

The Tat pathway as a biotechnological tool for the expression and export of heterologous proteins in *Escherichia coli*

Approximately 30% of all currently licensed, recombinantly expressed biotherapeutic products are produced in *Escherichia coli*, among which a significant proportion are exported to the periplasm by the general Secretory 'Sec' pathway. However, this pathway cannot handle many target proteins and the Tat pathway is emerging as a powerful alternative means of export. The Tat system exports fully folded proteins, moreover it preferentially exports correctly folded proteins and has been shown to export a range of biotherapeutics. This review will discuss our current understanding of the Tat pathway and its potential application for the industrial-scale production of biotherapeutics.

Keywords: biopharmaceutical • protein transport • signal peptide • Tat • twin-arginine

Escherichia coli-based platforms for expression of recombinant proteins: an overview

54 new biologics were approved in the period between January 2010 and July 2014. Of these, 31% were mAbs and 15% were hormones, the former having an estimated sales value of \$63 billion in 2013 [1]. Furthermore, the overall market value of the biopharmaceutical industry has risen steadily over this period and has now surpassed an approximate total annual sales value of \$140 billion [1]. Although mammalian systems have taken precedence for the recombinant expression of many therapeutic proteins, *Escherichia coli* is still responsible for the production of over 30% of all biotherapeutics licensed to date and is therefore highly valued as a recombinant expression system [2]. The popularity and longevity of *E. coli* expression systems can be accredited to ease of manipulation, rapid growth and low cost [3].

Historically, several strategies have been used to produce therapeutic proteins in *E. coli* including: first, expression of soluble protein in the cytoplasm; second, expression as insoluble inclusion bodies; third, expression in the cytoplasm followed by export to the

periplasm via the general Secretory or 'Sec' pathway or fourth, secretion into the culture medium with the help of OsmY fusions [4,5]. For the latter, the only molecular requirement is the presence of an N-terminal signal peptide and this is cleaved upon entry into the periplasm by a specific signal peptidase. Plasmid vectors containing these Sec signal sequences upstream of a multiple cloning region are commercially available.

The benefits of exporting the target protein into the periplasm are twofold. First, the protein can be more easily extracted by selectively rupturing the outer membrane [6], which minimizes the release of contaminating proteins, DNA and general debris. Consequently, clarification performance is also improved [7,8]. Second, the periplasm is the only oxidizing compartment in wild-type cells and this is essential for efficient disulfide bond formation. Since many proteins and almost all antibody fragments contain disulfide bonds, this 'protein export' strategy is an important one.

Unfortunately, the Sec pathway is unable to export many heterologous proteins, particularly those that fold rapidly within the cytoplasm. This is primarily because the Sec

Kelly L Walker¹, Alexander S Jones¹ & Colin Robinson^{*1}

¹School of Biosciences, University of Kent, Canterbury, CT2 7NJ, UK

*Author for correspondence:
C.Robinson-504@kent.ac.uk

Key terms

Protein export: the process by which proteins are translocated across the bacterial plasma membrane.

Signal peptide: cleavable N-terminal peptide that is necessary and sufficient to direct the export of a protein by the Sec or Tat pathways (removed on crossing the plasma membrane).

system transports proteins through a relatively narrow pore in an unfolded state, and folded substrates are often transport-incompetent due to the lack of a powerful unfolding capability [9]. Instead, they tend to form inclusion bodies within the cytoplasm, giving rise to cellular stress and frequently to a reduction in total yield. Other potential targets cannot efficiently fold in the periplasm due to a lack of suitable chaperone molecules and protein-folding catalysts. As a direct consequence, attention has focused on the other mainstream protein export pathway in bacteria: the Tat pathway.

Essential features of the Tat pathway

In *E. coli* and many other bacteria, the Tat pathway operates in parallel with the Sec pathway and is important for many cellular processes, including respiratory and photosynthetic energy metabolism, iron and phosphate acquisition, cell division and cell motility. As with Sec substrates, Tat substrates are synthesized with N-terminal signal peptides but for Tat, these contain a number of specific determinants including the presence of the twin-arginine motif that gives the pathway its full name (twin-arginine translocation system). Importantly, the Tat system transports proteins across the plasma membrane in a fully folded state; its mechanism is thus fundamentally different from that of the Sec system (see Figure 1). Translocation is reliant on the proton motive force and the system can transport a range of heterologous proteins if a Tat-specific signal peptide is present at the N-terminus (see [10–13]) (Figure 2).

Because the Tat pathway exports folded proteins, it clearly has the potential to transport proteins that cannot be exported by the Sec pathway. Another interesting characteristic of the Tat pathway that is particularly relevant for biotechnological exploitation is its unique ‘proofreading’ capacity (discussed below). Consequently, it is likely that this system has a natural tendency to produce high quality, active products.

In comparison to the Sec pathway, the Tat pathway has far fewer natural substrates and is somewhat more specialized. In enteric bacteria such as *E. coli*, for example, there are typically 20–30 known substrates that range in size from less than 10 kDa (high-potential iron–sulfur proteins) to 150 kDa (formate dehydrogenases) [14,15]. This variability indicates that the size and shape of the substrate protein may be relatively unim-

portant, which in turn suggests that the Tat pathway has the potential to export a correspondingly wide range of biotherapeutic products, potentially including relatively complex molecules.

The Tat pathway is responsible for the export of a number of periplasmic proteins that require cofactor insertion prior to folding in the cytoplasm, and also those that assemble into oligomeric complexes in the cytoplasm [16,17]. Like Sec substrates, Tat substrates are identified by the presence of an N-terminal signal which in addition to the twin-arginine motif, has a number of other features: it has a tripartite architecture with a polar N-terminal domain, a hydrophobic core (H domain) and a polar C-terminal domain. The conserved SRRxFLK motif that harbors the twin-arginine motif is located at the junction between the N and H domains. The presence of both arginine residues is important for efficient export, but in bacteria a single arginine is sufficient for export to take place albeit with reduced efficiency [18–20]. A notable difference between Sec and Tat signal peptides is the presence of basic residues in the C-terminal domain prior to the cleavage site, and there is some evidence that they carry out a ‘Sec avoidance motif’ that prevents mistargeting by, or perhaps futile interactions with, the Sec pathway [21,22]. Table 1 shows a number of Tat signal peptides, with the critical twin-arginine motif highlighted [23].

Tat components & complexes

Current models of the Tat system have the translocase consisting of two core complexes – a membrane-embedded substrate-docking complex and a pore-forming complex that likely assembles once substrate is docked. Depending on the organism, these core complexes are composed from a selection of components [24]. For example in the model Gram-negative bacterium, *E. coli*, the Tat system is composed of three components – the 9.6 kDa TatA, the 18.5 kDa TatB and the 28.9 kDa TatC proteins [25]. The genes encoding these proteins are all found in a single operon and are expressed constitutively. In the Gram-negative Tat system, TatB and TatC form the substrate-docking complex in the membrane where, upon substrate docking, it is thought that the homo-oligomeric TatA complex coalesces to form the pore, although the translocation mechanism is not fully understood [26–28]. In the TatABC system, all components are necessary for translocation to occur. However, *tatA* null mutants do have functioning Tat systems due to the presence of a TatA paralogue, TatE, located elsewhere in the genome. TatA and TatE share 53% sequence identity [29] with TatE thought to have arisen from a gene duplication event. Indeed TatE can functionally replace TatA; however, there is no evidence that TatE provides any additional role [29–31].

Bacillus subtilis is an industrially relevant Gram-positive bacterium that provides an example of twin-arginine translocases that differ from the ABC system [16]. Here, the translocation apparatus is composed of just two proteins – TatA-like and TatC-like – and has far fewer native substrates. Interestingly *B. subtilis* contains two discrete Tat systems that function in parallel (similar to many Gram-positive bacteria) with different substrate specificities [32,33]. The first is TatAyCy which, under all observations, is constitutively expressed and carries out the translocation of the Dyp-type proteins (e.g., YwbN peroxidase), the alkaline phosphatase YkuE and the Rieske Fe–S protein QcrA [33,34]. The other is TatAdCd produced from the *tatAdCd* operon that forms part of the *phoD* gene cluster and is expressed only under phosphate-limiting conditions [35]. It is unsurprising then that TatAdCd has only one known native substrate, PhoD. What is surprising, however, is the ability of either TatA-like protein to complement both *tatA/E* and *tatB* null mutants, indicating that these TatA-like proteins in Gram-positive bacteria are bifunctional [35]. As a consequence of this reduction in necessary components, these TatAC systems have been dubbed ‘minimal Tat systems’ and possibly offer a more simplistic approach to studying the Tat apparatus.

The Tat pathway's unique proofreading ability

While the Tat system's ability to export prefolded heterologous proteins has attracted considerable interest, the system has another trait that is equally important: it is not only capable of exporting folded proteins, it preferentially exports correctly folded proteins [36,37]. There is mounting evidence to suggest the Tat system operates with a complex proofreading and a quality control mechanism (here, proofreading refers to the overall-folded state of the protein, while quality control refers to the correct incorporation of cofactors). This characteristic of Tat has major implications for the biopharmaceutical industry because it has the potential to ensure the exclusive export of high-quality products that have high specific activities and minimal heterogeneity. The proofreading capacity of Tat is important for the normal functioning of the organism: the majority of its native substrates must fold, by necessity, in the cytoplasm prior to translocation – either because of the insertion of cofactors by accessory proteins not present in the periplasm, or a high-speed folding that precludes Sec-dependent export. As such the proofreading and/or quality control capabilities of Tat ensure there is no premature export of apoproteins.

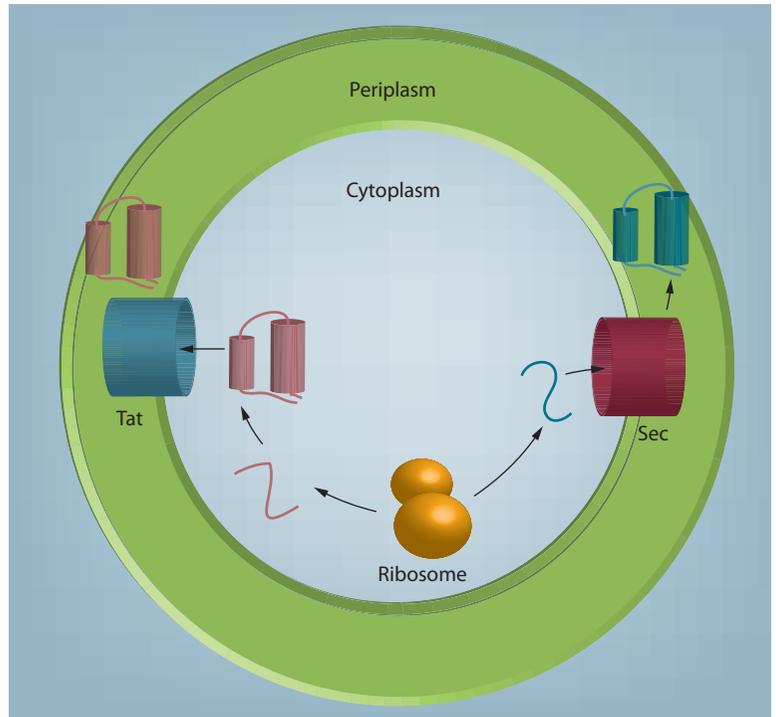


Figure 1. Protein export by the Tat and Sec *Escherichia coli* translocation pathways. Tat substrates are synthesized with an N-terminal signal peptide, after which they fold and are exported into the periplasm by the Tat translocon. The signal peptide is removed after entry into the periplasm. Proteins with a Sec signal peptide are synthesized with Sec-specific signal peptides but they are transported through the Sec translocon in an unfolded state. For Sec substrates, folding occurs after entry into the periplasm.

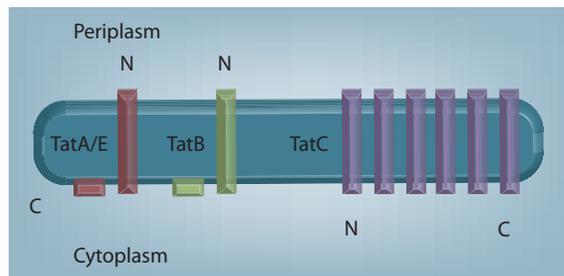


Figure 2. Organization of the Tat subunits in the bacterial plasma membrane. TatA and TatB are single-span membrane proteins with short periplasmic N-terminal regions and longer cytoplasmic C-terminal domains. They also contain short amphipathic regions that are believed to interact with the membrane bilayer. TatA and TatB are related but carry out different roles, but *E. coli* also contains a TatA paralogue, TatE, that is present in low quantities. TatE appears not to play a major role in the translocation process. TatC is a more complex, six-span protein that is found closely associated with TatB, while TatA is found as a separate homo-oligomeric complex at steady state. During translocation, the TatBC and TatA complexes coalesce to form the full translocon.

Table 1. Natural Tat signal peptides.

Signal peptide	Sequence
FdnG	MDV SRR QFFKICAGGMAGTTVAALGFAPKQALAQA
FdoG	MQV SRR QFFKICAGGMAGTTAAALGFAPSVALA
NapG	MSRS AKPQNGRRR FLRDVVRTAGGLAAVGVALGLQQQTARA
NrfC	MTW SRR QFLTGVGVLAAVSGTAGRVVA
HyaA	MNNEETFYQ AMRR QGVTRRSFLKYCSLAATSLGLGAGMAPKIAWA
YnfE	MSKNERMVGIS RR TLVKSTAIGSLALAAGGFSLPFTLRNAAA
WcaM	MPFK LSRR TFLTASSALAFHTPFARA
TorA	MNNNDLFQ SRRR FLAQLGGLTVAGMLGPSLLTPRRATA
NapA	M KL SRRSFMKANAVAAAAAAGLSVPGVARA
YagT	MSNQGEYPEDNRV GKHEPH DLSL RR DLIKVSAATAATA
YcbK	MDKFDAN RR KLLALGGVALGAAILPTPAFA

List of *Escherichia coli* Tat signal peptides from [23]. The signal peptides vary in length and in sequence apart from the conserved twin-arginine motif (in bold). Most of the peptides terminate with an Ala-Xaa-Ala consensus motif that specifies cleavage by signal peptidase following translocation.

Multiple reports abound of the Tat systems's ability to completely reject malformed substrates [11,36], in some cases appearing to target them for degradation [34,37]. In most cases, protein folding has been disrupted by inhibiting the formation of disulfide bonds, by expression of heterologous proteins in the reducing cytoplasm of *E. coli* or through the mutation of bridge-forming Cysteine residues [12,37], both scenarios resulting in the abolishment of export via the Tat apparatus. Some studies have sought to address the quality control mechanism of Tat through disruption of cofactor binding prior to export; however, these studies failed to establish the folded state of these cofactor-less species and thus it cannot be determined if these proteins were rejected due to substantial misfolding or a lack of cofactor. One study did find evidence to support a quality control/proofreading hierarchy; however – a Rieske Fe–S protein, QcrA that forms part of the cytochrome *bc₁* complex in Gram-positive bacteria, was mutated such that Fe–S binding was inhibited or a disulfide bridge was broken [34]. Cofactor-binding-deficient protein was observed associated with the Tat apparatus in the membrane but not in the periplasm, indicating translocation had been halted at a late stage, whereas protein lacking the disulfide bridge was degraded in the cytoplasm [34]. Some studies have sought to simplify the Tat proofreading process, indeed Richter *et al.* [38] showed that FG repeats taken from the small, hydrophilic yeast nuclear pore protein Nsp1p which lack secondary structure are exported via Tat, albeit with decreasing efficiency as size increased (constructs of 20–30 kDa were tested), and is rejected upon insertion of a hydrophobic patch (of at least five consecutive residues) from PhoA. Similarly, a recent

study found that IFN $\alpha 2\beta$ and hGH were shown to be exported without prior formation of their native disulfide bonds [39]. Taken together, these results seem to suggest a much more simplistic hydrophobicity sensor employed by the Tat apparatus rather than assessment of the folded state of substrates.

Progress is further underway to identify the proofreading and quality control mechanisms of Tat with the recent identification of residues in the apparatus itself important in the precision of proofreading capabilities. Dual mutations in either *tatA* or *tatC* (e.g., P97S and A133T in *tatC*) have been isolated that enable the export of previously export-incompatible, aggregation-prone, molten-globule state proteins while still maintaining the ability to reject proteins with Tat-incompatible signal peptides [40]. Uncovering the mechanism through which Tat carries out proofreading and/or quality control is beneficial to the biopharmaceutical industry as manipulation of these qualities in Tat can both simplify the apparatus involved and enable fine-tuned control over the quality of product produced.

Transport of heterologous proteins – including disulfide-bonded proteins – by the Tat pathway

A growing number of heterologous proteins have been shown to be exported by the Tat pathway and several are listed in Table 2, which excludes proteins that have been exported as fusion proteins [12,38,40–41]. The first demonstration of heterologous protein export [13] involved expression of GFP with an N-terminal TorA signal peptide. It was shown that the GFP could be efficiently exported to the periplasm and the protein

was shown to be fluorescent and thus active. In general, GFP cannot be exported by the Sec pathway, probably because it folds too rapidly, and these initial studies thus served as proof-of-principle that Tat can export 'Sec-incompatible' proteins.

Table 2 also shows that a number of disulfide-bonded proteins have been exported by the Tat pathway, including hGH, single chain variable fragments (scFvs) and interferons. Here, the story becomes more complex. In *E. coli*, disulfide bond formation occurs exclusively in the periplasm; the cytoplasm is more reducing because it lacks disulfide bond catalysts and also harbors a number of pathways that directly result in the reduction of disulfide bonds. This has led to the assumption that disulfide bond-containing proteins cannot be produced in the cytoplasm of the majority of prokaryotes including the industrial workhorse, *E. coli*.

Attempts to make the cytoplasm of bacteria less reducing and therefore more suitable for disulfide bond formation have met with mixed success [43–47]. Some strategies have favored disruption of either or both of the *gor* and *trxB* reducing pathways [12]. The disadvantage of this, particularly for the strain lacking both pathways, is that cells show significant growth defects, probably due to the requirement of the reducing pathways for other cellular processes [44]. This means that these strains are probably not suitable for large-scale industrial production of biopharmaceuticals, and to our knowledge they have yet to be used in this capacity. Despite these issues, the combination of *trxB* and *gor* knockouts with a mutation of *aphC* did give rise to a significant improvement in the production of some complex disulfide-bonded proteins including a truncated variant of vtPA [45–48]. An additional 20-fold improvement was seen by addition of DsbC, a disulfide isomerase [46,49].

CyDisCo strains (cytoplasmic disulfide bond formation in *E. coli*) are a recently engineered solution to this problem [50,51]. Instead of inactivating the reducing pathways of the cytoplasm, these strains express Erv1p and PDI, which is capable of acting as both an isomerase and also a protein-folding chaperone. Coexpression of a target protein with Erv1p and PDI facilitates efficient cytoplasmic disulfide bond formation, reduces inclusion body formation and can give rise to much higher yields of soluble product. For PhoA and AppA (which each have four disulfide bonds), a 1000-fold increase in yield was seen relative to a wild-type *E. coli* control (at shake flask scale; [51]). Additionally, for vtPA which has nine disulfide bonds, this system gave production yields 800-fold higher than those previously described for the $\Delta\text{gor } \Delta\text{trxB}$ strains with coexpression of DsbC [51].

The partnership of CyDisCo and the Tat export pathway has significant potential for the production of disulfide bond-containing biotherapeutics. Tests on two model disulfide bond-containing proteins (PhoA and AppA, described above) showed that they could not be exported by Tat in wild-type cells; apparently the Tat pathway rejects them because they are incorrectly folded [12,42]. However, these proteins were efficiently exported in CyDisCo strains [42] in keeping with Tat's preference for export of correctly folded proteins. The combination of CyDisCo-mediated disulfide bond formation and export via Tat thus offers a means of exporting a range of disulfide-bonded proteins and it will be interesting to test such strains with more complex molecules.

The combination of Tat and CyDisCo has also enabled efficient export of an anti-interleukin 1 β scFv. Following translocation, MS confirmed that the scFv contained the correct disulfide bonds and was also correctly processed [42]. Similar results were obtained for material produced in ΔdsbB cells (which cannot perform periplasmic disulfide bond formation) confirming the requirement of the CyDisCo system for disulfide bond formation. This work thus confirmed that the combination of the Tat and CyDisCo systems can be used to successfully produce and export biopharmaceutically relevant, active material.

For some proteins, however, prior disulfide bond formation is not essential for export via the Tat pathway; interferon $\alpha 2b$, hGH and two antibody fragments were all found to efficiently exported to the periplasm in both wild-type and CyDisCo strains [39]. For these substrates, disulfide bonds were acquired following export to the periplasm, probably catalyzed by the usual DsbABCD system [39]. A possible explanation for the export of these proteins in the reduced state is that they adopt near-native tertiary structures in the absence of disulfide bond formation, and thus 'fool' the Tat machinery into permitting export. hGH, for example, is known to be biologically active even after cysteine modifications that block disulfide bond formation [52,53]. In summary, it appears that Tat can export a significant number of disulfide bond containing proteins and the logical approach is to test for export of novel constructs in both wild-type and CyDisCo strains.

Tat-dependent export of heterologous proteins in bioreactors

The main benefits for the biopharmaceutical industry of exploiting the Tat pathway are: first, its natural proofreading ability which provides a quality control function; second, export of heterologous protein to the periplasm allows for simpler downstream processing

Table 2. Heterologous proteins exported by Tat.

Heterologous Tat substrate	Mw (kDa)	No. of disulfide bonds	Require CyDisCo for export?	Ref.
hGH	23	2	No	[42]
IFN $\alpha 2\beta$	21	2	No	[39]
VH domain construct	14	1	No	[39]
scFv against β -galactosidase	29	1	No	[39]
GFP	27	0	No	[13]
PhoA	47	4	Yes	[42]
AppA	50	4	Yes	[42]
scFv against interleukin-1 β	28	2	Yes	[42]
Top7	11	0	N/A	[40]
MBP/MaIE	43	0	N/A	[22]

The table shows a number of proteins that have been shown to be exported by the Tat pathway if a Tat signal peptide (usually from TorA) is attached. The M_w and number of disulfide bonds are indicated for each protein. For disulfide bond-containing proteins, the dependence on CyDisCo components for export is also shown; CyDisCo strains enable the generation of fully folded, disulfide-bonded proteins in the cytoplasm and several examples have been shown to be exported by Tat if a signal peptide is present (see text).

and finally, third, the Tat pathway may facilitate the export of proteins that are currently incompatible with the ubiquitous Sec pathway. The studies described in the above sections have shown that it can export a range of heterologous proteins, but these studies were mainly conducted at shake-flask scale and the overall export yields were accordingly low. However, a few tests have been carried out using fermentation systems and the results, while preliminary, are encouraging.

In the first of these studies [13], it was shown that GFP could be exported to the *E. coli* periplasm in wild-type cells with reasonable efficiency [54], and greater yields were obtained when the TatABC proteins were overexpressed. Presumably, the increased export capacity led to more efficient export. A yield of active GFP equal to 1.1 g/l was obtained [54]. This is comparable to yields obtained in industrial applications based on export by the Sec pathway [55], and this study thus provided evidence that the Tat system has potential for export of heterologous protein at industrially relevant levels.

Interestingly, overexpression of TatABC seems to have additional benefits in that it also increases overall growth rate and has a minimal effect on cell integrity or downstream processing [56]. For example, during the overexpression of a natural Tat substrate (FhuD), overexpression of TatABC improved growth rate by 25%, prevented saturation of the translocation machinery and consequently gave a 40% increase in periplasmic accumulation relative to wild-type Tat expression. Overexpression of TatABC also protects the integrity of the cell and at cell harvest, there was threefold less protein in the media than in cultures of wild-type cells. This is evidence of the detrimental effect of saturation

of the Tat machinery, which presumably results in secretion stress. Overexpression of the Tat machinery did not compromise downstream processing performance and therefore represents a useful and logical strategy for high-level export of commercially useful proteins.

A variant of the Tat export system involving expression of a *B. subtilis* Tat system in *E. coli*

An interesting variant of the Tat-dependent export strategy has also been described, involving the expression of a *B. subtilis* Tat system in an *E. coli* *tat* null mutant. One of the two *B. subtilis* TatAC-type systems, termed TatAdCd, is active in *E. coli* and able to partially complement an *E. coli* *tat* null strain [24,35,57]. The system was able to export a signal peptide-GFP construct as efficiently as the native TatABC system [58]. However, the bulk of the GFP was found to be released into the media during batch fermentation and approximately 80% of the mature protein is found there by cell harvest. This was equivalent to 0.35 g/l GFP [58]. The explanation for this 'periplasmic release' seems to be that the outer membrane of the background *tat* null mutant strain is inherently leaky, presumably due to the absence of one or more Tat-exported proteins, and expression of TatAdCd yields only a partial complementation; the system is able to export proteins bearing some *E. coli* signal peptides very effectively (including the TorA signal peptide), but fails to recognize others. The net result is that heterologous proteins can be exported to the periplasm, but the leaky outer membrane defect is not complemented by TatAdCd and the exported protein is released into the medium.

The spheroplast remains intact and there is little release of the cytoplasmic contents. Thus, the use of this TatAdCd-based system during fermentation offers the potential to harvest product direct from medium with potential advantages in terms of ease of purification and simplified DSP [58]. However, further work is required to characterize this strain and determine its suitability for fermentation systems.

Tat & glycosylation

Although less costly than other expression systems, a major limitation of *E. coli* for the expression of heterologous proteins is that they are unable to perform glycosylation. Asparagine-linked (N-linked) glycosylation is the most prevalent posttranslational modification [59,60] and many biotherapeutics require some form of glycosylation for biomedical applications. The correct attachment of glycan(s) to a eukaryotic protein can significantly affect a number of its properties including protein folding, stability and solubility. As a consequence, Chinese hamster ovary (CHO) cells and yeasts are usually used for the production of such proteins.

Although rarely seen in prokaryotes, the last decade has witnessed the identification and study of protein glycosylation systems in the enteropathogenic bacterium *Campylobacter jejuni*. This system, comprising of a number of *pgl* genes, has sequence similarity to eukaryotic glycosyltransferases but leads to the addition of a structurally dissimilar N-glycans [61,62]. Furthermore, in the *C. jejuni* system, glycosylation occurs independently of the protein translocation machinery [63]. Nevertheless, the expression of the *C. jejuni pgl* locus was sufficient for the glycosylation of recombinant glycoproteins in *E. coli* [61], raising the possibility that this system can be used for the glycosylation of at least some biotherapeutics. More recently, groups have worked to improve the efficiency of this process using *pgl* genes that have been codon optimized for *E. coli* expression [64].

Pgl-driven glycosylation in *E. coli* is becoming increasingly well characterized with the endogenous *C. jejuni* periplasmic glycoprotein, AcrA, being perhaps the most studied substrate [61,63]. It is also compatible with both Sec and Tat export machineries indicating that unlike in mammalian systems, glycosylation can occur post-translocation and for TorA-GFPmut2 (a variant of GFP) glycosylation had no effect on activity [42,62,64]. Ultimately, this implies that folded proteins can be exported by the Tat-pathway into the periplasm and glycosylated, and it will be interesting to determine whether such strategies can result in commercially attractive glycosylation platforms.

Conclusions

The Tat pathway has many unique properties that render it attractive for the production of heterologous proteins in *E. coli*. It is clearly able to export a range of globular proteins and has already been shown to export a series of biotherapeutics including antibody fragments, hGH and cytokines. Perhaps most importantly, it has an intrinsic proofreading system that enables it to preferentially export correctly folded protein to the periplasm.

Large-scale studies of Tat-dependent export have started to demonstrate its suitability for industrial application, with a model protein exported at industrially relevant levels. Encouragingly, the combination of cytoplasmic disulfide bond formation by CyDisCo, and Tat-dependent export of disulfide bond-containing proteins, offers clear potential for the industrial manufacture of the numerous biotherapeutics that contain disulfide bonds. This combination should prove attractive for the production of proteins with multiple disulfide bonds that are difficult to express in the *E. coli* cytoplasm. The development of Tat-based systems has thus reached an exciting point and it will be interesting to determine whether these systems can offer solutions to long-standing problems in the biotechnology industry.

Future perspective

Many of the current *E. coli*-based platforms are based on technology that is decades old. Plasmids, promoters and protein export strategies have been improved over the years but the general sense is of a technology that is mature. Yet, there is a clear need for new bacterial systems that can offer rapid, cost-effective solutions for the production of high-value proteins. The Tat system, especially in combination with cytoplasmic disulfide-bonding systems, may offer a solution to some of these problems and time will tell whether it enters mainstream use or is simply used as a niche technique for a subset of intransigent proteins.

Financial & competing interests disclosure

KL Walker is supported by funding from the Biotechnology and Biological Sciences Research Council 'Bioprocessing Research Industry Collaboration' initiative. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Executive summary

Escherichia coli-based platforms for expression of recombinant proteins: an overview

- Introduces the use of *E. coli*-based platforms for production of biopharmaceuticals.
- Emphasizes the historic use of the Sec pathway for exporting target proteins to the periplasm.

Essential features of the Tat pathway

- Transports large fully folded proteins across the plasma membrane.
- Radically different to the Sec pathway.
- Recognizes signal peptides bearing critical twin-arginine motif.

Tat components & complexes

- The Tat pathway is encoded by TatABC genes in Gram-negative bacteria.
- TatBC and TatA complexes coalesce to form a translocase capable of transporting large folded proteins.

The Tat pathway's unique proofreading ability

- The Tat system preferentially transports correctly folded proteins – even heterologous proteins.
- It thus offers a means of producing high-quality target protein with minimal heterogeneity.

Transport of heterologous proteins – including disulfide-bonded proteins – by the Tat pathway

- Numerous heterologous proteins can be exported by the Tat pathway.
- Disulfide-bonded proteins can be exported if the disulfide bonds can form in the cytoplasm – CyDisCo strains offer a powerful capability.

Tat-dependent export of heterologous proteins in bioreactors

- Very few pilot-scale or fermentation studies have been reported involving Tat-dependent export.
- Initial reports show that test proteins can be exported with high efficiency, with the potential for export fluxes comparable to those obtained using the Sec system.

Tat & glycosylation

- Glycosylation of proteins is now possible in *E. coli*, although the technology is at an early stage.
- It may be possible to combine this technology to glycosylate proteins that are exported by the Tat pathway.

References

- Walsh G. Biopharmaceutical benchmarks. *Nat. Biotechnol.* 32(10), 992–1000 (2014).
- Walsh G. Post-translational modifications of therapeutic proteins. *Drug Discov. Today* 15, 773–780 (2010).
- Walsh G. Biopharmaceutical benchmarks. *Nat. Biotechnol.* 24(7), 769–776 (2006).
- Georgiou G, Segatori L. Preparative expression of Secreted proteins in bacteria: Status report and future prospects. *Curr. Opin. Biotechnol.* 16, 538–545 (2005).
- Kotzsch A, Vernet E, Hammarström M *et al.* A secretory system for bacterial production of high-profile targets. *Protein Sci.* 20(3), 597–609 (2011).
- Pierce JJ, Turner C, Keshavarz-Moore E, Dunnill P. Factors determining more efficient large-scale release of a periplasmic enzyme from *E. coli* using lysozyme. *J. Biotech.* 58, 1–11 (1998).
- Balasundaram B, Harrison S, Bracewell DG. Advances in product release strategies and impact on bioprocess design. *Trends Biotechnol.* 27, 477–485 (2009).
- Harrison JS, Keshavarz-Moore E. Production of antibody fragments in *E. coli*. *Ann. NY Acad. Sci.* 782, 143–158 (1996).
- Saraogi I, Shan S. Co-translation protein targeting to the bacterial membrane. *Biochim. Biophys. Acta* 1843(8), 1433–1441 (2014).
- Müller M, Klösigen RB. The Tat pathway in bacteria and chloroplasts. *Mol. Membr. Biol.* 22, 113–121 (2005).
- Robinson C, Matos CFRO, Beck D *et al.* Transport and proofreading of proteins by the twin-arginine translocation (Tat) system in bacteria. *Biochim. Biophys. Acta* 1808, 876–884 (2011).
- DeLisa MP, Tullman D, Georgiou G. Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *Proc. Natl Acad. Sci. USA* 100(10), 6115–6120 (2003).
- Thomas JD, Daniel RA, Errington J, Robinson C. Export of active green fluorescent protein to the periplasm by the twin arginine translocase (Tat) pathway in *Escherichia coli*. *Mol. Microbiol.* 39, 47–53 (2001).
- Brüser T, Yano T, Brune DC, Daldal F. Membrane targeting of a folded and cofactor-containing protein. *Euro J. Biochem.* 270, 1211–1221 (2003).
- Jack RL, Buchanan G, Dubini A, Hatzixanthos K, Palmer T, Sargent F. Coordinating assembly and export of complex bacterial proteins. *EMBO* 23, 3962–3972 (2004).
- Monteferrante CG, Miethke M, van der Ploeg R, Glasner C, van Dijl JM. Specific targeting of the metallophosphoesterase YkuE to the bacillus cell wall requires the twin-arginine translocation system. *J. Biol. Chem.* 287, 29789–29800 (2012).
- Tottey S, Waldron KJ, Firbank SJ *et al.* Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. *Nature* 455(7216), 1138–1142 (2008).
- Buchanan G, Sargent F, Berks BC, Palmer T. 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 (2001).

- 19 DeLisa MP, Samuelson P, Palmer T, Georgiou G. Genetic analysis of the twin arginine translocator Secretion pathway in bacteria. *J. Biol. Chem.* 277(3), 29825–29831 (2002).
- 20 Mendel S, McCarthy A, Barnett JP *et al.* The *Escherichia coli* TatABC system and a *Bacillus subtilis* TatAC-type system recognise three distinct targeting determinants in twin-arginine signal peptides. *J. Mol. Biol.* 375(3), 661–672 (2008).
- 21 Bogsch E, Brink S, Robinson C. 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* 16(13), 3851–3859 (1997).
- 22 Blaudeck N, Kreutzenbeck P, Freudl R, Sprenger GA. Genetic analysis of pathway specificity during posttranslational protein translocation across the *Escherichia coli* plasma membrane. *J. Bacteriol.* 185(9), 2811–2819 (2003).
- 23 Tullman-Ercek D, DeLisa MP, Kawarasaki Y *et al.* Export pathway selectivity of *Escherichia coli* twin arginine translocation signal peptides. *J. Biol. Chem.* 282(11), 8309–8316 (2007).
- 24 Bogsch EG, Sargent F, Stanley NR, Berks BC, Robinson C, Palmer T. An essential component of a novel bacterial protein export system with homologues in plastids and mitochondria. *J. Biol. Chem.* 273, 18003–18006 (1998).
- 25 Patel R, Smith SM, Robinson C. Protein transport by the bacterial Tat pathway. *Biochim. Biophys. Acta* 1843, 1620–1628 (2014).
- 26 Alami M, Luke I, Deitermann S *et al.* Differential interactions between a twin-arginine signal peptide and its translocase in *Escherichia coli*. *Mol. Cell* 12, 937–946 (2003).
- 27 Bolhuis A, Mathers JE, Thomas JD, Barrett CML, Robinson C. TatB and TatC form a functional and structural unit of the twin-arginine translocase from *Escherichia coli*. *J. Biol. Chem.* 276, 20213–20219 (2001).
- 28 Mori H, Cline K. A twin arginine signal peptide and the pH gradient trigger reversible assembly of the thylakoid {delta} pH/Tat translocase. *J. Cell Biol.* 157, 205–210 (2002).
- 29 Sargent F, Bogsch EG, Stanley NR *et al.* Overlapping functions of components of a bacterial Sec-independent protein export pathway. *EMBO* 17, 3640–3650 (1998).
- 30 Yen MR, Tseng YH, Nguyen EH, Wu LF, Saier MH Jr. Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system. *Arch. Microbiol.* 177(6), 441–450 (2002).
- 31 Sargent F, Stanley NR, Berks BC, Palmer T. Sec-independent protein translocation in *Escherichia coli*. A distinct and pivotal role for the TatB protein. *J. Biol. Chem.* 274(51), 36073–36082 (1998).
- 32 Jongbloed JD, van der Ploeg R, van Dijl JM. Biofunctional TatA subunits in minimal Tat protein translocases. *Trends Microbiol.* 14(1), 2–4 (2006).
- 33 Jongbloed JD, Grieger U, Antelmann H *et al.* Two minimal Tat translocases in *Bacillus*. *Mol. Microbiol.* 54(5), 1319–1325 (2004).
- 34 Goosens VJ, Montefferrante CG, Maarten van Dijl J. Co-factor insertion and disulphide bond requirements for twin-arginine translocase-dependant export of the *Bacillus subtilis* rieske protein QcrA. *J. Biol. Chem.* 282(19), 13124–13131 (2014).
- 35 Barnett JP, Eijlander RT, Kuipers OP, Robinson C. A minimal Tat system from a gram-positive organism: a bifunctional TatA subunit participates in discrete TatAC and TatA complexes. *J. Biol. Chem.* 283(5), 2534–2542 (2008).
- 36 Maurer C, Panahandeh S, Moser M, Müller M. Impairment of twin-arginine-dependent export by seemingly small alterations of substrate conformation. *FEBS Lett.* 583(17), 2849–2853 (2009).
- 37 Matos CFRO, Robinson C, Di Cola AA The Tat system proofreads FeS protein substrates and directly initiates the disposal of rejected molecules. *EMBO* 27, 2055–2063 (2008).
- 38 Richter S, Lindenstrauss U, Lücke C, Bayliss R, Brüser T. Functional Tat transport of unstructured, small, hydrophilic proteins. *J. Biol. Chem.* 282, 33257–33264 (2007).
- 39 Alanen HI, Walker KL, Velez Suberbie LM *et al.* Efficient export of human growth hormone, interferon $\alpha 2\beta$ and antibody fragments to the periplasm by the *Escherichia coli* Tat pathway in the absence of prior disulfide bond formation. *Biochim. Biophys. Acta* 1853, 756–763 (2015).
- 40 Rocco MA, Waraho-Zhmeyev D, DeLisa MP. Twin-arginine translocase mutations that suppress folding quality control and permit export of misfolded substrate proteins. *Proc. Natl Acad. Sci. USA* 109(33), 13392–13397 (2012).
- 41 Lim HK, Mansell TJ, Linderman SW, Fisher AC, Dyson MR, DeLisa MP. Mining mammalian genomes for folding competent proteins using Tat-dependent genetic selection in *Escherichia coli*. *Prot. Sci.* 18(12), 2537–2549 (2009).
- 42 Matos CFRO, Robinson C, Alanen HI *et al.* Efficient export of prefolded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in *Escherichia coli* CyDisCo strains. *Biotechnol. Progress* 30(2), 281–290 (2014).
- 43 De Marco A. Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. *Microb. Cell Fact.* 8, 26 (2009).
- 44 Prinz WA, Åslund F, Holmgren A, Beckwith J. The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. *J. Biol. Chem.* 272, 15661–15667 (1997).
- 45 Stewart EJ, Åslund F, Beckwith J. Disulfide bond formation in the *Escherichia coli* cytoplasm: an *in vivo* role reversal for the thioredoxins. *EMBO* 17, 5543–5550 (1998).
- 46 Bessette PH, Åslund F, Beckwith J, Georgiou G. Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proc. Natl Acad. Sci. USA* 96, 13703–13708 (1996).
- 47 Ritz D, Lim J, Reynolds CM, Poole LB, Beckwith J. Conversion of a peroxiredoxin into a disulfide reductase by a triplet repeat expansion. *Science* 294, 158–160 (2001).
- 48 Obukowicz MG, Gustafson ME, Junger KD *et al.* Secretion of active kringle-2-serine protease in *Escherichia coli*. *Biochemistry* 29, 9737–9745 (1990).

- 49 Rietsch A, Belin D, Martin N, Beckwith J. An *in vivo* pathway for disulfide bond isomerization in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* 93, 13048–13053 (1996).
- 50 Hatahet F, Nguyen VD, Salo KEH, Ruddock LW. Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm *E. coli*. *Microbial Cell Fact.* 9, 67 (2010).
- 51 Nguyen VD, Hatahet F, Salo KEH, Enlund E, Zhang C, Ruddock LW. Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing protein expressed in the cytoplasm of *E. coli*. *Microbial Cell Fact.* 10, 1 (2011).
- 52 Bewley TA, Brovetto-Cruz J, Li CH. Human pituitary growth hormone. Physicochemical investigations of the native and reduced-alkylated protein. *Biochemistry* 8, 4701–4708 (1996).
- 53 Youngman KM, Spencer DB, Brems DN, DeFillipis MR. Kinetic analysis of the folding of human growth hormone. *J. Biol. Chem.* 270, 19816–19822 (1995).
- 54 Matos CFRO, Branston SD, Albinia A *et al.* High yield export of a native heterologous protein in the periplasm by the Tat translocation pathway in *Escherichia coli*. *Biotech. Bioeng.* 109(10), 2533–2542 (2012).
- 55 Chen C, Snedecor B, Nishihara JC *et al.* High-level accumulation of a recombinant antibody fragment in the periplasm of *Escherichia coli* requires a triple-mutant (degP prc spr) host strain. *Biotech. Bioeng.* 85, 463–474 (2004).
- 56 Branston SD, Matos CFRO, Freedman RB, Robinson C, Keshavarz-Moore E. Investigation of the impact of Tat export pathway enhancement on *E. coli* culture, protein production and early stage recovery. *Biotech. Bioeng.* 109(4), 983–991 (2012).
- 57 Barnett JP, van der Ploeg R, Eijlander RT *et al.* The twin-arginine translocation Tat systems from *Bacillus subtilis* display a conserved mode of complex organization and similar substrate recognition requirements. *FEBS J.* 276, 232–243 (2009).
- 58 Albinia AM, Matos CFRO, Branston SD, Freedman RB, Keshavarz-Moore E, Robinson C. High-level secretion of recombinant protein to the culture medium with a *Bacillus subtilis* twin-arginine translocation system in *Escherichia coli*. *FEBS J.* 280, 3810–3821 (2013).
- 59 Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* 1473, 4–8 (1999).
- 60 Wacker M, Linton D, Hitchen PG *et al.* N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* 298, 1790–1793 (2002).
- 61 Young N, Brisson JR, Kelly J *et al.* Structure of the N-linked glycan present on multiple glycoproteins in the gram-negative bacterium *Campylobacter jejuni*. *J. Biol. Chem.* 277(45), 42530–42539 (2002).
- 62 Kowarik M, Numao S, Feldman MF *et al.* N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science* 314, 148–150 (2006).
- 63 Pandhal J, Desai P, Walpole C *et al.* Systematic metabolic engineering for improvement of glycosylation efficiency in *Escherichia coli*. *Biochem. Biophys. Res. Comm.* 419(3), 40–43 (2012).
- 64 Fisher AC, Haitjema CH, Guarino C *et al.* Production of secretory and extracellular N-linked glycoproteins in *Escherichia coli*. *Appl. Environ. Microbiol.* 77(3), 871–888 (2011).