INTRODUCTION
CHAPTER - I

To elucidate the role of twin arginine (tat) motif in membrane targeting of OPH in *B. diminuta*
1.1. ORIGIN OF THE PROBLEM

The genesis of this work is identification of a structural motif that facilitates transport / targeting of prefolded proteins across / into the membrane in organophosphate hydrolase (OPH) isolated from a number of soil bacteria. The membrane associated OPH is a homodimer and possesses Zn ions as cofactor (Omburo et al., 1992; Benning et al., 1995). It hydrolyzes characteristic triester linkage found in structurally diverse group of neurotoxic organophosphates and nerve gas agents (Munnecke et al., 1976; Munnecke et al., 1982; Dumas et al., 1990). In view of its unusual catalytic properties and exceptional stability, a number of attempts have been made to produce recombinant OPH in large quantities. The recombinant OPH thus produced has successfully been used to develop biosensors and devices intending to use for detection and decontamination of op residues (Randhir et al., 2005).

The OPH producing bacteria have been isolated from diverse geographical regions (Sethunathan and Yoshida, 1973; Serdar et al., 1982; Rani and Lalitha kumari, 1994; Somara and Siddavattam, 1995; Horne et al., 2002). In all these isolates existence of highly conserved organophosphate degrading (opd) gene is shown to be responsible for coding OPH enzyme (Mulbry and Karns, 1989; Serdar and Gibson, 1985). The opd gene is constitutively expressed and in fact presence of op compounds in the culture medium failed to exert any influence on levels of OPH expression (Mulbry and Karns, 1986). In most of the soil bacteria the opd gene is found on large dissimilar plasmids (Mulbry et al., 1986; Harper et al., 1988; Somara and Siddavattam, 1995). As evidenced by hybridization
studies and restriction pattern a 7 kb DNA region containing opd gene was found to be conserved in plasmids pPDL2 and pCMS1 isolated from Flavobacterium sp. ATCC27551 and Brevundimonas diminuta respectively (Mulbry et al., 1986). However no homology was evident in these two plasmids outside this conserved 7 kb region. Such unusual observation has prompted us to undertake detailed sequence analysis of the conserved regions. In plasmid pPDL2, the conserved sequence showed feature of a complex transposon (Siddavattam et al., 2003). Mobile elements were found flanking opd gene and an adjacently located open reading frame (ORF) that codes for a protein of 306 amino acids (Fig. 1. 1).

**Fig. 1. 1**

Map of the sequenced region opd gene cluster from pPDL2. The eight ORFs identified are shown, together with the locations of the IR sequences that flank ISFisp1 (S Siddavattam et al., 2003. App. Environ. Microbiol.)

An IS elements that show similarity to IS21 class of IS elements was identified upstream of the opd gene. Downstream of orf306 a mobile element showing high similarity to transposon Tn3 was identified. Existence of identical plasmid borne transposon-like opd gene cluster supports horizontal transfer of opd gene cluster among soil bacteria. Though considerable information is available on biochemistry and catalytic properties of OPH, information pertaining to the mechanism of membrane targeting is scarce. The present
study is a meticulously designed effort to throw light on these aspects. Further a moderate effort is also made to identify the cis elements that contribute for regulation of opd gene expression.

1.2. REVIEW OF EXISTING LITERATURE

In order to highlight the importance of the current work to the reviewer an attempt is made to provide a brief account on biodegradation of organophosphates and on membrane targeting/transport of bacterial proteins. The organophosphates are extensively used across the globe as insecticides to control number of insects that damage economically important crops (Brajesh, 2008). Though they are susceptible for enzymatic hydrolysis persistent and indiscriminate use resulted in accumulation of their residues in various components of environment. As op compounds exclusively inhibit acetyl cholinesterase (AChE) microbes and plants which do not have this target molecule are least affected. A good number of soil bacteria that use op compounds as source of carbon and energy have been isolated from diverse geographical regions (Singh and Walker, 2006). In fact, isolation of methyl parathion degrading Flavobacterium sp. from the rice fields of International Rice Research Institute is the first report on biodegradation of organophosphates (Sethunathan and Yoshida, 1973). After this report a number of parathion and methyl parathion degrading microorganisms were isolated (Munnecke et al., 1974; Rosenberg and Alexander, 1979; Nelson, 1982; Serdar et al., 1982; Lewis et al., 1985; Mulbry et al., 1986; Chaudhry et al., 1988; Ou et al., 1989; Misra et al., 1992; Rani and Lalithakumari 1994, Somara and Siddavattam 1995). In most of these isolates a 40 KD membrane associated homodimer designated as organophosphate hydrolase was shown to
be responsible for degradation of organophosphates (Serdar et al., 1989; Mulbry et al., 1986; Mulbry et al., 1986).

1.2.1. Organophosphorus hydrolase (OPH)

OPH hydrolyses triester linkage found in structurally diverse group of op compounds and nerve gas agents (Benning et al., 1994). Its potential in detoxification of organophosphate nerve agents and pesticides has been well documented (Munnecke et al., 1976; Munnecke et al., 1982; Dumas et al., 1990). The OPH has been purified independently and characterized from number of sources (Sethunathan and Yoshida, 1973, Serdar et al., 1982). It is identified as one of the members of the amidohydrolase super family (Holm and Sander, 1997). Members of this superfamily utilize one or two divalent metal ions to activate a hydrolytic water molecule for a nucleophilic attack at tetrahedral phosphorus or trigonal carbon centers. OPH is a homodimeric metalloprotein and the active site of the native OPH enzyme contains two zinc ions per monomer (Omburo et al., 1992; Benning et al., 1994). Though Zn$^{2+}$ ions serve as natural cofactors they can be replaced with Co$^{2+}$, Cd$^{2+}$, Ni$^{2+}$, or Mn$^{2+}$ ions without compromising on catalytic properties of the enzyme (Omburo et al., 1992).

1.2.2. Genetics of op compound degradation

*Flavobacterium* sp. ATCC27551, *Brevundimonas diminuta* were used as model organisms to gain insights on genetics and molecular biology of op compound degradation. In these two organisms almost identical organophosphate degrading (opd) genes were identified on otherwise dissimilar indigenous plasmids (Serdar et al, 1982; Mulbry and Karns, 1986). Among opd plasmids, pPDL2 (40kb, isolated from
*Flavobacteium* sp. ATCC27551) and pCMS1 (66Kb, isolated from *Brevundimonas diminuta*) were used for further characterization (Mulbry et al., 1987).

**Fig. 1. 2**

**Organizations of three known organophosphorus-degrading genes.** The shapes indicate different gene locations and the direction of transcription. **a)** Structure of the opd (organophosphorus degrading) gene from the *Flavobacterium* sp. genome, which includes a complete *istAB* operon, the *tnpA* and *tnpR* genes and orf243, a gene that encodes for metabolite utilization. Two orfs on a complementary strand encode for a protein of unknown function (Siddavattam et al., 2003). **b)** Genomic structure of the *opdA* gene from the *Agrobacterium radiobacter* genome, which includes *tnpA* and inverted repeats (IRs) (left inverted repeats (LIRs) and right inverted repeats (RIRs)), *opdA* and two orfs of unknown function (Horne et al., 2003). **c)** Genomic structure of the *mpd* gene cluster from the *Ochrobactrum* sp. genome, which includes IRs, *tnpA* and three orfs of unknown function (Zhang, 2006). (The Figure was adapted from Brajesh, 2008. *Nature Reviews Microbiology*).

In these two dissimilar plasmids *opd* gene was found as part of complex transposon (Siddavattam et al., 2003). Interestingly similar organization is seen for the *opd* gene reported from *Agrobacterium radiobacter* P230 isolated from Australian and Chinese agricultural soils (Horne et al., 2002). In a typical transposition assay Oakeshott and his associates have successfully demonstrated the event of *opd* element transposition in *E. coli* (Horne et al., 2002). A diagrammatic representation is shown below to draw a better understanding on organization of *opd* elements found in soil bacteria (Fig. 1. 2). If the structural organization of *opd* elements is clearly seen most of the *opd* associated ORFs are transposases. However in the *opd* element found in *Flavobacterium* sp. a novel esterase is
seen adjacent to *opd* gene and is later shown to hydrolyze meta fission products generated during degradation of aromatic compounds (Khajamohiddin et al., 2006).

### 1.2.3. Protein secretion in bacteria

A cell needs to keep the integrity of its essential processes and retain its reactants and enzymes. Bacteria often live in harsh environments such as the intestinal lumen, the apoplast of plant leaves and a variety of other locales where exposure to high temperature, saturating levels of salt and even raw sewage are not uncommon. To survive in these realms, all bacteria have evolved a resolute outer membrane or wall to protect themselves against these severe surroundings. This strategy not only prohibits permeation of hydrophobic substances and other potentially hazardous chemicals that might otherwise be lethal, it also creates new extracytoplasmic locations (e.g., the periplasm, the inner and outer membrane, the extracellular space) that function as both a buffer to extracellular stress and create a new environment for protein accumulation and protein chemistries (Bronstein et al., 2004).

The transport of proteins across the cell envelope is a basic function found in all groups of bacteria. Evolution has produced a remarkable array of mechanisms to export proteins. As shown in Fig. 1, 3 sixteen such systems, which handle protein secretion, sorting and membrane integration, are present in Bacteria (Holland, 2004; Economou et al., 2006). Among these pathways T6S, T4S, Fla, T3S, CU, T4P, T2S, LOL, T5S, TPS, Omp, TAT, Sec, YidC were found in gram negative bacteria, whereas T4S, T4P, TAT, Sec, YidC, Sort, Esx transport mechanisms were identified in gram positive bacteria (Effrosyni et al., 2007).
Analysis of a large number of bacterial genomes indicate that up to 17% of *Proteobacteria* genomes encode for proteins with signal sequences for the general secretory pathway (GSP) (Bendtsen et al., 2005), with many additional secretion systems and substrate proteins being present in most species. Secreted proteins perform few functions and most important of them are biogenesis of the cell envelope, acquisition of nutrients, motility, intercellular communication etc. (Bendtsen et al., 2005). The GSP and its components are found in all three kingdoms of life: in bacteria, archaea and eukaryotic organelles (chloroplasts but also in the endoplasmic reticulum) and provides a general mechanism for the transport of proteins across the cytoplasm or organelle membrane (Albert, 2002). While this transport process is sufficient for the secretion of proteins in Gram-positive species, Gram-negative species are posed with a specific problem, the transport across a second membrane system, the outer membrane (OM). In fact, the
difference found in OM secretion mechanisms are basis for classification of secretion pathways in Gram-negative bacteria. In Fig. 1. 4 a schematic representation is given showing connections among different secretion pathways in Gram-negative bacteria.

**Fig. 1. 4**

Summary of known bacterial secretion systems. In this simplified view only the basics of each secretion system are sketched. HM: Host membrane; OM: outer membrane; IM: inner membrane; MM: mycomembrane; OMP: outer membrane protein; MFP: membrane fusion protein. ATPases and chaperones are shown in yellow. (Tseng et al., 2009. *BMC Microbiology*)

1.2.3.1. Type I secretion systems

Type one secretion systems (T1SS) or ATP-binding cassette (ABC) transporters are heterotrimeric complexes consisting of an inner membrane (IM) ABC exporter, a membrane fusion protein (MFP) and a pore-forming, outer membrane protein (OMP) (Delepelaire, 2004; Holland et al., 2005). T1SS allow the secretion of a wide range of substrates (proteinaceous and nonproteinaceous) from the cytoplasm to the extracellular space in a single step, without a stable periplasmic intermediate. Most protein substrates
described so far possess a C-terminal signal sequence which is characterized by loosely conserved secondary structures (Stanley et al., 1991) and is not cleaved off during secretion. This implies that co-translational secretion is not possible (Delepelaire, 2004). The mechanism of type I secretion was studied in great detail on the basis of the α-hemolysin (HlyA) secretion found in some uropathogenic *Escherichia coli* (UPEC) (Thanabal et al., 1998).

### 1.2.3.2. Type II secretion systems

The type II secretion system (T2SS) is also known as the Sec-dependent system as many proteins that pass through the T2SS must first reach the periplasm via the Sec pathway. Although Sec-dependent translocation is universal (Cao and Saier, 2003), the T2SS is found only in the Gram-negative proteobacteria phylum (Filloux, 2004; Cianciotto, 2005). It is found in species that extend from obligate symbionts (mutualistic, commensal and pathogenic) to free-living species, but is not universal among any particular group. It appears to be a specialized system that promotes functions specific to the interaction of a species with its biotic or abiotic environment. In Gram-negative bacteria, type II secretion is one of five protein secretion systems that permit the export of proteins from within the bacterial cell to the extracellular milieu and/or into target host cells (Nicholas, 2005). The T2SS is required for virulence of the human pathogens *Vibrio cholerae*, *Legionella pneumophila*, and enterotoxigenic *E. coli*, and of the plant pathogens *Ralstonia solanacearum*, *Pectobacterium atrosepticum* (*Erwinia carotovora* subsp. *atroseptica*) and *Xanthomonas campestris* pv. *campestris* (Filloux, 2004; Cianciotto, 2005).
1.2.3.3. Type III secretion system

Type III secretion system (T3SS) is most complex among all protein secretion systems. Its supramolecular structures span the inner membrane (IM), the periplasmic space, the outer membrane (OM), the extracellular space and a host cellular membrane (Daniela and Sheng, 2009). These complex assemblies are structurally and evolutionarily related to the flagella systems. T3SS have been isolated in species of several Gram-negative bacteria (Salmonella, Yersinia, Shigella, Escherichia, Pseudomonas) and were shown to consist at least 20 different subunits which enable these bacteria to translocate substrates (effectors) directly into the cytoplasm of the host cell to exert a broad range of virulence functions (Ghosh, 2004). Because of their shape and their ability to translocate proteins in a cell contact-dependent manner, they are also referred to as ‘injectisomes’ or ‘molecular needles’ (Cornelis, 2006). A characteristic of T3SS is the presence of cognate chaperones, small acidic proteins. These chaperones are considered to stabilize and prevent terminal folding of effector proteins. The energy of ATP hydrolysis by the ATPase is conducted by release of the chaperone from an effector chaperone complex and subsequent loading of the unfolded substrate into the T3SS apparatus (Akeda and Galan, 2005). Another function of the effector-specific chaperones could be the masking of domains needed for membrane targeting within the host cell (Letzelter et al., 2006).

1.2.3.4. Type IV secretion systems (T4SS)

T4SS are characterized by the ability to translocate proteins or complexes of protein and single-stranded DNA. Based on sequence similarities, T4SS are believed to have evolved from bacterial conjugation machineries (Cascales and Christie, 2003). The T-DNA
transfer system of *Agrobacterium tumefaciens* is the prototypical type A T4SS (Burns, 1999). This well-studied T4SS translocates protein–DNA complexes but is somehow distinct from the T4SS of pathogens of humans and animals that appear to translocate proteins only. Again by sequence comparison, T4SS are categorized into two subclasses: type IVA (T4ASS) and type IVB (T4BSS) (Christie and Vogel, 2000). For substrate recruitment and targeting to the inner membrane parts of the transenvelope protein complex, a homohexamer of VirD4 acts as the so-called ‘coupling protein’ (CP) in T4ASS. A stable interaction of the CP with homologs of VirB10, a part of the multi-subunit transenvelope protein complex, was demonstrated (Llosa et al., 2003). The components of the transenvelope complex are members of the mating-pair formation (Mpf) protein family. Different functions could be assigned to sets of VirB proteins which are responsible in formation of the channel traversing the periplasmic space, outer membrane pore (Ward et al., 2002) and pilus formation required in conjugation. Recent work demonstrated the roles of T4ASS in several important human bacterial pathogens (Henderson et al., 2000). Pertussis toxin of *Bordetella pertussis* is secreted in a contact independent manner, while CagA of *Helicobacter pylori* is a translocated T4SS effector protein that induces inflammatory responses and cytoskeletal alterations in the host cell (Henderson et al., 2000). T4ASS have also been identified in *Brucella* spp. and *Bartonella henselae*, and the translocated effectors have central functions in the intracellular lifestyle of these pathogens.

In contrast to the T4ASS, T4BSS are less well understood. One example of a T4BSS is the virulence associated dot/icm machinery of *Legionella pneumophila* (Segal et al., 1998; Vogel et al., 1998). The system was discovered by screening for mutants unable to survive within macrophages (Vogel and Isberg, 1999). Individual mutants were assigned
to dot (defect in organelle trafficking) or icm (intracellular multiplication) according to their respective phenotypes. To date only in *L. pneumophila* substrates of a T4BSS have been described (Segal et al., 2005).

1.2.3.5. Type V secretion systems

The type V secretion system is one of the most recently described pathways permitting the translocation of proteins into the extracellular milieu. Paradoxically, it possesses the simplest secretion apparatus and represents the largest family of protein-translocating outer membrane porins in Gram-negative bacteria (Yen et al., 2002). This system was first described for the IgA1 protease produced by *Neisseria gonorrhoeae* (Pohlner et al., 1987). Type V secretion system (T5SS) includes several mechanisms such as the two-partner system (TPS) and the oligomeric coiled-coil adhesin (Oca) system. A very large number of proteins are secreted via the T5SS, more even than the T2SS, over 500 in the T5aSS class alone (Jacob et al., 2004; Dautin and Bernstein, 2007; Bernstein, 2007). Most of the T5SS secreted proteins characterized to date contribute to the virulence of animal or human pathogens. Proteins secreted via the T5SS include adhesins such as AIDA-I and Ag43 of *E. coli*, Hia of *Haemophilus influenzae*, YadA of *Yersinia enterolitica* and Prn of *Bordetella pertussis*; toxins such as VacA of *Helicobacter pylori*; proteases such as IgA proteases of *Neisseria gonorrhoeae* and *Neisseria meningitides*, SepA of *Shigella flexneri* and PrtS of *Serratia marcescens*; and S-layer proteins such as rOmpB of *Rickettsia* sp. and Hsr of *Helicobacter pylori* (Tsai-Tien et al., 2009).
1.2.3.6. Type VI secretion system (T6SS)

The type VI secretion machinery (T6SS) is a recently characterized secretion system that appears to constitute a phage-tail-spike-like injectisome that has the potential to introduce effector proteins directly into the cytoplasm of host cells (Bingle et al., 2008; Shrivastava and Mande, 2008; Cascales, 2008), analogous to the T3SS and T4SS machineries. Investigation of the virulence-associated secretion cluster in *V. cholerae* led to the identification of type VI secretion system. Type VI secretion system (T6SS) manages the export of substrates at least in the extracellular space without the requirement of hydrophobic N-terminal signal sequences. Using the model host *Dictyostelium discoideum*, a virulence function of the T6SS secreted substrates have been shown, suggesting a translocation of these proteins in the cytosol of the amoebae (Pukatzk et al., 2006). Recently, another T6SS was identified in *P. aeruginosa* where translocation of an ATPase, ClpV1, was identified, presumably energizing the secretion mechanism (Mougous et al., 2006). The T6SS is required for virulence in human and animal pathogens such as *Vibrio cholerae*, *Edwardsiella tarda*, *Pseudomonas aeruginosa*, *Francisella tularensis*, and *Burkholderia mallei*, and also in plant pathogens such as *Agrobacterium tumefaciens*, *Pectobacterium atrosepticum* and *Xanthomonas oryzae* (Wu et al., 2008).

1.2.3.7. Type VII secretion system

Despite of many Gram-positive bacteria have only a single membrane, in certain species of Gram-positive bacteria, most notably the *Mycobacteria*, have a cell wall that is heavily modified by lipids, called a mycomembrane (Abdallah et al., 2007). The genomes of these species have evolved with a novel and specialized secretion systems also called as
Introduction

Type VII secretion systems (T7SS) for the transport of extracellular proteins across their hydrophobic, and highly impermeable, cell wall (Abdallah et al., 2007). The structure and operation of the T7SS and the various proteins involved in the inner membrane translocation channel is yet to be determined (Tsai-Tien et al., 2009).

Although most proteins are translocated via the Sec-pathway, proteins that are already folded in the cytoplasm, for example because they have to incorporate co-factors in cellular compartment, are translocated via the Tat-pathway also called as twin arginine translocation pathway (Berks et al., 1996). The proteins destined to the membrane or extracellular environment by using Tat secretion pathway depends on the characteristic features of the signal peptide and nature of substrate proteins (Driessen et al., 1998; Berks et al., 2000; Driessen et al., 2001; Mori and Ito, 2001; Berks et al., 2003). Though Tat pathway is presently best characterised for *E. coli*, but it was first described for thylakoids which is also designated as the ΔpH pathway in chloroplasts (Mould and Robinson, 1991; Chaddock et al., 1995). Four proteins, TatA, B, C, and E, are involved in Tat-dependent protein secretion in *E. coli*. The detailed mechanism of Sec and Tat pathways were described in the following sections.

1.2.3.8. Structural features - Signal peptide

Signal peptides control the entry of virtually all proteins to the secretory pathway, both in eukaryotes and prokaryotes (Gierasch, 1989; von Heijne, 1990; Rapoport, 1992). The signal peptide of secretory proteins function as both the targeting and recognition signal and ranges in length from 18 to about 30 amino acid residues. It is composed of three domain: the positively charged amino terminus (N region); the nonpolar, hydrophobic
core region (H region); and the more polar cleavage region (C region). Although the amino acid sequences of these domains are not well conserved among the many signal sequences, but they demonstrate similar physical characteristics (Izard and Kendall, 1994). In fact, it is well reported that the shared features facilitate different signal peptides to interact with common elements in a 'general secretory pathway' in *Escherichia coli* (Pugsley, 1993). Sec substrates that are translocated across the hydrophobic membrane possess amino-terminal signal peptides that resemble a trans membrane segment (TM), but they also contain a short positively charged amino-terminus and often a signal peptidase recognition site (Von, 1990). Upon cleavage, the substrate is either released from the membrane or anchored to the lipid bilayer (Fig. 1.5).

*N domain.* The presence of a net positive charge in the N region, introduced by lysine or arginine residues, enhances the processing and translocation rates of a precursor protein but is not essential. Preproteins with signal sequences that carry a neutral or even negatively charged N region can be processed, although at reduced rates (Gennity et al., 1990). With increasing positive charge at the N region of the signal sequence, the SecA requirement for translocation is reduced while the interaction of the preprotein with SecA is enhanced in proteins involved in Sec-dependent translocation pathway (Akita et al., 1990). This suggests that the N region is involved in the targeting of the preprotein to the translocase.
The N region has been suggested to bind the negatively charged surface of the lipid bilayer of the membrane (Phoenix et al., 1993). A reduction in the number of positive charges in the N region results in inefficient interaction with the membrane; this phenomenon can be compensated for by an increased hydrophobicity of the H region (Phoenix et al., 1993). Strikingly, this is accompanied by a restoration of the translocation defect, suggesting that the interaction of the signal sequence domain with the membrane is an important step in targeting and/or translocation.

**N-domain – role in membrane topology**

The positive charges in the N region are thought to orient the signal sequence of secretory proteins or the stop-transfer signal of membrane proteins correctly within the
lipid bilayer. The transmembrane electrical potential (Dc, inside negative) prevents the translocation of positively charged residues and facilitates that of negatively charged residues (Anderson and Heijne, 1994). In this way, the Dc would contribute to the realization of the correct topology of integral membrane proteins, which obey the “positive-inside rule” (Von, 1990). On the other hand, in acidophilic bacteria and archaea, the Dc has a reversed polarity, i.e., inside positive versus outside. The topology of the inner membrane proteins of these bacteria also obeys the positive-inside rule (Vossenberg et al., 1998), whereas the signal sequences of the secretory proteins identified thus far are indistinguishable from those of neutro- or alkalophiles. This challenges the electrophoretic mechanism and suggests that Dc may affect protein translocation and membrane insertion by another mechanism that would involve the translocation apparatus in a more direct manner.

**H domain**

The H domain is the hydrophobic core of a signal sequence and varies in length from 7 to 15 amino acids. It is the most important part of the signal sequence; this is best illustrated by the observation that an increase in the total hydrophobicity of this domain can overcome mutations in the other regions of the signal sequence (Phoenix et al., 1993; MacFarlane and Muller, 1995). To some extent, the total hydrophobicity of the H region determines the efficiency of translocation and the translocation efficiency increases with the length and hydrophobicity of the H region (Chou and Kendall, 1990). This relation is sigmoidal, and a minimum hydrophobicity is required for translocation (Doud and Kendall, 1993). The residues in the H region are responsible for the α-helical conformation which
extends from the N region. Frequently, a so-called helix breaker, i.e., a glycyl or prolyl residue, is found in the middle of the H region. This may allow the signal sequence to form a hairpin-like structure that can insert into the lipid bilayer. According to the unlooping model, the signal sequence inserts into the membrane by extension through unlooping of this hairpin (Shinde et al., 1989; Vrije et al., 1990). Indeed, nuclear magnetic resonance spectroscopy studies have shown that in a membranous environment, signal sequences may adopt a two-domain conformation consisting of an amino-terminal α-helix and a more flexible carboxy-terminal domain (Rizo et al., 1993; Wang et al., 1993; Chupin et al., 1995). In addition, when two cysteines are introduced into the signal sequence, translocation under oxidized conditions is hampered (Nouwen et al., 1994), indicating that the formation of the loop prevents translocation. Unlooping would be facilitated by the Dc to orient the signal sequence within the electrical field. In this respect, some replacements of the α-helix breakers with α-helix-forming residues render preprotein translocation less dependent on the proton motive force (Nouwen et al., 1996), although the effects are rather subtle. The unlooping model does not explain how the signal sequence inserts into the membrane while bound to SecA. It is more likely that both SecA and the signal sequence insert simultaneously at the translocation site upon the binding of ATP.

**C domain** The leader peptidase cleavage site (C domain) is the only part of the signal sequence that demands some primary sequence specificity. Two types of leader peptidases are known, type I, serving ordinary preproteins, and type II, cleaving the leaders of lipoproteins. For the signal sequences that rely on type I leader peptidases, the limitations are on the residues located at positions -1 and -3 relative to the start of the mature part (Von Heijne, 1984). This domain interacts with the leader peptidase which cleaves off the signal
sequence (Dalbey and Heijne, 1992). Usually, these residues have small neutral side chains, such as alanine, glycine, serine, and threonine (Von Heijne, 1984). Lipoproteins depend on type II leader peptidases, and the demands differ from those for type I peptidases only at the -3 and +1 positions (Sankaran and Wu, 1994). Precursors of lipoproteins contain larger hydrophobic amino acid residues at the -3 position, with a preference for leucine (Fekkes and Driessen, 1999). At the +1 position, a cysteine is always present and has to undergo modification prior to processing. After the signal sequence has been removed, it is degraded completely by a number of peptidases (Heijne, 1990).

1.2.3.9. Classification of Signal peptides

Based on the structural features the signal peptides have been classified into two major groups. These are designated as Sec signal peptides and Tat signal peptides (Pohlschroder, 2005).

1.2.3.9.1. Sec signal peptides

Though there are minor structural differences among Sec signal peptides, these are found mainly in proteins that are destined to target / translocate across the membrane in denatured form. Upon targeting / translocating they become active by acquiring native confirmation. However due to existence of minor differences they have been again classified into three sub groups.

- **Class 1 signal peptides** contain a class I signal peptidase (SPI) processing site following the hydrophobic stretch. These signal peptides are usually 18±35 amino
acids long and do not contain a strict consensus sequence. They share, however, a tripartite structure. The amino-terminal N domain, which is usually two to eight residues, contains one or more positively charged residues. The N-domain is important for interaction with the protein translocation machinery (Akita et al., 1990) and negatively charged lipid head groups on the cytoplasmic face of the lipid bilayer (de Vrije et al., 1990; Phoenix et al., 1993). The hydrophobic H-domain that follows the N-domain varies in length from 8 to 15 residues. This region has been proposed to form an α-helical conformation in the membrane (Briggs et al., 1986). The third domain (C domain) of the signal peptide contains the cleavage site for signal peptidase (SPase). The residues at positions -3 and -1 (relative to the start of the mature protein) are usually those with small neutral side chains, such as alanine, glycine, serine and threonine (von Heijne, 1984). Substrates processed by this universally conserved signal peptidase are either released from the membrane or anchored to the lipid bilayer by way of a carboxy-terminal membrane anchor (Pohlschroder et al., 2005).

- **Class 2 signal peptides** are found in bacterial lipoproteins. These signal peptides, which contain similar N and H-domains to Class I signal peptides, are characterized by a conserved lipobox in the C-domain with the consensus sequence L(A/S)(G/A)C (von Heijne, 1989). The invariable cysteine in this lipobox is lipid-modified by a diacylglyceryl transferase and becomes the first residue of the mature protein after cleavage by a lipoprotein-specific type II SPase. Due to the lipid-modified cysteine, the protein remains anchored to the cytoplasmic membrane.
These lipoproteins attach to the cytoplasm or outer membrane by way of a lipid-modified invariant amino-terminal cysteine (Pohlschroder et al., 2005).

- **Class 3 signal peptides** initially identified in bacterial type IV prepilin subunits, contain a prepilin peptidase processing site that precedes the hydrophobic stretch. These signal peptides also have a positively charged N-domain and a hydrophobic H-domain. In contrast to other signal peptides, prepilin signal peptides are cleaved just before the H-domain by a specific SPase that has its active site on the cytoplasmic face of the membrane (Lory, 1994). These are distinct from class 1 and class 2 signal peptides, this hydrophobic stretch remains part of the mature protein and is essential for initial membrane anchoring as well as for subunit–subunit interactions that are crucial for the biosynthesis of type IV pili and pilus like structures (Pohlschroder et al., 2005).

**1.2.3.9.2. Tat signal peptides**

As invariant twin arginines are found in the H region, they are designated as Tat (twin arginine transport) signal peptides. These twin arginines are part of a consensus sequence of tat motif (S/T)R\(x\)ØØ, where \(x\) is any residue (except polar residue or proline), and Ø is a hydrophobic residue (Berks, 1996; Cristobal et al., 1999).

Substrates of the twin arginine transport (Tat) protein translocation system possess a signal sequence with n-, h-, and c-regions like those of Sec pathway proteins and shows several distinct Tat system-specific features. The most important feature is the presence of a twin arginine motif, which is part of a larger conserved sequence (S/T)-R-R-Ø-Ø (where
x is any amino acid and ⋄ is a hydrophobic amino acid) found at the n-region / h-region boundary (Cristobal et al., 1999; Berks et al. 2000). In Gram-negative Bacteria, this sequence is usually (S/T-R-R-x-F-L-K) and these signal peptides are longer, on average, by 14 amino acid residues than Sec signal peptides (Bardy et al., 2003).

Based on signal peptide cleavage by signal peptidases, Tat signal peptides were classified into two groups (Bolhuis, 2002; Bardy et al., 2003; Pohlschroder et al., 2004; Pohlschroder et al., 2005). **Class I Tat signal peptides** shows similar features of typical Tat signal peptide having n-terminal, h-hydrophobic and c-terminal region which possess highly conserved tat motif and a signal peptidase I cleavage site. Whereas **Class II Tat signal peptides** also possess similar features as that of the class I except that it contains LIPO box at the c-terminal region (Bolhuis, 2002; Albers and Driessen, 2002; Bardy et al., 2003; Albers et al., 2004) and reports suggesting that some of these LIPO-box-containing archaeal proteins have lipid modifications at the conserved terminal cysteines (Kokoeva et al., 2002).

### 1.2.4. Protein translocation - Evolution

The cellular membrane is a basic feature that separates the interior of the cell from its external environment. This membrane must combine two antagonistic features:

(a) It must guarantee the maintenance of both the composition and concentration of cytoplasmic molecules; and

(b) it must allow for the controlled exchange of matter and information between the cell and its environment.
This functional paradox is solved by combining two different types of molecules, amphiphilic lipids and proteins. The lipids provide a flexible and impermeable barrier, and the introduction of proteins in the form of receptors and channels enables the cell to transduce information and to regulate the transport of molecules. Thus, mechanisms allowing the insertion of proteins into the cellular membrane are a prerequisite for the origin of cells. This need for transport was no longer limited to smaller molecules, such as signaling molecules, catabolites, and a need to secrete large, enzymatically active proteins, enabling primitive cells to alter their external environments. As a result, protein export mechanisms evolved that allowed cells to digest organic matter and form extracytoplasmic protective structures, such as the cell wall and outer membrane (Pohlschroder et al., 2005). The earliest mechanism of protein translocation likely involved the spontaneous insertion of membrane proteins.

The next translocation system evolved is the YidC/Oxa1/Alb3 membrane integration system which is present in prokaryotic domains and is relatively simplistic and requires only the YidC protein to facilitate substrate translocation. But it alone is incapable of transferring large domains across the membrane. The other candidate is the universally conserved Sec system, which is minimally composed of SecY and SecE proteins. The Sec pathway facilitates protein integration into and translocation across membranes. Final and the recently evolved translocation system is Tat (twin arginine translocation) system and most likely evolved after the Sec pathway, since Tat signal sequences appear to be Sec-like and hence may have directly evolved from preexisting Sec signals (Pohlschroder et al., 2005). A theoretical timeline of protein translocation evolution was shown in Fig. 1. 6.
1.2.4.1. Sec-dependent protein translocation

In cells from all three domains of life, more than one third of the proteome is secreted across, or inserted into, biological membranes. Secretory proteins include hydrolytic enzymes, periplasmic lipoproteins, toxins and surface appendages such as pilli and flagella. Integral membrane proteins mediate selective transport, energy conversion, cell division, extracellular signal sensing, and membrane and cell-wall biogenesis. To successfully localize polypeptides extracytoplasmically, the cell must tackle five daunting tasks (Effrosyni et al., 2007).

1. Discriminate the cytoplasmic-resident proteins from those that are destined for export.
2. Deal with the inherent tendency of polypeptides to fold rapidly.
3. Target exported proteins to the membrane with specificity and fidelity.
4. Achieve transmembrane crossing of these elongated, heteropolymeric substrates, which are several times as long as the membrane is wide.
5. Manage a second sorting event that releases membrane proteins into the lipid bilayer and secretory proteins to the *trans* side of the membrane.

The Sec pathway is the only known universally conserved protein translocation pathway. Through the course of their evolutionary divergence, the three domains of life have reached disparate solutions to the fundamental problem of translocating proteins in an unfolded fashion to extracytoplasmic destinations via the Sec pathway (Pohlschroder et al., 2005).

**1.2.4.2.1. Structural features of the Sec translocase**

In all domains of life the bulk of the structure of the translocation channel comprises two universally conserved proteins, Sec61α and Sec61γ, in eukaryotes, with the corresponding proteins in the prokaryotic domains designated SecY and SecE, respectively (Hartmann et al., 1994). In addition to these Sec subunits, the pore complexe contain a third subunit that differs between domains is Sec61β, is conserved in eukaryotes and archaea, and a distinct protein, SecG, completes the heterotrimeric complex in bacteria (Van et al., 2004). Translocase consists of a peripheral ATPase, SecA, and at least five integral membrane proteins (i.e. SecY, SecE, SecG, SecD and SecF). As shown in Fig. 1.7 this large complex most likely forms the translocation pore along which the preproteins pass the cytoplasmic membrane in an unfolded form (Eli et al., 2006).

![Fig. 1.7](image)

A schematic representation of the bacterial pre-protein translocase subunits (Effrosyni et al., 2007 *Nature Reviews Microbiology*).
1.2.4.2.2. The preprotein–SecB complex

SecB is a general cellular chaperone that appears to be important for the export of several secretory proteins, such as the outer-membrane porin lamb (Randall et al., 2002). It is not essential for viability and is not present in all bacteria, but has a dual role: it maintains secretory pre-proteins in a translocation-competent state and interacts specifically with membrane-bound SecA. SecB crystal structures (Xu, Z. et al., 2000; Zhou et al., 2003) revealed the structural features that could allow this chaperone to bind to extended polypeptides and SecA. SecB is a stable tetramer that probably binds to pre-proteins by recognizing exposed hydrophobic surfaces (Randall et al., 1998). A groove has been identified in the structures that could constitute the pre-protein binding site. Each SecB4 contains two such grooves, which, as these sites are mainly solvent exposed, could accommodate peptides without disturbing the stable tetramer (Fig. 1. 8c, pink cylinder). The grooves are lined by several aromatic residues at one end and hydrophobic residues at the opposing end (Fig. 1. 8c, part 2). The aromatic-residue-rich sub-site of SecB could accommodate the proposed SecB-binding motif in pre-proteins, which, it is suggested, is a nonapeptide that contains several aromatic and basic residues (Knoblauch et al., 1999). In an acidic region on the top surface of the SecB tetramer, several residues interact with the 25 C-terminal residues of the SecA C-terminal domain. (Fekkes et al., 1997; Woodbury et a., 2000) (Fig. 1. 8b). The structure of SecB and its interaction with SecA was shown in Fig. 1. 8.
Fig. 1. 8

The structure of SecB. (A) A ribbon Nature Reviews representation of the *Escherichia coli* SecB (ecSecB) tetramer, which is a dimer of dimmers (Dekker et al., 2003) (Protein Data Bank code: 1QYN). (B) The SecB–SecA interaction. A ribbon and space-filling model representation of the ecSecB tetramer (grey) (Dekker et al., 2003) in complex with the carboxy (C)-terminal peptide of *Haemophilus influenzae* SecA (yellow) (Xu et al., 2000; Zhou et al., 2003) is shown. (C) The SecB–pre-protein complex. A ribbon representation of the ecSecB4 tetramer is shown (Effrosyni et al., 2007. *Nature Reviews Microbiology*).

Several routes for how the polypeptide wraps its extended aminoacyl chain around SecB have been proposed, as would be expected for a chaperone that binds to numerous diverse substrates. SecB binds the preprotein at its mature domain (Randall and Hardy, 1995). The signal sequence provides no positive contribution to the binding energy or binding affinity of the interaction of the preprotein with SecB (Randall et al., 1998), but the signal sequence slows the folding of the mature domain (Park et al., 1988; Hardy and Randall, 1991). It has been suggested that this slowing process allows SecB to discriminate between the precursor proteins and other proteins in the cell, as formulated in the kinetic
partitioning model (Hardy and Randall, 1991). In this model, the cytosolic proteins would escape stable interaction by folding more rapidly than precursor proteins do. The final distribution of the precursor protein among the different pathways in the cell is then determined by partitioning that is dependent on the rate of folding or aggregation relative to the rate of binding to the chaperone (Randall and Hardy, 1986). Indeed, when the folding rate of the SecB-dependent precursors of maltose-binding protein (MBP) and galactose-binding protein (GBP) was increased, the amount of unbound ligand increased as well (Topping and Randall, 1997). However, several other observations are against the folding rate as the determining factor for SecB selectivity.

1.2.4.2.3. The Sec YEG complex

The translocase core consists of the integral membrane SecYEG heterotrimer. SecYEG retains limited but detectable sequence homology from bacteria to humans, and is remarkable in its ability to translocate proteins not only through, but also laterally into, the membrane (BOX 2), keeping the membrane barrier intact throughout these processes. SecYEG and SecA form the active holoenzyme (Brundage et al., 1990). SecA binds to the membrane with low affinity at acidic phospholipids but with high affinity (5–40 nM) at SecYEG (Hartl et al., 1990). Several cytoplasm-exposed loops of SecY are available for a possible interaction with SecA and/or the ribosome (Breyton et al., 2002; Van et al., 2004; Mitra et al., 2005; Bostina et al., 2005; Mori et al., 2006). SecYEG makes three connections with the ribosome — two between ribosomal-RNA hairpins and the cytoplasmic SecY loops and one between ribosomal proteins and the cytoplasmic region of SecG (Mitra et al., 2005). SecG, although not essential for the initial binding of SecA to
SecYEG, might interact at a later stage of the translocation reaction (Nishiyama et al., 1995; Nagamori et al., 2002). Pre-proteins cross the membrane, and a flexible molecular motor, the SecA ATPase, drives translocation at the expense of metabolic energy in the form of ATP and the protonmotive force (PMF). Other cytoplasmic and membrane subunits optimize the translocation reaction.

1.2.4.2.4. Protein sorting and targeting

Nascent pre-proteins are recognized directly by piloting factors, such as the ribonucleo protein signal-recognition particle (SrP) (Luirink and Sinning, 2004) or the SecB chaperone (Randall and Hardy, 2002; Ullers et al., 2004). Despite the fact that in bacteria the SrP is predominantly involved in the targeting of inner-membrane proteins (BOX 2), there are examples of long, strongly hydrophobic signal peptides of nascent secretory proteins, which are probably preferentially bound by the SrP (Schierle et al., 2003; Sijbrandi et al., 2003). Other signal peptides delay pre-protein folding and thus allow SecB to bind to the mature region of the pre-protein. This process can occur while the polypeptide chain is still being synthesized, but it is not mechanistically coupled to elongation (Randall, 1983). In both cases, the resulting SrP–pre-protein and SecB–pre-protein complexes are targeted to the translocase at the membrane. For SrP, this is achieved by docking to its membrane receptor FtsY3, and for SecB, by docking to the SecA subunit of the translocase (Hartl et al., 1990). Although SecA does not contribute to the SrP targeting route (Scotti et al., 1999), when long, hydrophilic segments are encountered, SecA is recruited to catalyse their export (Neumann et al., 2000).
1.2.4.2.5. Translocation

In the presence of preprotein, SecB binds SecA with higher affinity (Fekkes et al., 1997; Hartl et al., 1990). Only the signal sequence is needed to stimulate SecA for this high-affinity SecB recognition (Fekkes et al., 1998). Upon interaction of SecB with SecA, SecB releases the precursor protein, which is subsequently transferred to SecA (Fekkes et al., 1997; Fekkes et al., 1998). This reaction involves binding of the signal sequence of the preprotein to SecA. Upon binding SecB, SecA might lower the affinity of SecB for the preprotein by modulating the conformation of preprotein binding site of SecB. This site is located at the opposite face of the putative b strand that constitutes the SecA-binding site. According to this scenario, the SecA-bound SecB will be unable to interact with a new preprotein, and therefore the binding reaction precludes other proteins from entering the translocation pathway at a site that is already occupied. Only upon initiation of translocation, i.e., after the binding of ATP to SecA, SecB released from the membrane to bind a new preprotein in the cytosol (Fekkes et al., 1997). And finally pre-proteins are converted into mature proteins once after reaching the trans side of the membrane upon cleavage of signal peptide by signal peptidase (Paetzel et al., 2002; Mogensen and Otzen, 2005). A schematic representation of general Secretory pathway in *E. coli* was shown in Fig. 1. 9.
1.2.4.2. Twin Arginine Translocation (Tat) pathway

The discovery of the Tat pathway dates to the early 1990s, when it was noticed that a subset of polypeptides in chloroplasts could be translocated independently of ATP hydrolysis and instead depend exclusively on the proton gradient (Mould and Robinson, 1991; Cline et al., 1992). For this reason, it was initially designated the ∆pH pathway but more recently has been termed the cpTat pathway (chloroplast Tat/∆pH pathway). In 1995 Creighton and his coworkers (Creighton et al., 1995) presented the first evidence that the cpTat pathway enables the translocation of prefolded proteins. Shortly thereafter, Berks (Berks, 1996) observed that a group of bacterial periplasmic proteins containing various cofactors share a unique type of signal peptide containing a consensus “double arginine” motif also found in substrates of the chloroplast pathway. The existence of a bacterial pathway analogous to the one in chloroplasts was thus established; it was initially termed...
Introduction

The bacterial cytoplasmic membrane and the chloroplast thylakoid membrane have a common evolutionary origin and contain equivalent protein transport pathways. The Sec pathway translocates proteins as unstructured chains (Rapoport, 2007), whereas the twin-arginine translocation (Tat) system is dedicated to transporting folded proteins (Berks et al., 2003; Berks et al., 2005; Cline and Theg, 2007). In both cases substrates must be translocated without compromising the membrane permeability barrier. This translocation is particularly challenging for the Tat apparatus because folded proteins are larger and more variable in size than the linear peptides transported by the Sec system.

**Fig. 1. 10**

Features of a typical Tat signal peptide, ssTorA from *E. coli* (Philip et al., 2006. *Annual Reviews of Microbiology*).

The mechanism of Tat transport has not been established but appears unrelated to that of other membrane transporters (Leake et al., 2008). As shown in Fig. 1. 10 another unique feature of the Tat system is its substrates possess a conserved SRRxFLK “twin-
arginine” motif in their N-terminal signal sequences in which the twin arginines are highly invariant in nature (Berks et al., 1996).

1.2.4.2.1. Organization and structural features of tat genes

The first tat gene to be characterized was the maize hcf106 gene, following the characterization of a mutant line (Voelker and Barkan, 1995) defective in the targeting of a subset of thylakoid lumen proteins. It was well established that proteins were transported across the thylakoid membrane by two distinct pathways, one of which was related to bacterial Sec-type systems. The discovery of a chloroplastic Sec pathway was not unexpected given that these organelles evolved from endosymbiotic cyanobacteria, but the sequencing of the hcf106 gene (Settles, 1997) provided the first structural information on the Sec-independent pathway used to transport other thylakoid lumen proteins. As expected, this gene was found to encode a novel component of the thylakoid protein transport apparatus. Most significantly, however, the hcf106 gene was also found to be homologous to many open reading frames in bacterial genomes—the first concrete indication of a mainstream, Sec-independent bacterial protein export system. The first bacterial tat mutants were isolated in 1998 (Weiner et al., 1998; Sargent et al., 1998; Bogsch et al., 1998) and this field is just evolved when compared with most of the other protein transport systems.

The first tatABC operon was identified in E. coli since the initial isolation of tat mutants were discovered in this organism (Colin and Albert, 2004). As shown in Table 1.1 the minimal integral membrane protein components TatA, TatB, and TatC were identified in E. coli.
Table 1. Components of the TAT pathway

<table>
<thead>
<tr>
<th>Protein</th>
<th>Predicted size (kDa)</th>
<th>Characteristic</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TatA</td>
<td>9.6 or 11.3 (a)</td>
<td>60% Homologous to TatE; its expression is higher than other tat genes</td>
<td>Gouffi et al., 2004; Porecelli et al., 2002; Sargent et al., 1998.</td>
</tr>
<tr>
<td>TatB</td>
<td>18.4 (a)</td>
<td>Complexes with TatC and prevents its degradation</td>
<td>Sargent et al., 1999</td>
</tr>
<tr>
<td>TatC</td>
<td>28.9 (a)</td>
<td>Likely to be a signal peptide binding component</td>
<td>Allen et al., 2002; Behrendt et al., 2004; Buchanan et al., 2002.</td>
</tr>
<tr>
<td>TatD</td>
<td>29.5 (a)</td>
<td>No effect on protein translocation; presents DNase activity</td>
<td>Wexler et al., 2000.</td>
</tr>
<tr>
<td>TatE</td>
<td>6.9 (b)</td>
<td>Can partially substitute TatA</td>
<td>Berks et al., 2000.</td>
</tr>
</tbody>
</table>

(a) Two possible translation initiation sites exist for tatA (Sargent et al., 1998).
(b) DNA sequence retrieved from Genbank accession no. NP_308692. Molecular mass calculated by the program PROTPARAM (http://www.expasy.ch/tools/protparam.html).

The Tat components are required for the twin arginine translocation pathway to transport prefolded proteins across the biological membrane and organization of tat genes in *E. coli*, *Bacillus subtilis* and *Pseudomonas* sp. were depicted in Fig. 1.11. (Weiner et al., 1998; Sargent et al., 1998; Bogsch et al., 1998; Sargent et al., 1999; Jongbloed et al., 2000; Jan Maarten et al., 2002). The first three genes belong to the same operon, while tatE is located elsewhere on the bacterial chromosome (Bogsch et al., 1998; Sargent et al., 1998b; Weiner et al., 1998).
Chromosomal organisation of the tat genes in *E. coli*, *B. subtilis* and *Pseudomonas sp.* **A]** The tatABCD genes of *E. coli* are organised in an operon with the tatE gene located elsewhere on the chromosome. **B]** The genes for TatAd and TatCd that make up the TatAdCd translocase are located downstream from the *phoD* gene encoding the TatAdCdB-dependent phosphodiesterase D. The genes for the second Tat translocase, tatAy and tatCy, are similarly organised in an operon elsewhere on the chromosome. The tatAc gene is monocistronic and does not have a tatC counterpart. This figure is adapted from Jongbloed et al., 2000). **C]** Physical maps of the DNA regions carrying tat genes in *Pseudomonas aeruginosa* and *Pseudomonas stutzeri*. Transcriptional directions are indicated by the open arrows (Jan Maarten et al., 2002 *Current Opinion in Microbiology*).

The *tatA* and *tatB* encode small proteins of 9.6 and 18.4 kDa, respectively, each of which contains a single transmembrane (TM) span (Sargent et al., 1998; Bolhuis et al.,
2001; Porcelli et al., 2002). TatC is a 28.9-kDa protein that was predicted to contain 6 trans membrane (TM) spans (Bogsch et al., 1998) (Fig. 1. 12).

**Fig. 1. 12**

![The predicted structure and topology of the *E. coli* Tat components. Predicted helical regions are shown as boxes (Philip et al., 2006. *Annual Reviews of Microbiology*)](image)

Deletion of either *tatB* or *tatC* leads to a complete loss of Tat-dependent protein export (Sargent et al., 1998; Bogsch et al., 1998), whereas the Δ*tatA* mutant can support a minimal level of export activity. This is due to the presence of fourth *tat* gene in *E. coli*, *tatE*, which encodes a TatA paralogue that is expressed at very low levels. Disruption of both *tatA* and *tatE* blocks export by the Tat pathway, but overexpression of *tatE* complements the Δ*tatA* mutant (Sargent et al., 1998; Sargent et al., 1999), confirming a similar basic function. However, the evolutionary rationale for maintaining both genes in *E. coli* is unclear since the TatE protein is expressed at such very low levels, and many Gram-negative bacteria possess only one *tatA*-type gene (Yen et al., 2002). As shown in the Fig. 1. 11 the *tat* operon encodes a fourth gene, designated *tatD*, but studies have shown this gene to encode a soluble protein with DNase activity that plays no apparent role in Tat-dependent translocation (Wexler et al., 2000).
TatA

The TatA protein is the most abundant of the Tat components and is estimated to be present at around 20 times more than TatB and TatC (Jack et al., 2001; Sargent et al., 2001). The TatA protein is predicted to have a structure similar to TatB, with an N-terminal trans membrane (TM) domain followed by a short hinge region leading to an amphipathic helical region and an unstructured soluble C-terminal tail (Fig. 1.13). The *E. coli* tatA gene encodes a 9.6-kDa polypeptide of 89 amino acid residues. Structural predictions (Fig. 1.13) suggest an N-terminal hydrophobic α-helix to residue 20, followed by a short hinge region and a longer amphipathic α-helix extending to amino acid 42. TatA is thought to function at a late stage in translocation and likely forms the major component of the Tat pore itself (Alami et al., 2003; Gohlke et al., 2005).

The TM domain of TatA is critical for export and is important for interactions with TatB (Lee et al., 2002; Dabney et al., 2003; Barrett and Robinson, 2005). The *tatE* gene encodes a protein of 67 residues with a predicted mass of 7.0 kDa. It shares almost 60% amino acid identity with TatA along its entire length and can functionally substitute for TatA. In keeping with the high percentage of amino acid identity, the positions of the predicted hydrophobic and amphipathic α-helical regions in TatE are indistinguishable from those in TatA (Fig. 1.13).
Philip et al., identified that although the C-terminal domain of TatA protein is not strictly necessary for its activity, but the transmembrane and adjacent helical regions play a critical role in its function (Philip et al., 2006). Both proteins have C-terminal regions that are predicted to be mainly random coil. Because of the high sequence and structural similarity between the two proteins, coupled with the previous studies that TatE is poorly expressed but functionally equivalent to TatA (Sargent et al., 1998; Sargent et al., 1999; Jack et al., 2001). In *E. coli* TatA is present at an approximately 20-fold molar excess over the other essential Tat components, TatB and TatC (Sargent et al., 2001; Jack et al., 2001). Chemical cross-linking studies have shown that TatA forms at least tetrameric homooligomers in the cytoplasmic membrane (Leeuw et al., 2001), while purification experiments suggest that TatA associates with TatB and TatC in at least two types of large (approximately 600 kDa) macromolecular complexes (Sargent et al., 2001; Bolhuis et al., 2001).
The oligomeric state of TatA in Tat translocation

The TatA protein of *E. coli* is an 89-amino acid membrane protein. The secondary structural elements of this protein were predicted (Fig. 1. 13) using the program PSIPRED (Jones, 1999). TatA is strongly predicted to contain a transmembrane α-helix at the amino terminus (helix α-1) followed by a polypeptide segment containing a basic amphipathic α-helix (helix α-2). These predicted helical regions exhibit a high level of sequence conservation between species. However, only a glycine residue between the predicted helices is invariant (Fig. 1. 13) (Settles et al., 1997; Berks et al., 2000).

Detergent-solubilized TatA is found as large, homooligomeric complexes of variable size (Porcelli et al., 2002; Oates et al., 2005; Gohlke et al., 2005; McDevitt et al., 2006). Low resolution structures of purified TatA show doughnut-shaped particles with an internal cavity large enough to accommodate substrates (Gohlke et al., 2005). This and other evidence (Cline and Mori, 2001; Mori and Cline, 2002) suggest that TatA forms the translocating element of the Tat system. The number of chemical cross-links that can be formed between TatA molecules increases under transport conditions (Dabney et al., 2006). These observations have led to the suggestion that substrate interaction with TatBC triggers TatA polymerization (Dabney et al., 2006; Cline and Theg, 2007).

Since cross-linking studies are unable to define the stoichiometry and stoichiometry distribution of the TatA complexes that are present in the membrane, Leake et al., from University of Oxford have used “*in vivo* single-molecule imaging” technique to visualize individual yellow fluorescent protein-labeled TatA (TatA-YFP) complexes expressed at native levels in living *E. coli* cells by using fluorescence microscopy to determine TatA
Introduction

Stoichiometry (Leake et al., 2008). Their experimental studies also suggest that the TatA-YFP complexes are made up of tetrameric units (4.3 ± 0.6 molecules) of TatA molecules and forms a ring model of translocating pore (Fig. 1.14). Since it was previously shown that TatA interaction with substrate-bound TatBC complexes requires the transmembrane proton motive force (pmf) (Mori and Cline, 2002; Alami et al., 2003) and does substrate induced structural reorganization of TatA (Dabney et al., 2006), Leake et al., extended their studies and determined the role of pmf on stiochiometry of the TatA complex and its structural features in the membrane by treating the cells with protonophore carbonyl cyanide 4-(trifluoromethoxy) phenylhydrazone (FCCP) which will abolish the transmembrane pmf in the membrane (Leake et al., 2008) and determined that the oligomeric state of TatA is unaffected by the removal of the pmf shows that pmf is not required to maintain the polymerized state of TatA.

Fig. 1.14

Schematic illustration of the ring model of TatA complex (Leake et al., 2008 Proceeding of the Natural Academy of Sciences)

In addition, the substrate-bound TatBC complex triggers the TatA and TatA polymerizes around the substrate molecule to produce a pore that matches the size of the substrate and this observation is in accordance with different-sized TatA oligomers were
determined in membrane using YFP labeled TatA molecules by applying in vivo single molecule imaging analysis technique using fluorescence microscopy (Leake et al., 2008).

**Topological organization of TatA**

Previous observations based on electron microscopy studies revealed that TatA has domains on either side of the membrane (Sargent et al., 2001). To test this hypothesis, Porcelli et al., attempted to determine the topological organization of TatA using protease accessibility measurements by isolating spheroplasts and the inner membrane vesicles which were prepared from cells expressing TatA. These membrane preparations were subjected to proteinase K treatment, and the fate of the TatA molecules present was assessed by immunoblotting.

The TatA in spheroplasts was insensitive to proteinase K treatment and suggests that TatA is not accessible from the periplasmic side of the membrane. Control experiments demonstrate that permeabilization of the spheroplast membrane with the detergent Triton X-100 renders TatA susceptible to proteolysis by proteinase K. These observations suggest that TatA is exposed at the cytoplasmic face of the membrane (Porcelli et al., 2002) and obeying the Positive-Inside rule (von Heijne, 1992). Recently Philip et al., obtained a higher-resolution three-dimensional image of a Tat complex using random canonical tilt electron microscopy (Fig. 1. 15) which also obeys the positive-Inside rule (Philip et al., 2006).
Fig. 1. 15

Three-dimensional density maps of TatA complexes. (a) TatA complexes viewed from the closed end of the channel proposed to be at the cytoplasmic side of the membrane (C-face). (b) TatA complexes viewed from the open end of the channel proposed to be at the periplasmic side of the membrane (P-face). (c) Side views of TatA. The front half of each molecule has been cut away to reveal internal features. (d) Views of TatA parallel to the membrane plane. The proposed position of the lipid bilayer is indicated in gray. (Scale bar, 100 Å.) (Philip et al., 2006. *Annual Reviews of Microbiology*)

**TatB**

The *E. coli* tatB gene encodes a protein of 171 amino acids with a predicted molecular mass of 18.4 kDa. Structural prediction analysis of TatB (Fig. 1. 16) indicates an N-terminal hydrophobic α-helix, similar to those of TatA and TatE, extending to residue 20. This is followed by a helical region extending to amino acid 81, of which at least the first section is predicted to be strongly amphipathic. The C-terminal portion is predicted to be mainly random coil.
Fig. 1. 16

Structural prediction for TatB. α-Helical regions are represented as cylinders, and β-sheets are represented as arrows. The transmembrane α-helical domains of TatB are black, and the strongly amphipathic α-helical domains are white. (Philip et al., 2006 *Journal of Bacteriology*).

TatB shares some sequence homology with TatA/E (approximately 25% identity at the amino acid level) but is considerably longer. TatB is an essential component of the *E. coli* Tat machinery. Thus, an in-frame deletion of the tatB gene leads to a complete block in the twin arginine translocation protein export (Philip et al., 2006). But this is exceptional in the case of *Bacillus subtilis* where it doesn’t contain any tatB genes (Fig. 1.11) and, *tatA* acts as bifunctional which is able to carry out the function of both *E. coli* TatA and TatB (Jongbloed et al., 2006; Barnett et al., 2008).

TatB was shown *in vitro* to contact the entire length of the signal peptide and also the mature protein more than 20 residues away from the signal peptide cleavage site (Alami et al., 2003). Such interactions were seen only when TatC was present, suggesting that substrate targeting involves a series of sequential interactions, with TatC forming the primary recognition site before the substrate is transferred to TatB. In turn TatB could be considered a mediator between TatC and TatA, contacting the substrate after initial recognition by TatC and then potentially involved in transfer to a complex consisting mainly of the TatA protein proceeding to translocation (Sargent et al., 2001; Alami et al., 2003).
**TatC**

TatC is an essential, largest and most highly conserved component of the Tat system in bacteria and chloroplasts (Bogsch et al., 1998). The *E. coli* TatC is a 258 amino-acid polytopic membrane protein with six Trans Membrane Domains (TMD) with the amino and carboxyl termini located at the cytoplasmic face of the membrane (Behrendt et al., 2004; Ki et al., 2004). Chloroplast TatC (cpTatC) has been demonstrated to bind the twin-arginine motif of Tat signal sequences, and the signal sequence can remain bound to TatC during transport without affecting the functionality of the Tat system (Gerard et al., 2006). Similar to the plant system, bacterial TatC binds the region of the twin-arginine motif of Tat substrates (Alami et al., 2003). Important residues have been identified in TatC, but so far the exact functions of these positions are unknown (Allen et al., 2002; Buchanan et al., 2002; Barrett et al., 2005).

*E. coli* TatC has several weakly conserved protonatable residues in its TMDs. At the position E170 also glutamine residues can be found at the E227 position, tryptophan and phenyl alanine residues are present in orthologs of some phyla. In *E. coli*, the TatC E170A and E227A mutants did not turn into the functional Tat system (Buchanan et al., 2002). The mutation of a more conserved glutamate residue (E103) in cytoplasmic loop of TatC was reported to inactivate the Tat system (Buchanan et al., 2002). However, in other assay systems, this mutation did not affect transport (Barrett et al., 2005). One conserved aspartate (D211) has been reported to be essential for Tat transport of TMAO reductase (Buchanan et al., 2002). This residue is not positioned within a TMD and its mutation does not affect SDS-sensitivity. Of all of the components of the Tat pathway it is TatC that
shows the highest level of amino acid conservation. Twenty-one amino acids, several of which are polar residues, are strictly conserved amongst the eubacterial TatC proteins and seven of these are also conserved amongst the eukaryotic homologues (Buchanan et al., 2002; Allen et al., 2002). The majority of these conserved residues fall within predicted cytoplasmic loops of the protein. Recent site-directed mutagenesis experiments have confirmed an essential role for some of these residues in the operation of the Tat pathway (Allen et al., 2002; Buchanan et al., 2002).

The cytoplasmic side of TatC seems to be particularly important for function. Several residues within cytoplasmic loops are completely or functionally conserved across prokaryotes, chloroplasts, and plant mitochondria, and site-directed mutagenesis of a number of these residues has revealed that they are important or essential for TatC function (Buchanan et al., 2002; Allen et al., 2002; Barrett and Robinson, 2005). In vitro biochemical studies have revealed that TatC serves as the initial docking site for Tat signal peptides. Signal peptides from bacteria and chloroplasts interact with TatC either alone or in a complex with TatB (Cline and Mori, 2001; De Leeuw et al., 2002; Alami et al., 2003). It is likely that more than one TatC monomer is required for each targeting event, since two mutant versions of TatC that alone blocked Tat-dependent transport support transport when expressed together (Buchanan et al., 2002). Based on cross-linking of in vitro translocated proteins into inverted membrane vesicles, Alami et al. discovered that TatC in particular exhibits extensive contact with the signal peptide and recognizes the RR motif (Cline and Mori, 2001; Alami et al., 2003).
1.2.4.2.2. The Translocation Pore

The evolution of specific protein translocators solved an essential problem for cells, allowing proteins to be moved across or inserted into the lipid bilayers. The Tat translocation pore of both the *E. coli* cytoplasmic membrane and the chloroplast thylakoid membrane comprises three integral membrane proteins, TatA, TatB, and TatC (in chloroplasts also called Tha4, Hcf106, and cpTatC, respectively). In *E. coli* and some enterobacteria, an additional protein highly identical to TatA is encoded by the *tatE* gene (Wu et al., 2000) which is most probably the result of a cryptic gene duplication (Jack et al., 2001). TatA and TatB both are anchored in the membrane by one N-terminal transmembrane helix. Structural analyses predict at least one subsequent amphiphilic helix for each protein. The amphiphilic helix of TatA being distinctly positive interacts with anionic phospholipids and forms upon interaction with liposomes (Porcelli et al., 2002).

An invariant Glycine seems to be part of a flexibility-conveying hinge region located in between the transmembrane and amphiphilic helices of TatA and TatB (Barrett et al., 2003). A transmembrane Glu residue that is found conserved among the TatA and TatB proteins of plastids and that is lacking from the bacterial TatA, seems to play a role in the assembly of TatA into the translocase (Dabney-Smith et al., 2003). The relative stoichiometric ratio between TatA:TatB:TatC in the *E. coli* cytoplasmic membrane has been determined to be approximately 20-30:1:0.4 (Berks et al., 2003). Complexes of about 600-700 kDa containing varying amounts of TatA, TatB, and TatC were isolated from solubilized cytoplasmic membranes of a variety of bacterial organisms as well as from the
thylakoids of plant chloroplasts (Berghofer and Klosgen, 1999; Bolhuis et al., 2001; Cline and Mori, 2001; de Leeuw et al., 2002; Oates et al., 2003; Sargent et al., 2001).

It has been predicted that the Tat pore may be able to attain a diameter of up to 50–60 Å when fully open which helps to explain how the Tat system can accommodate a wide range of folded, native-like proteins with molecular weights as large as 120 kDa (Robinson and Bolhuis, 2001; Sargent et al., 2001; Berks et al., 2000). The isolation of different Tat protein complexes, together with experimental evidence from both the bacterial and cpTat pathways, has led to the idea of a modular and highly dynamic system whereby the Tat proteins exist as separate complexes in the resting state and come together to form a complete translocation pore upon substrate binding (Alami et al., 2003) and complexes of varying sizes may form pores that match the different sizes of folded Tat substrate proteins (Lee et al., 2006).

The first structural view of the Tat translocon was obtained by Sargent et al. (Sargent et al., 2001) using negative stain electron microscopy. Pore-like structures derived from a purified TatAB complex that contains an approximate 20 times molar excess of TatA to TatB and no detectible TatC were observed. The external diameter of these structures was 155 to 160 Å, with an internal diameter of around 65 Å containing one to two density features, possibly forming a gate to the pore. Single-particle electron microscopy was also used to obtain low-resolution structures of TatABC complexes from three different bacteria expressed in E. coli. Similar structures were observed for the TatABC proteins from the different species, indicating structural conservation.
1.2.4.2.3. **Tat Substrates**

The most prominent feature of the Tat pathway is its capability to transport fully folded proteins and the protein which has consensus invariant twin arginine motif (S/TRR*FLK*) in their signal peptides (Berk et al., 2005). Prior to export, the majority of bacterial Tat substrates undergo a complex cytosolic incorporation of cofactors like molybdopterin, FAD, NADP, iron-sulphur and iron-nickel clusters, copper and others (Berks et al., 2003). Cofactor insertion is, however, not a prerequisite for being exported by the Tat pathway. Several cofactor-less hydrolases were shown to use the Tat pathway. Rapid folding kinetics inherent in these enzymes might be the reason why they evolved as passengers of the Tat pathway (Jongbloed et al., 2000; Angelini et al., 2001; Voulhoux et al., 2001).

1.2.4.2.4. **Tat mechanism**

The Tat pathway is an alternate secretory pathway present in chloroplasts and many prokaryotic organisms. It is important for a wide range of biological processes, including cell division, bacterial pathogenesis, anaerobic respiration, degradation and acquisition of organic and inorganic compounds (Santini et al., 1998; Ochsner et al., 2002; Pop et al., 2002; Ding et al., 2003; Ize et al., 2003). Unlike the Sec pathway, Tat is neither ubiquitous nor essential for viability in most organisms tested (Jongbloed et al., 2000; Wu et al., 2000; Voulhoux et al., 2001; Ding et al., 2003).

Proteins using Tat pathway contain a typical highly invariant twin-arginine motif in their signal peptide (Chaddock et al., 1995; Berks, 1996) and, therefore, it was designated
as the twin-arginine translocation (Tat) pathway. In contrast to the Sec pathway, the Tat system has the unique ability to transport fully folded proteins and does not depend on the presence of nucleoside triphosphates (Robinson and Bolhuis, 2001). This pathway is involved in the export of proteins that either have to fold before translocation, such as certain co-factor containing proteins, or just fold too quickly. These folded proteins are incompatible with the Sec machinery and can only be exported via the Tat system. Most bacteria and chloroplasts contain three major components that are required for the Tat pathway. These are TatA (Tha4 in chloroplasts), TatB (Hcf106 in chloroplasts) and TatC (Bogsch et al., 1998; Sargent et al., 1998; Weiner et al., 1998; Walker et al., 1999).

As and when the pre-proteins of Tat substrates emerges from the ribosome the preprotein must avoid targeting to other pathways such as Sec, which is possible by the characteristics of the signal peptide and mature protein (Cristobal et al., 1999) or the binding of Tat-specific chaperones (Jack et al., 2004; Turner et al., 2004). After folding in the cytoplasm, any cofactors or additional subunits are added to the pre-proteins prior to targeting to the TatBC receptor complex (Cline and Mori, 2001; De Leeuw et al., 2002; Alami et al., 2003). The proton motive force drives the formation of an active translocase and the substrate is transported through a pore consisting mainly of TatA (Mori and Cline, 2002; Alami et al., 2003; Gohlke et al., 2005). Finally the folded Tat substrates will be reached to the trans side of the membrane or anchored to the membrane depending upon removal of the signal peptide by the signal peptidases. The model Tat transport cycle in *E. coli* was represented in Fig. 1.17.
1.3. Definition of the problem

Mulbry and his associates have shown existence of a 29 amino acid long signal peptide in the OPH purified from *Flavobacterium* sp. (Mulbry and Karns, 1989). Similar situation is shown to exist in *B. diminuta* (Mulbry et al., 1986). On careful perusal of its structure we have noticed existence of a sequence that resembles a typical twin arginine transport (Tat) motif (MQTRRVVLK). The Tat motif is seen only in membrane and extracellular proteins targeting or transporting across the membrane in a prefolded form.
These are mainly proteins requiring large cofactors (Berks, 1996; Palmer et al., 2005). The OPH is an esterase depending on metal ions for its activity. In fact existence of a Tat motif in a protein that uses small molecules like metal ions as cofactor is rather unusual. However the presence of Tat motifs in most of the reported OPH sequences signifies its functional relevance. The present study is designed to investigate the requirement of Tat motif for membrane targeting of OPH in *B. diminuta*. 