Soon after synthesis, proteins are transported to their respective destinations to perform various cellular activities needed for the survival of the cell. Protein trafficking is a complex process. In prokaryotes, 16 different mechanisms are involved in transporting the proteins to their respective destinations. Interestingly, except one, the rest of them transport unfolded proteins. A number of proteins acquire folded conformation before targeting/transporting across membranes. Such prefolded proteins are transported through a novel pathway, known as twin arginine transport (Tat) pathway. The Tat pathway is found in both Gram-positive as well as Gram-negative bacteria and plays a key role in various cellular activities including pathogenesis. The review provides a comprehensive picture of the mechanism of the Tat pathway and describes its role in pathogenesis.

Keywords: Protein transport, signal peptide, signal recognition particle, twin arginine transport pathway.

Bacteria need to transport proteins across their membrane to have constant interaction with their extracellular milieu. While struggling to survive in harsh environment, they have designed a number of mechanisms to send certain proteins across/into the membranes. These proteins perform important functions such as cell envelope biogenesis, nutrient acquisition, motility, cell to cell communication, etc. and play a key role in cell viability. In prokaryotes, about 16 distinct protein transport systems are known to contribute towards translocation/targeting of proteins across/into the membranes. Some of them, such as T4S, T4P, Tat, Sec, Yide are found operational in both Gram-negative and Gram-positive bacteria, whereas the other transport systems are found exclusively either in Gram-negative (T6S, Fla, T3S, Cu, T2S, LOL, T5S, TPS and OmpB5) or in Gram-positive (Sort and Esx) bacteria. Most of the extracellular proteins acquire functional conformation only after reaching their destination. However, certain proteins can acquire folded conformation only in the cytoplasm. The periplasmic or extracellular environment, due to a variety of reasons, is unsuitable for their folding. These include (i) dependence of proteins on large cofactors for activity; (ii) acquisition of folded conformation while they are still in cytoplasm; (iii) unfavourable extracellular environment, such as high temperature, salt concentration, etc. for folding; (iv) failure to fold in a relatively more oxidizing environment prevailing in periplasmic space of Gram-negative bacteria and (v) multi-subunit protein complexes. However, folded conformation acts as a structural hinderance for using the conventional protein transport pathways evolved exclusively for transport of unfolded proteins. Such a complex task of transporting/targeting of prefolded proteins across the hydrophobic phospholipid bilayer, without damaging its integrity, is operated through unique protein translocases and chaperones. In addition to these two requirements, unique structural features found in the signal peptides of prefolded proteins contribute towards chaperone-specific interactions through a novel mechanism known as Sec-avoidance. A signature sequence motif with twin arginines are found in the N-terminal positive region and it is shown to be essential for membrane targeting of all prefolded proteins. Therefore, the unique transport pathway is named as twin arginine transport (Tat) pathway. Originally, the Tat pathway was discovered in plants while studying the transport of nuclear genome coded proteins to the thylakoid membrane. Later, Weiner et al. have shown its existence in Escherichia coli while unravelling the mechanism of membrane targeting of redox proteins, such as nitrate reductase (NapA), trimethylamine N-oxide reductase (TorA) and molybdoenzyme dimethyl sulphoxide reductase (DmsABC). Since then, a number of Tat substrates were shown to exist both in Gram-negative and Gram-positive bacteria. A few examples are shown in Table 1. Initially the Tat pathway is assumed to have evolved exclusively to transport/target proteins containing large cofactors. However, in recent times a number of cofactor-less proteins or proteins with small cofactors are shown to take Tat route for targeting/translocating across the membrane. The focus of this review is to highlight the mechanism of transport and its role in pathogenesis.
### Table 1. Examples of Tat pathway translocated proteins, their function and cofactor involved

<table>
<thead>
<tr>
<th>Organism</th>
<th>Gene/translocated protein</th>
<th>Function</th>
<th>Cofactor/ions involved</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Escherichia coli</em></td>
<td>hyaA; hydrogenase-1 subunit</td>
<td>Energy metabolism</td>
<td>Iron sulphur clusters</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>torA; TMAO reductase catalytic subunit</td>
<td>Anaerobic respiration</td>
<td>Molybdenum guanine dinucleotide</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>dmsA; DMSO reductase catalytic subunit</td>
<td>Anaerobic respiration</td>
<td>MGD and iron sulphur clusters</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>amicC; cell wall amidase</td>
<td>Cell wall biogenesis</td>
<td>Not known</td>
<td>58</td>
</tr>
<tr>
<td><em>Pseudomonas aeruginosa</em></td>
<td>napF; ferrodoxin</td>
<td>Cell growth</td>
<td>Iron</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>nosZ; nitrous oxide reductase</td>
<td>Cell growth</td>
<td>Copper</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>plcN; phospholipaseC</td>
<td>Infection</td>
<td>Calcium</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>plcH; phospholipaseC</td>
<td>Infection</td>
<td>Calcium</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>psdN; putative aminotransferase</td>
<td>Iron uptake</td>
<td>Pyridoxal phosphate</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>napA; nitrate reductase catalytic subunit</td>
<td>Cell growth</td>
<td>Molybdenum guanine dinucleotide</td>
<td>59</td>
</tr>
<tr>
<td><em>Mycobacterium tuberculosis</em></td>
<td>fecB2; iron transport protein Fe(III) dicitrate transporter</td>
<td>Iron uptake</td>
<td>Not known</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>plcD; phospholipase C precursor</td>
<td>Infection</td>
<td>Calcium</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>plcC; phospholipase C precursor</td>
<td>Infection</td>
<td>Calcium</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>ugpB; glycerol-3-phosphate transport</td>
<td>Nutrient uptake</td>
<td>Not known</td>
<td>68</td>
</tr>
<tr>
<td><em>Agrobacterium tumefaciens</em></td>
<td>Nitrate reductase large subunit</td>
<td>Cell growth</td>
<td>Flavin adenine dinucleotide</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Rieske FeS sulphur cluster</td>
<td>Energy metabolism</td>
<td>Iron</td>
<td>37</td>
</tr>
</tbody>
</table>

Figure 1. Sect(A) and Tat (B) specific signal peptides. The signal peptidase I (i), the signal peptidase II site (ii) cleavage sites are shown by inverted arrows. The Tat motif found at N-region of Tat specific signal peptides are shown in bold letters.

### Tat signal peptides

The Sec and Tat-signal peptides show significant differences despite sharing a number of structural similarities (Figure 1). In view of striking structural similarities between Sec and Tat signal peptides, the Tat pathway has been considered to be a recently evolved protein translocation machinery13. As in the signal peptides of Sec substrates, the Tat signal peptides contain three major domains, i.e. a positively charged amino terminus (N-region), the hydrophobic core region (H-region) and the more polar cleavage region (C-region)14. The positively charged N-region contains a consensus signature sequence with conserved twin arginine residues (S/T–R–R–X–θ–θ). Substitutions to the twin arginines of the Tat motif, even with similarly charged lysines hamper transport process. The signal recognition particle (SRP) specifically interacts with Tat motif before taking the prefolded proteins to Tat-specific translocases. As twin arginines are an absolute requirement for transport of prefolded proteins, the process is named as twin arginine Transport pathway7,15. The H-region of the signal peptide is the hydrophobic core of the signal sequence and varies in length from 7 to 15 amino acids. Hydrophobicity at this region determines the efficiency of the translocation process and it follows a sigmoidal relation (requirement of a minimum hydrophobicity) for the translocation16. The helical structure that starts at N-region continues through H-region till it is disturbed by the presence of prolyl or glycyl residues and is known to help in insertion of the substrate into the lipid bilayer. The C-domain contains the leader peptidase cleavage site with conserved residues at –1 and –3 positions. Analysis of the C-domain revealed the existence of conserved residues required for both type I (SPase I) and type II (SPase II) leader peptidase17,18.

### Tat translocases

The mechanism of Tat transport has not been well established but appears unrelated to that of other membrane transporters19. In *E. coli*, the Tat translocase is made of three proteins TatA, TatB and TatC. These three proteins are coded by an operon consisting of four open reading frames (ORFs) tatA, tatB, tatC and tatD20. Although tatABC are shown essential, deletion of tatD had no influence on transport of Tat-substrates21. In addition to the tat operon, the tatA paralogue, tatE is found elsewhere in the chromosome. TatE, due to its homology to TatA complements TatA functions. TatA and TatB have considerable sequence similarity and are associated22. In certain
Figure 2. The complete translocation mechanism with individual steps highlighting the entire Tat process is shown.

1. The interactions between the SRP and the Tat motif.
2. Assembly of the ‘Hitchhiking’ partners, metal ions or Co factors.
3. TatBC complex will then interact with the above described assembly leading to oligomerization of TatA.
4. Oligomerized TatA will form the pore component and the protein to be translocated will be transported to the exterior.

Cases, tatA has been shown to complement tatB. Supporting this situation in a few Gram-positive bacteria, the Tat-transport machinery contains only TatA and TatC proteins. TatA is the most abundant protein of all Tat-translocases and is estimated to be present at around 20 times more than TatB and TatC. TatA is a 9.6 kDa polypeptide with 89 amino acids and contains a N-terminal hydrophobic \( \alpha \)-helix region followed by a short hinge region and a longer amphipathic \( \alpha \)-helix region. The TatA forms tetrameric homooligomers in the cytoplasmic membrane generating an intramolecular pore. Such a proposal is supported by the chemical cross-linking studies and purification of TatA oligomer with a molecular mass of 600 kDa (refs 26,28,29). TatA is found at the cytoplasmic face of the inner membrane and functions only at the last stage in translocation process. The TatA upon interacting with the protein complex consisting of TatB, TatC and Tat-substrate, through its trans membrane domain (TM) acquires, oligomeric state (Figure 2). The product of tatB is a 18.4 kDa polypeptide with 171 amino acids. The TatB has a structure very similar to that of TatA but has a considerably longer C-terminal portion with a predicted random coil. TatB plays an important role in translocation process since deletion of the tatB gene abolishes the substrate translocation. TatB, in presence of TatC, is shown to interact with the entire length of the signal peptide and also with the mature protein almost 20 residues away from the signal peptide cleavage site.

The product of tatC is an essential component of the Tat pathway. This 28.3 kDa protein is found to be a highly conserved component of the Tat system in bacteria and chloroplasts. The TatC contains a six TM domain with the amino and carboxyl termini located at the cytoplasmic face of the membrane and binds to the signal peptide of the substrate protein. As shown by the docking and chemical cross-linking studies, the N-terminal portion of TatC facing the cytoplasmic side plays a critical role in recognition of the Tat substrates, particularly recognizing the twin arginines of the Tat motif. The TatB and TatC form an initial complex which recognizes the Tat substrates. As in tatB mutants, the tatC null mutants have failed to transport prefolded proteins indicating their importance in the transport process.

Signal recognition particle

The SRP and its membrane-associated receptor help in targeting nascent secretory and membrane proteins. The role of SRP in targeting the substrates is (i) to recognize the substrate and the translocation machinery to be utilized, (ii) to protect the ribosome-nascent chain complex from protease degradation, (iii) to efficiently deliver the protein to the translocation channel. The process begins with the recognition of the signal motif by the SRP. Although the signal sequence is degenerate, the SRP is still capable of recognizing its cognate partner and culminate in targeting the proper partner to its specific site. This remarkable property is due to the recognition of the threshold level hydrophobicity of the signal peptide by the M-domain (rich in methinone residues) of the SRP. This
mechanism is similar among different levels of organisms and is thus known to have been conserved evolutionarily. The Tat substrates have specific targeting chaperones or Tat motif binding proteins that recognize the cognate substrates efficiently. These chaperones and also the TatC have been shown to bind to the Tat signal sequence and presumably regulate folding, substrate addition and complex assembly prior to the translocation. This step is known to be unique for the Tat system and is considered to be the major folding quality checkpoint in the process. These are basically cytoplasmic proteins ranging between 20 and 30 kDa across the species. Studies on transport of Tat-specific redox enzymes have identified the interactions between redox enzyme maturation protein (REMP) which acts as SRPs to TorA, DmsD, etc. The REMP proteins help in substrate maturation and have been shown in vivo to bind to the TatBC complex implicating their direct involvement in the targeting process. Studies conducted on trimethylamine nitrate oxidoreductase (TMAO)-specific SRPs TorD and TorA have revealed its involvement in protecting its cognate substrate (TMAO) from protease degradation, cofactor insertion as well as in regulation of transport process. Hatzixanthis et al. have shown interaction of TMAO with TorD and TorA covering the entire length of signal peptide which includes the twin arginine motif (signal peptide residues 2–22), the h-region and the Sec-avoidance motif (signal peptide residues 10–36). These interactions are believed to take Tat-substrates away from Sec transport machinery and appear to be the major quality control aspect in Tat-pathway.

The translocation process

The details of translocation process are shown in Figure 2. Upon recognition of the cognate substrate by the chaperone or the SRP, the SRP bound to the RR motif is recognized by the TatC protein. The TatB and TatC then form an initial association with many copies of each of the constituent subunits. This TatBC complex is the membrane binding site for Tat substrates which then recruits TatA to form the active translocation site. TatA forms homo-oligomeric ring-like structures that constitute the protein translocating channels of the Tat system (Figure 2). The proton motive force generated due to pH gradient across the membrane appears to facilitate the transport of prefolded proteins. As a result of a series of co-ordinated events, the prefolded proteins successfully translocate to the cell exterior. Recent studies conducted by Kostecki et al. have shown that the Tat substrate–chaperone interactions are independent and are shown to occur even in the absence of TatABC and E proteins. They have also shown the permissive residues involved in the recognition process.

Role in pathogenesis

The Tat translocase mutants have shown a wide response on the physiology of the cells. They mainly affect permeability, cell wall integrity, cause deformations in the biofilm formation, defects in iron acquisition, copper homeostasis systems, cell invasion, symbiosis and aerobic respiration. These defects have been shown to affect the pathogens ability to colonize and survive in the host system which is critical for the existence of the pathogen.

The involvement of the Tat pathway is also seen in the secretion of two major virulent factors in Pseudomonas aeruginosa, in Mycobacterium tuberculosis as well as in Staphylococcus aureus. Following this discovery, the total genome sequences of most of the pathogenic bacteria were analysed using the bioinformatic tool, TATFIND developed specifically to identify Tat-substrates. The involvement of the Tat pathway is also seen in the secretion of two major virulent factors in Pseudomonas aeruginosa, in Mycobacterium tuberculosis as well as in Staphylococcus aureus. Following this discovery, the total genome sequences of most of the pathogenic bacteria were analysed using the bioinformatic tool, TATFIND developed specifically to identify Tat-substrates (http://signalfind.org/tatfind.html). Mining of the M. tuberculosis genome using TATFIND has revealed existence of 31 Tat substrates. Among these 31 proteins, only 11 withstood more stringent bioinformatic procedures. The list consists of important proteins, including those that contribute for virulence. Although in silico predictions were made to identify Tat substrates very few in vitro studies are available to validate these bioinformatic predictions. Logically, if tat negative mutants are generated, comparative analysis of wild type and mutant extracellular proteome is expected to provide direct evidence for Tat substrates. However, attempts made to obtain tat mutants of M. tuberculosis were unsuccessful. The tat knockout of M. tuberculosis have only been obtained in the merodiploid strains, suggesting that the Tat is an essential component for survival of the organism.

In shiga toxin producing enteroohaemorrhagic E. coli O157:H, the causative agent of uncomplicated diarrhoea, haemorrhagic colitis and haemolytic-uremic syndrome, the tatABC mutants have shown attenuated toxicity due to substantial reduction in the amount of Stx1. Similar situation was seen in Salmonella enterica serovar Enteritidis, a causative agent of gastroenteritis in humans. S. enterica is known to propagate through ingestion of infected eggs and chicken. The tatB and tatC mutants of S. enterica have shown impaired invasion into Caco-2 cells and caused considerable reduction in systemic spread in chickens. The plant pathogen, Dickeya dadantii causes the soft rot diseases in many plant species. When tatC is deleted, the mutant strains have shown reduced virulence and hypersensitivity to copper. They are also found to be defective in both iron uptake and motility. In total, tatC mutants of D. dadantii have shown reduced fitness and virulence capabilities. Supporting the role of Tat-pathway in infection and pathogenesis, the tat negative strains of Agrobacterium tumefaciens failed to induce tumours in plant tissues.
Role in transport of phospholipase C

The phospholipase Cs (PLCs) are present in a variety of prokaryotic and eukaryotic organisms. In eukaryotes PLCs are important in signalling, apoptosis, inflammation and oncogenesis. In prokaryotic pathogens, they are shown to be responsible for gas gangrene, haemolysis and cell to cell spread of pathogens. In *M. tuberculosis*, four putative PLC coding genes are identified and all of them shown to be upregulated during the first 24 h of macrophage infection. Recent studies have demonstrated PLCs of *P. aeruginosa* and *M. tuberculosis* as Tat substrates, highlighting the importance of Tat-pathway in the pathogenesis.

Future perspectives

After discovering the Tat-pathway in bacteria and thylakoids, a number of attempts have been made to identify similar protein transport systems in animal cells, particularly in mitochondria. However, till date no discovery is yet made to show the existence of Tat homologues or any other system for transport of prefolded proteins in animal cells. In the absence of Tat homologues in animal cells, the transport mechanism appears to be unique to prokaryotes and plant thylakoids. Given the significance of Tat machinery in transport of virulence factors it can be viewed as a potential drug target. Rational designing of its inhibitors is expected to serve as a potential drug molecule. Hence, with some more progress in our understanding on chaperone–Tat motif interactions and on Tat translocase–substrate interactions, it would be possible to design molecules that specifically inhibit Tat translocases. Such inhibitors, in the light of bacteria developing resistance to most of the known antibiotics are expected to serve as a new generation drug candidates for controlling bacterial infections. As there exist no Tat homologues in animals, these novel drug molecules are expected to cause less or no side effects in hosts.

Heterologous expression of recombinant proteins is a major challenge in bacteria. The conventional expression systems produce considerable amounts of mis/unfolded proteins causing difficulties for the purification of active recombinant proteins from the cytoplasmic proteins of the host. If fully folded, active proteins are alone transported into culture media, it would be advantageous. First, every recombinant molecule to be transported into the medium will be proof read and only those molecules that acquired folded conformation will be sent into extracellular medium, leaving behind the misfolded proteins in the cytoplasm. As extracellular proteins are relatively few, purification of recombinant proteins, especially if they have C-terminal affinity tag would be a single-step process. If an expression system that facilitates expression of fusion proteins with appropriate Tat signal peptide is developed it would exploit the unique features of the Tat machinery for transport of folded proteins. Initial success has been seen while achieving expression of green fluorescence protein as active periplasmic protein. If we gain clear understanding on Tat-specific translocation process in generally regarded as safe organisms, it can be exploited for designing new generation expression vectors. Using these vectors it would be possible to selectively direct only the active, folded proteins into culture media.

41. Zopf, D., Bernstein, H. D., Johnson, A. E. and Walter, P., The methionine-rich domain of the 54 kDa protein subunit of the signal recognition particle contains an RNA binding site and can be crosslinked to a signal sequence. EMBO J., 1990, 9, 4511–4517.


ACKNOWLEDGEMENTS. S.P. is the recipient of the Shantha Bio-technics Excellence Junior research fellowship. Research in D.S. laboratory, is supported by DST, DBT, CSIR and DRDO.

Received 10 May 2010; revised accepted 7 January 2011.