Citation for published version


DOI

https://doi.org/10.1128/ecosalplus.ESP-0040-2018

Link to record in KAR

https://kar.kent.ac.uk/74552/

Document Version

Author's Accepted Manuscript

Copyright & reuse
Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research
The version in the Kent Academic Repository may differ from the final published version. Users are advised to check http://kar.kent.ac.uk for the status of the paper. Users should always cite the published version of record.

Enquiries
For any further enquiries regarding the licence status of this document, please contact: researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at http://kar.kent.ac.uk/contact.html
The Twin-arginine pathway for protein secretion

Kelly M. Frain¹, Jan Maarten van Dijl²* and Colin Robinson¹

¹The School of Biosciences, University of Kent, Canterbury, United Kingdom
²University of Groningen, University Medical Center Groningen, Department of Medical Microbiology, Groningen, The Netherlands

*Correspondence to: J.M. van Dijl, University of Groningen (UMCG), Department of Medical Microbiology, Hanzeplein 1, P.O. Box 30001, 9700 RB Groningen, the Netherlands. E-mail: j.m.van.dijl01@umcg.nl

Running title: The bacterial Tat pathway
Summary

The Tat pathway for protein translocation across bacterial membranes stands out for its selective handling of fully folded cargo proteins. In this review, we provide a comprehensive summary of our current understanding of the different known Tat components, their assembly into different complexes and their specific roles in the protein translocation process. In particular, this overview focuses on the Gram-negative bacterium *Escherichia coli* and the Gram-positive bacterium *Bacillus subtilis*. Using these organisms, we discuss structural features of Tat complexes alongside mechanistic models that allow for the Tat pathway’s unique protein proofreading and transport capabilities. Finally, we highlight recent advances in exploiting the Tat pathway for biotechnological benefit, the production of high value pharmaceutical proteins.
INTRODUCTION

About 20-30% of proteins synthesised in the bacterial cytoplasm are destined for extra-cytoplasmic locations (1). They pass the cytoplasmic membrane using specialised transport systems, involving gated pores, energy, and signal peptides to direct protein export. Two major protein export systems are known, namely the general Secretory (Sec) pathway and the Twin-Arginine Translocation (Tat) pathway (Figure 1). Most proteins use the Sec pathway common to all domains of life. The Tat pathway, the focus of this review, is more exclusive. For example, it has only ~30 native substrates in the Gram-negative bacterium Escherichia coli, and is not universally conserved (2).

Twin-arginine signal peptides

Specific N-terminal signal peptides direct proteins to the Sec or Tat pathways. Upon membrane translocation, the signal peptide is removed by signal peptidase to release the mature protein (3,4,5,6,7). Bacterial Sec and Tat signal peptides have a core tripartite structure: a positively charged N-terminal domain, a hydrophobic H-domain and a C-terminal domain (3,8). The signal peptidase cleavage site, usually Ala-x-Ala, resides in the C-domain (Figure 2). Tat signal peptides are defined by the twin-arginine motif S-R-R-x-F-L-K (x is any polar amino acid), joining the N- and H- domains (9,10). The RR-residues are essential for protein export (11,12) although an RR>KR substitution may be tolerated (12,13,14,15,16). Tat signal peptides are around 30 residues long, while Sec signal peptides range between 17-24 residues (17). To avoid Sec, Tat signal peptides are less hydrophobic than Sec signal peptides (18), and often contain basic residues as ‘Sec avoidance motif’ (19,20).

THE TAT PATHWAY

The Tat pathway was discovered in the 1990’s (21,22). Three thylakoidal membrane proteins were identified as essential for protein translocation via Tat, namely Tha4, Hcf106 and cpTatC.
Homologues were found in bacteria, archaea and mitochondria (26,27). In *E. coli*, these were, respectively, named TatA, TatB and TatC (9,28,29). Their importance for protein export was highlighted by protein mislocalisation in mutant strains (27,28).

Studies on thylakoidal and bacterial Tat pathways showed that they transport complex fully-folded proteins requiring cofactor insertion or oligomerisation (30). Accordingly, Tat facilitates many processes, including cell division and cell envelope biogenesis in *E. coli* (31,32). Tat mutants of other organisms, e.g. *Agrobacterium* and *Salmonella*, are impaired in quorum sensing, motility and pathogenesis (33,34). Unlike Sec, Tat even exports complex proteins from eukaryotes, as first evidenced for tightly folded dihydrofolate reductase or GFP provided with Tat signal peptides (35,36,37,38,39).

**Tat Genes**

The *E. coli* Tat system is constitutively expressed from the *tatABCD* operon, where *tatA* is expressed ~25- and ~50-fold more than *tatB* and *tatC*, respectively (40). Elsewhere in the genome, *tatE* is expressed. Presumably, *tatB* and *tatE* originate from gene duplications of *tatA* (26,41). Δ*tatABCDE* strains are viable, but display pleiotropic phenotypes (42).

Tat systems minimally function with TatA-like and TatC-like proteins, although this varies between organisms (43). The Gram-positive bacterium *Bacillus subtilis* expresses two minimal Tat pathways from distinct *tatAdCd* and *tatAyCy* operons (5). TatAdCd is co-expressed with its only known substrate, PhoD, during phosphate-starvation (44,45). TatAyCy is expressed constitutively as is the case for its substrates EfeB (YwbN), QcrA and YkuE (5,46,47,48,49). The *tatAc* gene is constitutively expressed elsewhere in the genome (50).

**Escherichia coli** Tat Components
TatA channels proteins across the inner membrane. Additionally, *E. coli* has two TatA-like proteins, TatB and TatE, involved in protein translocation (51). In contrast, *B. subtilis* has three TatA proteins, TatAd, TatAy and TatAc, but no TatB (5,45,52). TatA-like proteins can often be interchanged, functioning both intra- and inter-bacteria. In *E. coli*, TatE can substitute TatA (41), *B. subtilis* TatAd can substitute TatA and TatB (53), and TatAc forms active translocases with TatCd and TatCy (54) that functionally replace *E. coli* TatA and TatE (55). This functional overlap suggests a universal Tat translocation mechanism (53,56) (Table 1).

TatA-like proteins share structural features, including a short N-terminal domain exposed to periplasm or cell wall (57), a single transmembrane (TM) helix linked to an amphipathic helix (APH) positioned against the cytoplasmic membrane side (58), and an unstructured cytoplasmic tail (59,60,61,62)(Figure 3). Only few substitutions in TatA block its export function. In *E. coli* TatA, Gly33 in the “hinge region” connecting the TM helix and the APH is essential for export (63) as is Phe39 that anchors the APH to the membrane (64). Other APH residues and Gln8 in the TM helix are also imperative (65).

TatE shares 57% similarity with TatA (9). Despite its low abundance (40), TatE can substitute TatA (41), being a regular constituent of the Tat translocase (66). TatE interacts with Tat signal peptides, partially preventing premature signal peptide cleavage (67). Homologues function in many other bacteria as exemplified by *B. subtilis* TatAc, showing 45% amino acid sequence similarity to TatE (50).

TatB is functionally different from TatA. It binds Tat signal peptides and subsequent mature proteins. TatB’s APH and C-terminal region are longer than those of TatA (64,68,69)(Figure 3), and only mutations in “the hinge” and APH cause translocation defects (70). Single N-terminal amino acid substitutions in TatA allow complementation of ΔtatB strains (71,72). This bifunctionality of TatA mirrors the minimal TatAC systems in Gram-positive bacteria (5,73).

TatC is the largest Tat protein aiding substrate binding (74,75). It has 6 TM helices and an N-in C-in topology (Figure 3)(76). Crystallisation of *Aquifex aeolicus* TatC revealed that this protein resembles
a baseball glove or a cupped hand (77). TatC shows restricted structural flexibility, and is unlikely to undergo major conformational changes during translocation (78). A notable surface feature of TatC is the conserved Glu165 in A. aeolicus or Glu170 in E. coli (12,78). This residue is close to the binding pocket for Tat signal peptides, and its position in the membrane may perturb bilayer structure (79). Other important residues reside in TatC’s cytoplasmic N-region and first cytoplasmic loop (75,80).

**Bacillus subtilis** Tat Components

*B. subtilis* is the model organism for Gram-positive bacteria. Like all Gram-positive bacteria except *Actinomycetes* (81), *B. subtilis* has only TatA and TatC subunits (5,45). Accordingly, *B. subtilis* TatA is bifunctional, and able to functionally replace both both *E. coli* TatA and TatB.

TatAdCd exists as a complex of ~230 kDa, while TatAd alone assembles into complexes of ~270 kDa (53). TatAd has essentially the same structure as *E. coli* TatA (58,82). Mutations in the N-terminus of heterologously expressed TatAd block translocation in ΔtatB *E. coli* cells, implying their importance for TatAd’s bi-functionality. Further, electron microscopy (EM) showed that TatAd complexes are too small and homogeneous to serve as pores for passage of differently-sized Tat substrates (44). TatCd is very similar to *E. coli* TatC with 6 TM helices, and it both binds and stabilises TatAd (83,84).

TatAyCy and TatAy form complexes of ~200 kDa (52). The conserved hinge region of TatAy is equally important as in TatAd and *E. coli* TatA (82,85). The Pro2 residue in the N-terminal extracytoplasmic region of TatAy is key for bifunctionality and Pro2 mutations interfere with TatAy’s protein export function (50,82,85). The P2A mutation causes formation of fibril-like TatAy-TatCy assemblies of varying lengths, as shown by EM and atomic force microscopy (86). Compared to other TatC proteins, TatCy’s C-terminus is extended by five amino acids that are important for EfeB export (87). Further, depending on the substrate, the N-terminus and the first cytoplasmic loop of TatC, TatCd and TatCy are important for translocation (88).
TatAc is dispensable for Tat function in *B. subtilis* (5). However, when expressed in *E. coli*, TatAc interacts with TatCd and TatCy facilitating export of Tat substrates, like AmiA, AmiC and TorA (54). While TatAc can substitute TatA and TatB in *E. coli* (55), it cannot replace TatAd or TatAy in *B. subtilis*. Nonetheless, TatAc enhances TatAyCy function in *Bacillus* (50).

**Tat Complexes**

The largest native Tat substrates are heterodimeric formate dehydrogenases of ~150 kDa (89). However, Tat components themselves are relatively small, so the translocase must coalesce with multiple copies of each individual subunit. Live cell imaging fluorescence microscopy showed the TatABC system assembling on demand to newly synthesised Tat substrates (90). Two major complexes exist: TatBC and TatA. TatBC forms a receptor that binds Tat substrates at the membrane. This triggers proton-motive force (PMF-)dependent TatA recruitment and oligomerisation to form an active translocase (91,92). Starting from substrate-bound TatBC, the translocation event is slow, taking 1-3 minutes (93). Disassembly of TatA from TatBC requires substrate export (94).

TatB and TatC interact in a 1:1 stoichiometry (95), and several TatBC complexes bind substrate(s) (96). Low resolution EM structures of TatBC revealed a hemispherical morphology, with an internal cavity that could accommodate the signal peptide (74). Seven copies of TatBC fit into the 11-17nm reconstruction, and apparently more than one substrate binds at once (97). Cross-linking showed TatC is the primary interaction site of the signal peptide (92,98,99,100), which inserts deeply into TatC by adopting a hairpin-like conformation (77,98). Subsequently, the signal peptide exposes its C-region at the extracytoplasmic membrane side to signal peptidase (101,102). The signal peptide's H-domain interacts with TatB's TM helix (92) and Tat precursors initially bind to TatB at multiple sites (103). NMR data indicate that TatB's extended C-terminus is flexible to facilitate substrate binding (68). Each TatC monomer has two TatB contact sites (75,102,103,104,105). Intriguingly, the TM helix of TatB seems close to where TatA initiates translocase oligomer assembly, suggesting TatB is a
regulatory surrogate of TatA (106).

Coevolution analysis predicts TatA and TatB binding along TM5/6 of TatC, indicating a cluster of polar amino acids on TatC forms contact with polar side chains of TatA or TatB (107,108). Cross-linking studies indicate that, upon substrate docking, TatA binding to the TM6 of TatC is reduced, while TatA and TatB binding at the polar site of TatC increased. As the polar cluster site is adjacent to the docking site of the signal peptide (66), this docking event could cause the conformational rearrangements. Signal peptide binding also alters TatC’s resting state arrangement of head-to-tail to tail-to-tail. This opens the complex, allowing TatA access to the vacated polar cluster site and placing TatA adjacent to the concave face of TatC where TatA nucleation has been suggested (104).

TatA facilitates protein translocation, either by forming a pore (109) or by weakening the membrane (110). While TatA is not obligatory for TatBC formation or substrate docking (111), trace amounts are associated with TatBC, possibly representing nucleation points for TatA oligomers (112,113). TatA complexes in *E. coli* are in the 100-500 kDa size range (114), which can be resolved at 34 kDa intervals suggesting modular formation of three or four complexes at a time. Self-oligomerisation involves TatA’s TMH (59,68,115).

**MECHANISM OF TAT TRANSLOCATION**

The mechanism of Tat translocation is still not fully resolved. The ‘translocation pore model’ theorises that TatA forms a channel/pore for protein passage (109). More recent data favor the ‘membrane-destabilisation model’ (110,116).

Low-resolution images of TatA revealed a pore-like complex of varying diameters (8.5–13 nm) that would accommodate varying substrate sizes (109,117). A lid-like feature was identified which presumably resides on the cis-side of the inner membrane, as if the transporter were a trap door (118). Thus, TatA’s APH could form a hairpin to align with its TM helix, forming a pore for hydrophilic
cargo passage (119). Upon substrate docking onto TatBC, TatA protomers would assemble to suit the substrate size in an oligomeric ring format (120,121). Each TatA protein could be ‘zipped together’ by 7 salt bridges. Indeed, salt-dependence for the translocase was reported (85). However, this model is less attractive as APH insertion into the bilayer was not observed in other studies (57,122), while solution NMR indicated the APH fans outwards (59).

The second model proposes TatA complexes weaken/destabilise the membrane due to TatA’s short TM domain (59,110), restricting membrane thickness to its own length. Destabilisation only happens when Tat substrate is bound. When no substrate is bound, TatA’s APH immerses itself into the membrane, elongating the membrane. Upon substrate association, TatA re-orientates its APH outside of the membrane (122). The ‘switch’ from one state to another relates also to interactions between substrate and TatA’s APH. TatA does not recognise the ‘RR’ motif, but its APH interacts with the C-domain of signal peptides (123,124), suggesting the mature protein also interacts with TatA (92,103,125).

**Energy Requirements**

The Tat system is powered by the PMF, consisting of a pH gradient (ΔpH) and the electric potential (Δψ) across the membrane and Tat can use both (126,127,128,129,130). Accordingly, a protein-H⁺ antiporter mechanism was suggested (131). It was estimated that a counter flow of $7.9 \times 10^4$ protons is needed per protein (132). The PMF is not required for protein targeting or binding to TatBC, but is necessary for TatA oligomerisation and the actual export process (92,112,133,134).

**Tat Proofreading**

Generally, Tat substrates only exit the cytosol once they are fully folded and contain their cofactor (138,139,140,141,142). The respective proofreading is extremely tight in *B. subtilis* (166,167), but
less stringent in the thylakoidal Tat system as unfolded proteins are also transported (35). Requirements for conformational stability were studied using non-native Tat substrates, like PhoA and antibody fragments. These are only exported in oxidising conditions where disulphide bonds form (38,148,149,152). Nevertheless, human growth hormone (hGH) and interferon α2b are also exported via Tat without disulphide bonds formed (143), suggesting that they assume near-native states as shown for E. coli CueO, which is exported without its copper cofactor bound (144). Further, unstructured and hydrophilic polypeptides of 100-120 residues are tolerated by Tat, but a short hydrophobic stretch stops export (145,146,147). It thus seems that proofreading senses conformational flexibility. Mutations with less stringent proofreading were identified in E. coli TatABC (153), indicating that proofreading is undertaken by the Tat translocon. Proofreading may also involve chaperones known as Redox Enzyme Maturation Proteins (REMPs) (135). This is exemplified by TorD, facilitating Tat-dependent export of the trimethylamine-N-oxide (TMAO) reductase TorA in E. coli (136,137). Lastly, Tat may recruit quality control proteases to clear the translocase, as evidenced in B. subtilis for TatAyCy and the wall protease WprA (150,151).

FUTURE PERSPECTIVES FOR BIOTECHNOLOGY

Export of fully folded proteins is advantageous for biotechnological applications. Owing to their rapid growth, high yield, ease of scale-up and cost-effectiveness, ~30% of biotherapeutics are synthesised in bacteria (154). Preferably, these proteins are exported to the oxidising periplasm with an N-terminal signal peptide. This is the usual approach in B. subtilis, where most translocated proteins are secreted into the medium (155,156). In E. coli, export to the periplasm decreases proteolysis and eases downstream purification as the periplasm only contains 4% of the proteome (157,158). Here, proteins are released by osmotic shock (159). Recently it was shown that biotherapeutics can be exported via Tat (143,160). Moreover, Tat overexpression in so-called TatExpress cells resulted in 5-fold increased export of hGH (161). The ‘gold standard’ of bacterial
protein production, protein 'secretion' to the medium, was achieved when *B. subtilis* TatAdCd was expressed in ΔtatABCDE strains, as this led to protein release from the periplasm (162). Current efforts to apply Tat in biotechnology focus on different biotherapeutics and increased protein yields. It is anticipated that this area will expand rapidly once the mechanisms underlying Tat-dependent protein export are fully understood.

REFERENCES


140. **Goosens VJ, Monteferrante CG, Van Dijl JM.** 2014. Co-factor insertion and disulfide bond


<table>
<thead>
<tr>
<th></th>
<th>Molecular mass of gene product (kDa)</th>
<th>Molecular mass of complex (kDa)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E. coli</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TatA</td>
<td>11</td>
<td>100-500</td>
<td>(114)</td>
</tr>
<tr>
<td>TatB</td>
<td>18</td>
<td>100</td>
<td>(111)</td>
</tr>
<tr>
<td>TatC</td>
<td>30</td>
<td>220</td>
<td>(111)</td>
</tr>
<tr>
<td>TatE</td>
<td>7</td>
<td>50-110</td>
<td>(41)</td>
</tr>
<tr>
<td>TatABC</td>
<td>440</td>
<td>440-580</td>
<td>(165)</td>
</tr>
<tr>
<td><strong>B. subtilis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TatAd</td>
<td>7</td>
<td>270*</td>
<td>(53)</td>
</tr>
<tr>
<td>TatCd</td>
<td>27</td>
<td>100</td>
<td>(54)</td>
</tr>
<tr>
<td>TatAy</td>
<td>6</td>
<td>200*</td>
<td>(166)</td>
</tr>
<tr>
<td>TatCy</td>
<td>28</td>
<td>66</td>
<td>(54)</td>
</tr>
<tr>
<td>TatAc</td>
<td>6</td>
<td>100</td>
<td>(54)</td>
</tr>
<tr>
<td>TatAdCd</td>
<td>230</td>
<td></td>
<td>(54)</td>
</tr>
<tr>
<td>TatAyCy</td>
<td>200</td>
<td></td>
<td>(54)</td>
</tr>
</tbody>
</table>

Table 1. Molecular mass of Tat proteins and complexes in *E. coli* and *B. subtilis*

The predicted molecular mass of *E. coli* and *B. subtilis* Tat proteins is according to gene sizes. The molecular mass of complexes was approximated mostly by blue-native gel electrophoresis, except those determined by gel filtration chromatography, as indicated with a *
Figure legends:

Figure 1. Two alternative protein translocation pathways, Sec and Tat.

Proteins always originate from translating ribosomes (R). Their N-terminal signal peptide (OmpA or TorA in this overview) will direct the nascent polypeptide chain to the correct translocase (Sec or Tat, respectively), which may be aided by chaperones. The unfolded Sec protein is transferred to SecA where it is threaded through the SecYEG channel in the plasma membrane, powered by repeated cycles of ATP binding and hydrolysis. In the oxidising periplasm, the unfolded protein assumes its tertiary fully folded state. The Tat-dependently translocated protein is fully folded within the cytoplasm where it may also acquire its cofactor. Once directed to TatBC, TatA protomers are recruited to translocate the protein across the cytoplasmic membrane. Energy required for this process is created by the proton-motive force. Please note that mRNA molecules are schematically represented by an interrupted line, synthesized proteins by uninterrupted lines, and translocase subunits by cylinders.

Figure 2. Comparison of Tat (TorA) and Sec (OmpA) signal peptides

The structure of Tat and Sec signal peptides includes three regions, namely a basic N-domain (blue), a hydrophobic H-domain (red) and a polar C-domain (green). A signal peptidase cleavage site (AxA) is positioned prior to the mature protein (pink). The amino acid sequences of the TorA and OmpA signal peptides are specified. Tat signal peptides (above) have a consensus motif containing 'twin-arginines', while Sec signal peptides do not contain this motif. Sec signal peptides tend to be shorter, with fewer residues in their N and H-domains, than Tat signal peptides.

Figure 3. Schematic representation of E. coli TatABC components in the plasma membrane

Three essential components form the Gram-negative bacterial Tat complex, namely TatA, TatB and TatC. TatA/E and TatB have similar topologies in that they have one TM helix domain with a short periplasmic N-terminal region, a tilted APH and an unstructured C-terminus on the cytoplasmic side.
of the plasma membrane. Notably, TatB is larger than TatA with a longer C-terminal tail. TatC is significantly bigger as it containing 6 membrane-embedded helices with both the C- and N-terminal ends residing in the cytoplasm. The helices 5 and 6 do not fully span the membrane, which may contribute to TatC’s function.
Figure 1
Figure 2