831.

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**(10): 6115-6120.

Derman, A. I., Prinz, W. A., Belin, D. and Beckwith, J. (1993). "Mutations that allow disulfide bond formation in the cytoplasm of Escherichia coli." *Science* **262**(5140): 1744-1747.

Di Cola, A., Bailey, S. and Robinson, C. (2005). "The thylakoid Delta pH/Delta Psi are not required for the initial stages of Tat-dependent protein transport in tobacco protoplasts." *J. Biol. Chem.* **280**(50): 41165-41170.

Dilks, K., Rose, R. W., Hartmann, E. and Pohlschroder, M. (2003). "Prokaryotic utilization of the twin-arginine translocation pathway: a genomic survey." *J. Bacteriol.* **185**(4): 1478-1483.

Dilks, M., Gimenez, M. I. and Pohlschroder, M. (2005). "Genetic and biochemical analysis of the twin-arginine translocation pathway in halophilic archaea." *J. Bacteriol.* **187**(23): 8104-8113.

Ding, Z. Y. and Christie, P. J. (2003). "Agrobacterium tumefaciens twin-arginine-dependent translocation is important for virulence, flagellation, and chemotaxis but not type IV secretion." *J. Bacteriol.* **185**(3): 760-771.

Droge, M. J., Boersma, Y. L., Braun, P. G., Buining, R. J., Julsing, M. K., Selles, K. G., van Dijl, J. M. and Quax, W. J. (2006). "Phage display of an intracellular carboxylesterase of Bacillus subtilis: Comparison of Sec and Tat pathway export capabilities." *Appl. Environ. Microbiol.* **72**(7): 4589-95.

Economou, A. and Wickner, W. (1994). "SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion." *Cell* **78**(5): 835-843.

Eilers, M. and Schatz, G. (1986). "Binding of a specific ligand inhibits import of a purified precursor protein into mitochondria." *Nature* **322**(6076): 228-232.

Endemann, H. and Model, P. (1995). "Location of filamentous phage minor coat proteins in phage and in infected cells." *J. Mol. Biol.* **250**(4): 496-506.

Feilmeier, B. J., Iseminger, G., Schroeder, D., Webber, H. and Phillips, G. J. (2000). "Green fluorescent protein functions as a reporter for protein localization in Escherichia coli." *J. Bacteriol.* **182**(14): 4068-76.

Finazzi, G., Chasen, C., Wollman, F. A. and de Vitry, C. (2003). "Thylakoid targeting of Tat passenger proteins shows no delta pH dependence in vivo." *EMBO J.* **22**(4): 807-15.

Fisher, A. C., Kim, W. and DeLisa, M. P. (2006). "Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway." *Protein Sci.* **15**(3): 449-458.

Gao, C. S., Mao, S. L., Lo, C. H. L., Wirsching, P., Lerner, R. A. and Janda, K. D. (1999). "Making artificial antibodies: A format for phage display of combinatorial heterodimeric arrays." *Proc. Natl. Acad. Sci. U. S. A.* **96**(11): 6025-6030.

Geller, B., Zhu, H. Y., Cheng, S., Kuhn, A. and Dalbey, R. E. (1993). "Charged residues render pro-OmpA potential dependent for initiation of membrane translocation." *J. Biol. Chem.* **268**(13): 9442-7.

Gennis, R. B. and Stewart, V. (1996). Respiration. Escherichia coli and Salmonella: cellular and molecular biology. Neidhardt, F. C., III., C. R., Ingraham, J. L., Lin, E. C. C. and Low, K. B., American Society for Microbiology. **1:** 217-261.

Georgiou, G. and Segatori, L. (2005). "Preparative expression of secreted proteins in bacteria: status report and future prospects." *Curr. Opin. Biotechnol.* **16**(5): 538-545.

Gohlke, U., Pullan, L., McDevitt, C. A., Porcelli, I., de Leeuw, E., Palmer, T., Saibil, H. R. and Berks, B. C. (2005). "The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter." *Proc. Natl. Acad. Sci. U. S. A.* **102**(30): 10482-10486.

Gon, S., Patte, J. C., Mejean, V. and Iobbi-Nivol, C. (2000). "The torYZ (yecK bisZ) operon encodes a third respiratory trimethylamine N-oxide reductase in Escherichia coli." *J. Bacteriol.* **182**(20): 5779-5786.

Griffiths, A. D., Williams, S. C., Hartley, O., Tomlinson, I. M., Waterhouse, P., Crosby, W. L., Kontermann, R. E., Jones, P. T., Low, N. M., Allison, T. J., Prospero, T. D., Hoogenboom, H. R., Nissim, A., Cox, J. P. L., Harrison, J. L., Zaccolo, M., Gherardi, E. and Winter, G. (1994). "Isolation of high-affinity human antibodies directly from large synthetic repertoires." *EMBO J.* **13**(14): 3245-3260.

Guzman, L. M., Belin, D., Carson, M. J. and Beckwith, J. (1995). "Tight regulation, modulation, and high-level expression by vectors containing the arabinose pBAD promoter." *J. Bacteriol.* **177**(14): 4121-30.

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 binding." *Eur. J. Biochem.* **263**(2): 543-551.

Hamilton, C. M., Aldea, M., Washburn, B. K., Babitzke, P. and Kushner, S. R. (1989). "New Method For Generating Deletions And Gene Replacements In Escherichia Coli." *J. Bacteriol.* **171**(9): 4617-4622.

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

Hinsley, A. P. and Berks, B. C. (2002). "Specificity of respiratory pathways involved in the reduction of sulfur compounds by Salmonella enterica." *Microbiology* **148**(Pt 11): 3631-8.

Hinsley, A. P., Stanley, N. R., Palmer, T. and Berks, B. C. (2001). "A naturally occurring bacterial Tat signal peptide lacking one of the 'invariant' arginine residues of the consensus targeting motif." *FEBS Lett.* **497**(1): 45-49.

Hoogenboom, H. R., de Bruine, A. P., Hufton, S. E., Hoet, R. M., Arends, J. W. and Roovers, R. C. (1998). "Antibody phage display technology and its applications." *Immunotechnology* **4**(1): 1-20.

Huber, D., Boyd, D., Xia, Y., Olma, M. H., Gerstein, M. and Beckwith, J. (2005). "Use of thioredoxin as a reporter to identify a subset of Escherichia coli signal sequences that promote signal recognition particle-dependent translocation." *J. Bacteriol.* **187**(9): 2983-91.

Hufton, S. E., Moerkerk, P. T., Meulemans, E. V., de Bruine, A., Arends, J. W. and Hoogenboom, H. R. (1999). "Phage display of cDNA repertoires: the pVI display system and its applications for the selection of immunogenic ligands." *J. Immunol. Methods* **231**(1-2): 39-51.

Huston, J. S., Levinson, D., Mudgetthunter, M., Tai, M. S., Novotny, J., Margolies, M. N., Ridge, R. J., Bruccoleri, R. E., Haber, E., Crea, R. and Oppermann, H. (1988). "Protein engineering of antibody-binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analog produced in Escherichia coli." *Proc. Natl. Acad. Sci. U. S. A.* **85**(16): 5879-5883.

Hutcheon, G. W. and Bolhuis, A. (2003). "The archaeal twin-arginine translocation pathway." *Biochem. Soc. Trans.* **31**: 686-689.

Hynds, P. J., Robinson, D. and Robinson, C. (1998). "The Sec-independent twin-arginine translocation system can transport both tightly folded and malfolded proteins across the thylakoid membrane." *J. Biol. Chem.* **273**(52): 34868-34874.

Ignatova, Z., Hornle, C., Nurk, A. and Kasche, V. (2002). "Unusual signal peptide directs penicillin amidase from Escherichia coli to the tat translocation machinery." *Biochem. Biophys. Res. Commun.* **291**(1): 146-149.

Izard, J. W. and Kendall, D. A. (1994). "Signal peptides: exquisitely designed transport promoters." *Mol. Microbiol.* **13**(5): 765-73.

Ize, B., Gerard, F. and Wu, L. F. (2002). "In vivo assessment of the Tat signal peptide specificity in Escherichia coli." *Arch. Microbiol.* **178**(6): 548-553.

Ize, B., Gerard, F., Zhang, M., Chanal, A., Voulhoux, R., Palmer, T., Filloux, A. and Wu, L. F. (2002). "In vivo dissection of the tat translocation pathway in Escherichia coli." *J. Mol. Biol.* **317**(3): 327-335.

Ize, B., Porcelli, I., Lucchini, S., Hinton, J. C., Berks, B. C. and Palmer, T. (2004). "Novel phenotypes of Escherichia coli tat mutants revealed by global gene expression and phenotypic analysis." *J. Biol. Chem.* **279**(46): 47543-47554.

Ize, B., Stanley, N. R., Buchanan, G. and Palmer, T. (2003). "Role of the Escherichia coli Tat pathway in outer membrane integrity." *Mol. Microbiol.* **48**(5): 1183-1193.

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**(20): 3962-3972.

Jack, R. L., Sargent, F., Berks, B. C., Sawers, G. and Palmer, T. (2001). "Constitutive expression of Escherichia coli tat genes indicates an important role for the twin-arginine translocase during aerobic and anaerobic growth." *J. Bacteriol.* **183**(5): 1801-1804.

Jongbloed, J. D. H., Antelmann, H., Hecker, M., Nijland, R., Bron, S., Airaksinen, U., Pries, F., Quax, W. J., van Dijl, J. M. and Braun, P. G. (2002). "Selective contribution of the twin-arginine translocation pathway to protein secretion in Bacillus subtilis." *J. Biol. Chem.* **277**(46): 44068-44078.

Kajava, A. V., Zolov, S. N., Kalinin, A. E. and Nesmeyanova, M. A. (2000). "The net charge of the first 18 residues of the mature sequence affects protein translocation

across the cytoplasmic membrane of gram-negative bacteria." *J. Bacteriol.* **182**(8): 2163-9.

Kang, D. G., Lim, G. B. and Cha, H. J. (2005). "Functional periplasmic secretion of organophosphorous hydrolase using the twin-arginine translocation pathway in Escherichia coli." *J. Biotechnol.* **118**(4): 379-385.

Kebir, M. O. and Kendall, D. A. (2002). "SecA specificity for different signal peptides." *Biochemistry (Mosc).* **41**(17): 5573-80.

Keegstra, K. and Cline, K. (1999). "Protein import and routing systems of chloroplasts." *Plant Cell* **11**(4): 557-570.

Ki, J. J., Kawarasaki, Y., Gam, J., Harvey, B. R., Iverson, B. L. and Georgiou, G. (2004). "A periplasmic fluorescent reporter protein and its application in high-throughput membrane protein topology analysis." *J. Mol. Biol.* **341**(4): 901-9.

Kim, J. and Kendall, D. A. (2000). "Sec-dependent protein export and the involvement of the molecular chaperone SecB." *Cell Stress Chaperones* **5**(4): 267-275.

Kim, J. Y., Fogarty, E. A., Lu, F. J., Zhu, H., Wheelock, G. D., Henderson, L. A. and DeLisa, M. P. (2005). "Twin-arginine translocation of active human tissue plasminogen activator in Escherichia coli." *Appl. Environ. Microbiol.* **71**(12): 8451-8459.

Kouzarides, T. and Ziff, E. (1988). "The role of the leucine zipper in the fos-jun interaction." *Nature* **336**(6200): 646-51.

Kuhn, A. (1995). "Major coat proteins of bacteriophage-Pf3 and bacteriophage-M13 as model systems for Sec-independent protein transport." *FEMS Microbiol. Rev.* **17**(1-2): 185-190.

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

Lee, P. A., Tullman-Ercek, D. and Georgiou, G. (2006). "The bacterial twin-arginine translocation pathway." *Annu. Rev. Microbiol.* **60**.

Lequette, Y., Odberg-Ferragut, C., Bohin, J. P. and Lacroix, J. M. (2004). "Identification of mdoD an mdoG paralog which encodes a twin-arginine-dependent periplasmic protein that controls osmoregulated periplasmic glucan backbone structures." *J. Bacteriol.* **186**(12): 3695-3702.

Levy, R., Weiss, R., Chen, G., Iverson, B. L. and Georgiou, G. (2001). "Production of correctly folded Fab antibody fragment in the cytoplasm of Escherichia coli trxB gor mutants via the coexpression of molecular chaperones." *Protein Expr. Purif.* **23**(2): 338-347.

Li, P., Beckwith, J. and Inouye, H. (1988). "Alteration of the amino terminus of the mature sequence of a periplasmic protein can severely affect protein export in Escherichia coli." *Proc. Natl. Acad. Sci. U. S. A.* **85**(20): 7685-7689.

Li, S. Y., Chang, B. Y. and Lin, S. C. (2006). "Coexpression of TorD enhances the transport of GFP via the TAT pathway." *J. Biotechnol.* **122**(4): 412-421.

Lopilato, J., Bortner, S. and Beckwith, J. (1986). "Mutations in a new chromosomal gene of Escherichia coli K-12, pcnB, reduce plasmid copy number af pBR322 and its derivatives." *Mol. Gen. Genet.* **205**(2): 285-290.

Mabry, R., Brasky, K., Geiger, R., Carrion, R., Jr., Hubbard, G. B., Leppla, S., Patterson, J. L., Georgiou, G. and Iverson, B. L. (2006). "Detection of anthrax toxin in the serum of animals infected with Bacillus anthracis by using engineered immunoassays." *Clin. Vaccine Immunol.* **13**(6): 671-7.

Manheim, B. S., Granahan, P. and Dow, K. J. (2006). "'Follow-on biologics': Ensuring continued innovation in the biotechnology industry." *Health Aff. (Millwood).* **25**(2): 394-404.

Maynard, J. and Georgiou, G. (2000). "Antibody engineering." *Annu. Rev. Biomed. Eng.* **2**: 339-376.

McDonough, J. A., Hacker, K. E., Flores, A. R., Pavelka, M. S. and Braunstein, M. (2005). "The twin-arginine translocation pathway of Mycobactetium smegmatis is functional and required for the export of mycobacterial beta-lactamases." *J. Bacteriol.* **187**(22): 7667-7679.

Mori, H. and Cline, K. (2002). "A twin arginine signal peptide and the pH gradient trigger reversible assembly of the thylakoid Delta pH/Tat translocase." *J. Cell Biol.* **157**(2): 205-210.

Mould, R. M. and Robinson, C. (1991). "A proton gradient is required for the transport of two lumenal oxygen-evolving proteins across the thylakoid membrane." *J. Biol. Chem.* **266**(19): 12189-93.

Musser, S. M. and Theg, S. M. (2000). "Proton transfer limits protein translocation rate by the thylakoid DeltapH/Tat machinery." *Biochemistry (Mosc).* **39**(28): 8228-33.

Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997). "A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites." *Int. J. Neural Syst.* **8**(5-6): 581-99.

Nouwen, N., de Kruijff, B. and Tommassen, J. (1996). "prlA suppressors in Escherichia coli relieve the proton electrochemical gradient dependency of translocation of wild-type precursors." *Proc. Natl. Acad. Sci. U. S. A.* **93**(12): 5953-7.

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

Oates, J., Mathers, J., Mangels, D., Kuhlbrandt, W., Robinson, C. and Model, K. (2003). "Consensus structural features of purified bacterial TatABC complexes." *J. Mol. Biol.* **330**(2): 277-286.

Ochsner, U. A., Snyder, A., Vasil, A. I. and Vasil, M. L. (2002). "Effects of the twin-arginine translocase on secretion of virulence factors, stress response, and pathogenesis." *Proc. Natl. Acad. Sci. U. S. A.* **99**(12): 8312-8317.

Ohkawa, I. and Webster, R. E. (1981). "The orientation of the major coat protein of bacteriophage-f1 in the cytoplasmic membrane of Escherichia coli." *J. Biol. Chem.* **256**(19): 9951-9958.

Olson, J. W. and Maier, R. J. (2002). "Molecular hydrogen as an energy source for Helicobacter pylori." *Science* **298**(5599): 1788-90.

Oresnik, I. J., Ladner, C. L. and Turner, R. J. (2001). "Identification of a twin-arginine leader-binding protein." *Mol. Microbiol.* **40**(2): 323-331.

Ostroff, R. M., Vasil, A. I. and Vasil, M. L. (1990). "Molecular comparison of a nonhemolytic and a hemolytic phospholipase C from Pseudomonas aeruginosa." *J. Bacteriol.* **172**(10): 5915-23.

Palmer, T., Sargent, F. and Berks, B. C. (2005). "Export of complex cofactor-containing proteins by the bacterial Tat pathway." *Trends Microbiol.* **13**(4): 175-180.

Paschke, M. and Hohne, W. (2005). "A twin-arginine translocation (Tat)-mediated phage display system." *Gene* **350**(1): 79-88.

Peters, E. A., Schatz, P. J., Johnson, S. S. and Dower, W. J. (1994). "Membrane insertion defects caused by positive charges in the early mature region of protein pIII of filamentous phage-fd can be corrected by PrlA suppressors." *J. Bacteriol.* **176**(14): 4296-4305.

Pohlschroder, M., Gimenez, M. I. and Jarrell, K. F. (2005). "Protein transport in Archaea: Sec and twin arginine translocation pathways." *Curr. Opin. Microbiol.* **8**(6): 713-9.

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

Pradel, N., Ye, C. Y., Livrelli, V., Xu, H. G., Joly, B. and Wu, L. F. (2003). "Contribution of the twin arginine translocation system to the virulence of enterohemorrhagic Escherichia coli O157: H7." *Infect. Immun.* **71**(9): 4908-4916.

Rader, C. and Barbas, C. F. (1997). "Phage display of combinatorial antibody libraries." *Curr. Opin. Biotechnol.* **8**(4): 503-508.

Ransone, L. J., Visvader, J., Sassone-Corsi, P. and Verma, I. M. (1989). "Fos-Jun interaction: mutational analysis of the leucine zipper domain of both proteins." *Genes & Development* **3**(6): 770-81.

Raynaud, C., Guilhot, C., Rauzier, J., Bordat, Y., Pelicic, V., Manganelli, R., Smith, I., Gicquel, B. and Jackson, M. (2002). "Phospholipases C are involved in the virulence of Mycobacterium tuberculosis." *Mol. Microbiol.* **45**(1): 203-17.

Ritz, D. and Beckwith, J. (2001). "Roles of thiol-redox pathways in bacteria." *Annu. Rev. Microbiol.* **55**: 21-48.

Robinson, C. and Bolhuis, A. (2001). "Protein targeting by the twin-arginine translocation pathway." *Nat. Rev. Mol. Cell Biol.* **2**(5): 350-356.

Rodrigue, A., Chanal, A., Beck, K., Muller, M. and Wu, L. F. (1999). "Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial Tat pathway." *J. Biol. Chem.* **274**(19): 13223-13228.

Rose, R. W., Bruser, T., Kissinger, J. C. and Pohlschroder, M. (2002). "Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway." *Mol. Microbiol.* **45**(4): 943-950.

Rossier, O. and Cianciotto, N. P. (2005). "The Legionella pneumophila tatB gene facilitates secretion of phospholipase C, growth under iron-limiting conditions, and intracellular infection." *Infect. Immun.* **73**(4): 2020-2032.

Ruiz, N. and Silhavy, T. J. (2005). "Sensing external stress: watchdogs of the Escherichia coli cell envelope." *Curr. Opin. Microbiol.* **8**(2): 122-6.

Russel, M., Linderoth, N. A. and Sali, A. (1997). "Filamentous phage assembly: Variation on a protein export theme." *Gene* **192**(1): 23-32.

Sambasivarao, D., Scraba, D. G., Trieber, C. and Weiner, J. H. (1990). "Organization of dimethyl-sulfoxide reductase in the plasma membrane of Escherichia coli." *J. Bacteriol.* **172**(10): 5938-5948.

Sambasivarao, D., Turner, R. J., Simala-Grantz, J. L., Shaw, G., Hu, J. and Weiner, J. H. (2000). "Multiple roles for the twin arginine leader sequence of dimethyl sulfoxide reductase of Escherichia coli." *J. Biol. Chem.* **275**(29): 22526-22531.

Sanders, C., Wethkamp, N. and Lill, H. (2001). "Transport of cytochrome c derivatives by the bacterial Tat protein translocation system." *Mol. Microbiol.* **41**(1): 241-246.

Santini, C. L., Bernadac, A., Zhang, M., Chanal, A., Ize, B., Blanco, C. and Wu, L. F. (2001). "Translocation of jellyfish green fluorescent protein via the Tat system of Escherichia coli and change of its periplasmic localization in response to osmotic up-shock." *J. Biol. Chem.* **276**(11): 8159-8164.

Santini, C. L., Ize, B., Chanal, A., Muller, M., Giordano, G. and Wu, L. F. (1998). "A novel Sec-independent periplasmic protein translocation pathway in Escherichia coli." *EMBO J.* **17**(1): 101-112.

Sargent, F., Bogsch, E. G., Stanley, N. R., Wexler, M., Robinson, C., Berks, B. C. and Palmer, T. (1998). "Overlapping functions of components of a bacterial Sec-independent protein export pathway." *EMBO J.* **17**(13): 3640-3650.

Sargent, F., Gohlke, U., de Leeuw, E., Stanley, N. R., Palmer, T., Saibil, H. R. and Berks, B. C. (2001). "Purified components of the Escherichia coli Tat protein transport system form a double-layered ring structure." *Eur. J. Biochem.* **268**(12): 3361-3367.

Sargent, F., Stanley, N. R., Berks, B. C. and Palmer, T. (1999). "Sec-independent protein translocation in Escherichia coli - A distinct and pivotal role for the TatB protein." *J. Biol. Chem.* **274**(51): 36073-36082.

Sassetti, C. M., Boyd, D. H. and Rubin, E. J. (2003). "Genes required for mycobacterial growth defined by high density mutagenesis." *Mol. Microbiol.* **48**(1): 77-84.

Schaerlaekens, K., Van Mellaert, L., Lammertyn, E., Geukens, N. and Anne, J. (2004). "The importance of the Tat-dependent protein secretion pathway in Streptomyces as revealed by phenotypic changes in tat deletion mutants and genome analysis." *Microbiology* **150**: 21-31.

Schatz, G. and Dobberstein, B. (1996). "Common principles of protein translocation across membranes." *Science* **271**(5255): 1519-1526.

Schnell, D. J. and Blobel, G. (1993). "Identification of intermediates in the pathway of protein import into chloroplasts and their localization to envelope contact sites." *J. Cell Biol.* **120**(1): 103-115.

Settles, A. M. and Martienssen, R. (1998). "Old and new pathways of protein export in chloroplasts and bacteria." *Trends Cell Biol.* **8**(12): 494-501.

Shimohata, N., Akiyama, Y. and Ito, K. (2005). "Peculiar properties of DsbA in its export across the Escherichia coli cytoplasmic membrane." *J. Bacteriol.* **187**(12): 3997-4004.

Singh, S. K., Grass, G., Rensing, C. and Montfort, W. R. (2004). "Cuprous oxidase activity of CueO from Escherichia coli." *J. Bacteriol.* **186**(22): 7815-7.

Sone, M., Kishigami, S., Yoshihisa, T. and Ito, K. (1997). "Roles of disulfide bonds in bacterial alkaline phosphatase." *J. Biol. Chem.* **272**(10): 6174-6178.

Stanley, N. R., Findlay, K., Berks, B. C. and Palmer, T. (2001). "Escherichia coli strains blocked in Tat-dependent protein export exhibit pleiotropic defects in the cell envelope." *J. Bacteriol.* **183**(1): 139-144.

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**(16): 11591-11596.

Stanley, N. R., Sargent, F., Buchanan, G., Shi, J. R., Stewart, V., Palmer, T. and Berks, B. C. (2002). "Behaviour of topological marker proteins targeted to the Tat protein transport pathway." *Mol. Microbiol.* **43**(4): 1005-1021.

Stemmer, W. P. C. (1994). "DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution." *Proc. Natl. Acad. Sci. U. S. A.* **91**(22): 10747-10751.

Sturm, A., Schierhorn, A., Lindenstrauss, U., Lilie, H. and Bruser, T. (2006). "YcdB from Escherichia coli reveals a novel class of Tat-dependently translocated hemoproteins." *J. Biol. Chem.* **281**(20): 13972-13978.

Thomas, G., Potter, L. and Cole, J. A. (1999). "The periplasmic nitrate reductase from Escherichia coli: a heterodimeric molybdoprotein with a double-arginine signal sequence and an unusual leader peptide cleavage site." *FEMS Microbiol. Lett.* **174**(1): 167-171.

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**(1): 47-53.

Tinker, J. K., Erbe, J. L. and Holmes, R. K. (2005). "Characterization of fluorescent chimeras of cholera toxin and Escherichia coli heat-labile enterotoxins produced by use of the twin arginine translocation system." *Infect. Immun.* **73**(6): 3627-3635.

Topell, S., Hennecke, J. and Glockshuber, R. (1999). "Circularly permuted variants of the green fluorescent protein." *FEBS Lett.* **457**(2): 283-289.

Tullman-Ercek, D., DeLisa, M.P., Kawarasaki, Y., Iranpour, P., Ribnicky, B., Palmer, T., Georgiou, G. (Submitted). "Export pathway selectivity of *Escherichia coli* twin-arginine translocation signal peptides."

Turner, R. J., Papish, A. L. and Sargent, F. (2004). "Sequence analysis of bacterial redox enzyme maturation proteins (REMPs)." *Can. J. Microbiol.* **50**(4): 225-238.

Uchida, K., Mori, H. and Mizushima, S. (1995). "Stepwise movement of preproteins in the process of translocation across the cytoplasmic membrane of Escherichia coli." *J. Biol. Chem.* **270**(52): 30862-8.

Villaverde, A. and Carrio, M. M. (2003). "Protein aggregation in recombinant bacteria: biological role of inclusion bodies." *Biotechnology Letters* **25**(17): 1385-1395.

von Heijne, G. (1985). "Signal sequences. The limits of variation." *J. Mol. Biol.* **184**(1): 99-105.

Voulhoux, R., Ball, G., Ize, B., Vasil, M. L., Lazdunski, A., Wu, L. F. and Filloux, A. (2001). "Involvement of the twin-arginine translocation system in protein secretion via the type II pathway." *EMBO J.* **20**(23): 6735-6741.

Walker, K. W. and Gilbert, H. F. (1994). "Effect of redox environment on the in vitro and in vivo folding of Rtem-1 beta-lactamase and Escherichia coli alkaline phosphatase." *J. Biol. Chem.* **269**(45): 28487-28493.

Walsh, G. (2005). "Biopharmaceuticals: recent approvals and likely directions." *Trends Biotechnol.* **23**(11): 553-558.

Walsh, G. (2005). Current status of biopharmaceuticals: approved products and trends in approvals. Modern Biopharmaceuticals. Knablein, J. Weinheim, Wiley-VCH. **1:** 1-34.

Weiner, J. H., Bilous, P. T., Shaw, G. M., Lubitz, S. P., Frost, L., Thomas, G. H., Cole, J. A. and Turner, R. J. (1998). "A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins." *Cell* **93**(1): 93-101.

Wickner, W. and Schekman, R. (2005). "Protein translocation across biological membranes." *Science* **310**(5753): 1452-6.

Winter, G., Griffiths, A. D., Hawkins, R. E. and Hoogenboom, H. R. (1994). "Making antibodies by phage display technology." *Annu. Rev. Immunol.* **12**: 433-455.

Yahr, T. L. and Wickner, W. T. (2001). "Functional reconstitution of bacterial Tat translocation in vitro." *EMBO J.* **20**(10): 2472-2479.

Yamane, K. and Mizushima, S. (1988). "Introduction of basic amino acid residues after the signal peptide inhibits protein translocation across the cytoplasmic membrane of Escherichia coli. Relation to the orientation of membrane proteins." *J. Biol. Chem.* **263**(36): 19690-6.

Yen, M. R., Tseng, Y. H., Nguyen, E. H., Wu, L. F. and Saier, M. H. (2002). "Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system." *Arch. Microbiol.* **177**(6): 441-450.

You, L. and Arnold, F. H. (1996). "Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide." *Protein Eng.* **9**(1): 77-83.

Zeytun, A., Jeromin, A., Scalettar, B. A., Waldo, G. S. and Bradbury, A. R. M. (2003). "Fluorobodies combine GFP fluorescence with the binding characteristics of antibodies." *Nat. Biotechnol.* **21**(12): 1473-1479.

Vita

Danielle Tullman Ercek was born in Rome, New York on May 29, 1979, the daughter of Maureen Sharkey Tullman and Edward Julius Tullman Jr. As a child in a military family, she attended schools all around the country. She received her high school education at Green Valley High School in Henderson, Nevada, in 1997, where she graduated as valedictorian with an International Baccalaureate diploma. She then entered Illinois Institute of Technology in Chicago, Illinois, where she received the degree of Bachelor of Science in Chemical Engineering with a specialization in Biotechnology in December 2000. In January 2001 she entered the Graduate School of The University of Texas at Austin to work towards her Doctorate in Chemical Engineering. In 2004, she married James Michael Ercek, who is also a graduate of Illinois Institute of Technology. She has one sister, Jennifer Tullman, a chemical engineering graduate student at Johns Hopkins University, and one brother, Edward Tullman, a chemistry major at the University of Nevada, Las Vegas.

Permanent address: 2128 Club Meadows Drive, Henderson, Nevada 89074

This dissertation was typed by the author.