CHAPTER I

INTRODUCTION

1.1 Statement and significance of the problem

Phage display is a powerful technique for engineering proteins or peptides. By this technique, the bacteriophage is used as a vehicle to express diverse polypeptides as a part of phage coat protein, i.e., gpIII or gpVIII. The strategy for expressing the fusion protein is to insert the DNA coding fragment of interest between the signal sequence and phage coat protein genes. Upon viral infection, the recombinant proteins are synthesized in the bacterial host together with other coat proteins and directly translocated *via* the general secretory pathway to the inner membrane, where the signal peptide is cleaved before incorporating into phage progeny.

The general secretory (Sec) pathway is a general bacterial protein translocation system. The proteins exported *via* Sec machinery are synthesized with an N-terminal signal peptide that directs translocation in an unfolded state by the apparatus composing of soluble and membrane bound Sec gene products. The phagemids which are applied in phage display technique usually contain Sec signal sequence such as OmpA and PelB. However, the Sec-dependent pathway is suitable for only a group of proteins, since there are some limitations associated with this translocation pathway. The Sec translocon transports its substrates in unfolded form,

so this machinery is incapable of secreting proteins that fold in the cytoplasm. The translocation rate of protein delivered by the Sec pathway is quite rapid, about 10,000 amino acids per minute. However, the folding rate of some heterologously expressed chimeric proteins can be very fast compared with the translocation, thus resulting in an incompatible substrate for the Sec channel. Additionally, the bacterial periplasm does not contain ATP, so ATP-dependent molecular chaperones and/or cofactors which require for the folding of many proteins are not present. Consequently, the unfolded protein, that the molecule have not yet buried their hydrophobic amino acid stretches, that passes through the Sec channel may be not able to form the functional conformation in the periplasm since the appropriate chaperone and the cofactor are unavailable and they tend to form inclusion bodies which occur by the intermolecular hydrophobic interaction. Finally, the stop transfer signal or the local accumulation of positively charged residues in a protein's amino acid sequence can result in incomplete translocation. Taken together, an efficient protein secretion pathway is required for displaying an active protein on the phage particle.

In recent years, it has become clear that most bacteria possess a second general protein translocation pathway that is quite distinct from the Sec apparatus. This Sec-independent pathway has been termed the twin-arginine translocation (Tat) system. The signal peptide contains a peculiar motif including two consecutive arginine residues. The most remarkable characteristic of the Tat pathway is that it functions in transporting folded proteins of varying dimensions across the cytoplasmic membrane. By this pathway, proteins are exported in their native conformation since the folding process occurs before translocation thus enabling cytoplasmic chaperones and foldases to be used to reduce the amount of aggregation either in the cytoplasm or

periplasm. In this study, the Tat system was used to improve the quality of the displayed molecule, CD147, on a phage particle. CD147 is a member of human leukocyte surface molecule of the immunoglobulin superfamily and found on the surface of various cell types e.g. cancer cells and activated T-lymphocyte. Although, certain studies demonstrated the involvement of CD147 molecule in different cellular activations but the information of mechanism in signal transduction *via* CD147 is not clearly proposed since the surface ligand or receptor of CD147 has not yet been identified. A suitable method to produce a functional CD147 is needed for molecular and cellular discovering of CD147 functions.

1.2 Literature review

1.2.1 The human CD147

1.2.1.1 The structure of CD147

CD147 is a cell surface transmembrane glycoprotein with a molecular weight of 50-60 kDa. It is also known as human basigin (Miyauchi *et al.*, 1991), leukocytes activation-associated M6 antigen (Kasinrerk *et al.*, 1992) and extracellular matrix metalloproteinase inducer (EMMPRIN) (Biswas *et al.*, 1995). Homologues of CD147 in other species have also been discovered, such as basigin or gp42 in mouse (Miyauchi *et al.*, 1991), OX47 in rat (Fossum *et al.*, 1991; Nehme *et al.*, 1995) and 5A11, HT7 or neurothelin in chicken (Fadool and Linser, 1993; Seulberger *et al.*, 1992). The gene name for CD147 given in the genome project is basigin both in mouse (Bsg) and human (BSG), the latter presents on human chromosome 19, p13.3 (Muramatsu and Miyauchi, 2003). CD147 has typical features of type I integral membrane protein and it is composed of two immunoglobulin domains in the extracellular region, a single transmembrane domain and a short cytoplasmic domain containing 39 amino acid (Biswas *et al.*, 1995; Miyauchi *et al.*, 1991; Muramatsu and Miyauchi, 2003) (Figure1.1). The extracellular region contains three Asn glycosylation sites (Muramatsu and Miyauchi, 2003) and the glycosylation was shown to determine its MMP stimulating activity (Guo *et al.*, 1997; Sun and Hemler, 2001). Endoglycosidase F treatment of immunoprecipitates resulted in a mobility shift from 54 kDa to 28 kDa demonstrating that the majority of the oligosaccharide chains are N-linked. The first Ig domain is involved in matrix metalloproteinase (MMP) production (Biswas *et al.*, 1995; Sun and Hemler, 2001) and homo-oligomerization on

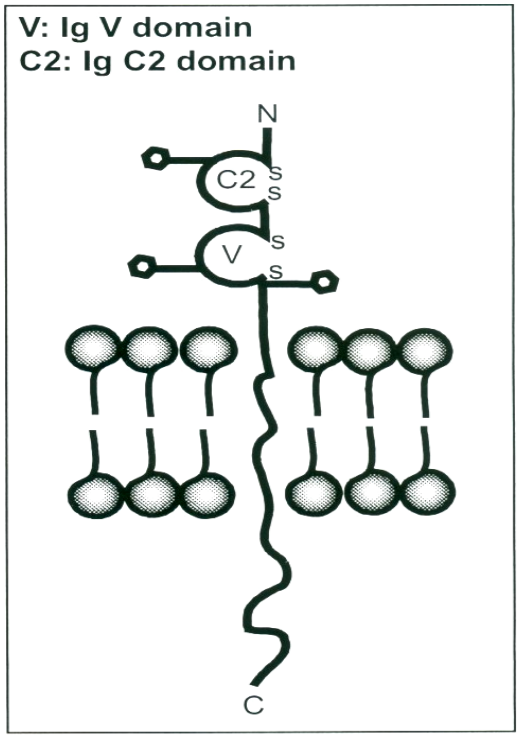

Figure 1.1 Schematic represents the structure of CD147.

the cell surface (Yoshida *et al.*, 2000) while, the second Ig domain is required for association with caveolin-1 which leads to decreased self-association on the cell surface (Muramatsu and Miyauchi, 2003). A stretch of 24 amino acids in the transmembrane region is completely conserved among human, mouse and chicken (Miyauchi *et al.*, 1991). The presence of the charged amino acid, glutamic acid, in the middle of this region is not commonly encountered in single transmembrane domain (Green, 1991), but usual for a protein with multiple transmembrane domains (Saier, 1994). This implies the engagement of CD147 in association with another transmembrane protein.

1.2.1.2 Cell and Tissue distribution

CD147 widely expresses on many cell types, such as activated T cells (Kasinrerk *et al.*, 1992), differentiated macrophage (Major *et al.*, 2002), retinal pigment epithelium (Marmorstein *et al.*, 1998), endometrium (Noguchi *et al.*, 2003), and human keratinocytes (DeCastro *et al.*, 1996). Elevated CD147 expression is often correlated with tumor progression of gliomas (Sameshima *et al.*, 2000), hepatomas (Jiang *et al.*, 2001), squamous cell carcinomas (Bordador *et al.*, 2000) and melanomas (Kanekura *et al.*, 2002). As CD147 expression is elevated on human tumor cells and was shown to increase tumor cell invasion, most studies focused on its role in cancer progression but the presence of CD147 in non-tumoral tissues suggests a role in other physiological situations.

1.2.1.3 The functions of CD147

In *in vitro* study by Biswas demonstrated that CD147 expressed by cultured tumor cells stimulates fibroblasts to produce very high levels of collagenase activity, MMP-1, which likely facilitates tumor metastasis (Biswas *et al.*, 1995). The later studies confirmed the capacity of recombinant EMMPRIN to stimulate the production of MMP-1 (interstitial collaginase), MMP-2 (gelatinase A), MMP-3 (stromelysin 1) by fibroblasts *in vitro* (Guo *et al.*, 1997; Kataoka *et al.*, 1993). Consistent with these studies, the degree of MMP expression by stromal fibroblasts in a wide range of tumors has been shown to be correlated with CD147 expression level (Caudroy *et al.*, 1999; Dalberg *et al.*, 2000; Thorns *et al.*, 2002), suggesting a key role for CD147 in the induction of MMPs at the site of tumor-stroma interaction. Studies *in vitro* have shown that CD147 effect is not limited to tumor-fibroblast or tumor-endothelial cell heterophilic interaction. The ability of CD147 to stimulate MMPs production within the same population of cells was also demonstrated in both tumor cells and fibroblasts (Caudroy *et al.*, 2002; Li *et al.*, 2001; Sun and Hemler, 2001; Zucker *et al.*, 2001). CD147 also can promote hyaluronan production and mammary carcinoma cell growth (Marieb *et al.*, 2004), affect the activation and development of T cells (Renno *et al.*, 2002; Staffler *et al.*, 2003), require for correctly plasma membrane expression and function of monocarboxylate transporters (MCTs) (Kirk *et al.*, 2000) and act as a receptor for cyclophilin A (Yurchenko *et al.*, 2002).

Additional studies demonstrated the role of CD147 in development and reproduction. Deletion of basigin gene in mice leads to defects in spermatogenesis, female fertilization, and retinal development (Hori *et al.*, 2000; Igakura *et al.*, 1998; Kuno *et al.*, 1998). Many basigin-null mutants are lost around the time of

implantation (Igakura *et al.*, 1998) and those that survive are sterile. These observations in the mutant mice are also consistent with the dose dependent inhibition of fertilization that resulted using basigin blocking antibodies. The role of CD147 was also described in human placenta and fetal membrane, suggesting that the stimulation of MMP production by CD147 may facilitate fetal membrane rupture and detachment from maternal uterus at the time of parturition (Li *et al.*, 2004). Furthermore, CD147 implicates in the human reproductive process by the cyclic modification of endometrium tissue from the proliferative to secretory phases during the normal menstrual cycle (Noguchi *et al.*, 2003).

1.2.1.4 The CD147 mAbs

In the Sixth International Workshop and Conference of Human Leukocyte Differentiation Antigens, several mAbs were submitted and clustered. The submitted mAbs were recognized several distinct epitopes of the human CD147 i.e. AAA6 (Felzmann *et al.,* 1991), UM-8D6, HI197, HIM6 and H84. All of these mAbs recognize determinants in the N-terminal Ig domain (D1) of the CD147 molecule (Stockinger *et al.,* 1997).

In the recent years, Kasinrerk and colleagues produced the CD147 mAbs i.e. M6-1B9 (IgG3), M6-2B1 (IgM), M6-1D4 (IgM), M6-1E9 (IgG2a), M6-1F3 (IgM) and M6-2F9 (IgM) (Kasinrerk *et al.,* 1999). Three of six mAbs (M6-1B9, M6-1D4 and M6-1E9) recognize the linear epitopes of human CD147 other three mAbs recognize the conformational epitopes (Pengin, 2003). Moreover, three CD147 mAbs (M6-1D4, M6-1F3 and M6-2F9) induce cell aggregation (Kasinrerk *et al.,* 1999).

1.2.2 Protein translocation across bacterial cytoplasmic membrane

The targeting and transport of proteins across biological membrane into the periplasm is one of the fundamental features of cellular life. Proteins located within the periplasmic space perform many crucial roles. For example, the detoxifying enzymes play a role in the inhibition of the activity of molecules which are toxic to the cell. Nucleases, peptidase and other scavenging enzymes metabolize large complex molecules into simpler ones that can be utilized by the cell. Proteins that are exported to the bacterial periplasm are usually synthesized with cleavable N-terminal signal sequences, termed signal peptides, which direct the protein to a specific transporter complex in the cytoplasmic membrane. The signal sequences in general have a tripartite structure where a short, basic n-region precedes a longer hydrophobic stretch of amino acid (h-region) and follows by the c-region, which normally contains a recognition sequence for the enzyme signal peptidase. In bacteria, two major routes are used to achieve protein translocation across the cytoplasmic membrane.

1.2.2.1 The general secretory (Sec) pathway

The vast majority of periplasmic proteins are exported by the well-characterized Sec apparatus (Danese and Silhavy, 1998; Manting and Driessen, 2000; Pugsley, 1993). A defining feature of Sec mechanism is that proteins are translocated in an extended conformation by binding with the Sec-cytoplasmic protein to prevent folding before export (Figure 1.2). Proteins transported by this pathway have in common a signal peptide of 18-26 amino acids in their amino-terminal region (Cristobal *et al.*, 1999; Izard and Kendall, 1994). These signal peptides have three

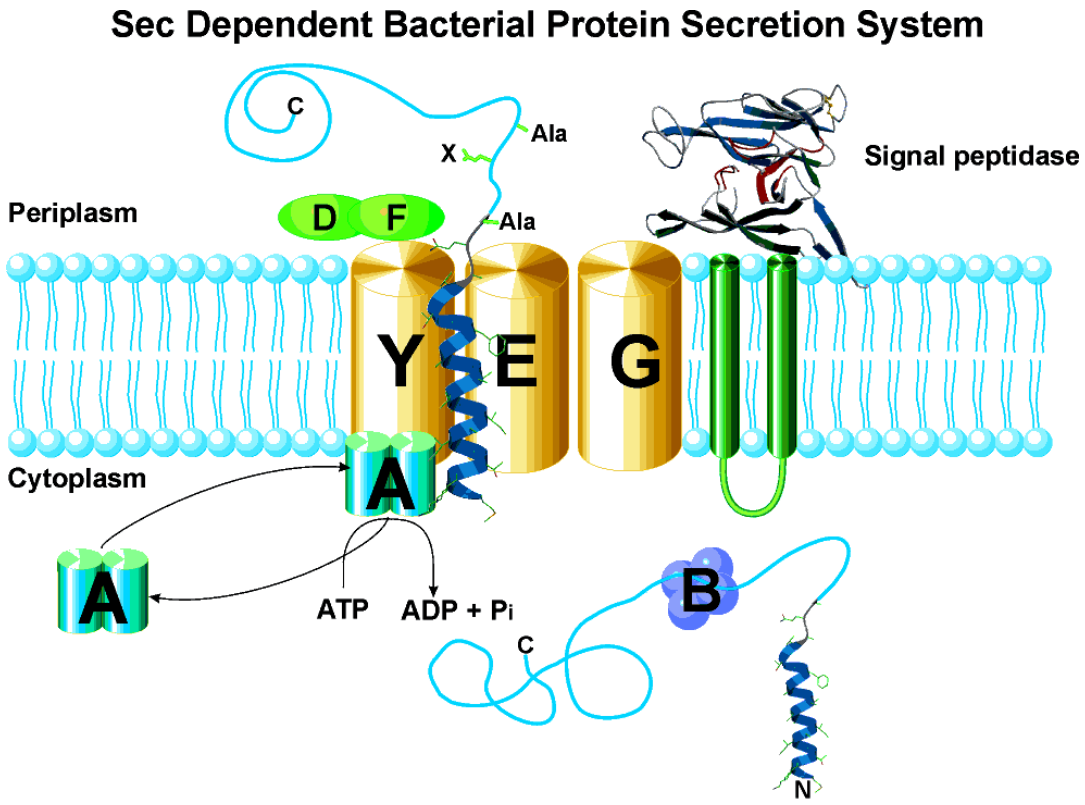

Figure 1.2 The general secretory (Sec) pathway in bacteria. Translocation of unfolded proteins involves SecA and a membrane-bound SecYEG complex. ATP hydrolysis by SecA drives the translocation of unfolded intermediate through the Sec apparatus. After exportation, substrates are processed to the mature form by signal peptidase (http://byron.biochem.ubc.ca/research2.sec2.htm).

characteristic regions: a positively charged amino acid at the N-terminus, a highly hydrophobic region, and a polar region containing the signal peptidase cleavage site. The Sec machinery involves a membrane-embedded SecYEG translocation complex, together with SecD, SecF and yajC (Danese and Silhavy, 1997; Duong and Wickner, 1997). The Sec chaperone protein, SecB, recognizes the translocation substrates and, with ATP hydrolysing SecA protein, targets them to the inner membrane. Hydrolysis of ATP causes both the dissociation the preprotein from SecA and SecA deinsertion from the plasma membrane (Wickner, 1994). Cycles of substrate binding and ATP hydrolysis drive the substrate through the Tat apparatus. The driving force for Sec-dependent translocation is dependent on both ATP and the membrane proton electrochemical gradient. (de Keyzer *et al.*, 2003; Schiebel *et al.*, 1991; Yoshida *et al.*, 2000).

1.2.2.2 The twin-arginine translocation (Tat) pathway

In recent years, a second general transport pathway, designated Tat (for twin-arginine translocation) system, has been described (Berks, 1996; Santini *et al.*, 2001; Sargent *et al.*, 1998) (Figure 1.3). It works in parallel with the *E. coli* Sec translocation system to transport folded proteins across the cytoplasmic membrane. The name was appointed from the characteristic twin-arginine motif present in the N-region of Tat substrates (Sargent *et al.*, 1998). The Tat pathway was originally discovered in plant chloroplasts where it involves in the transport of proteins from stroma into thylakiods (Chaddock *et al.*, 1995; Summer *et al.*, 2000). The most remarkable feature of the Tat pathway is that it apparently functions to transport folded proteins of variable dimensions across the cytoplasmic membrane without

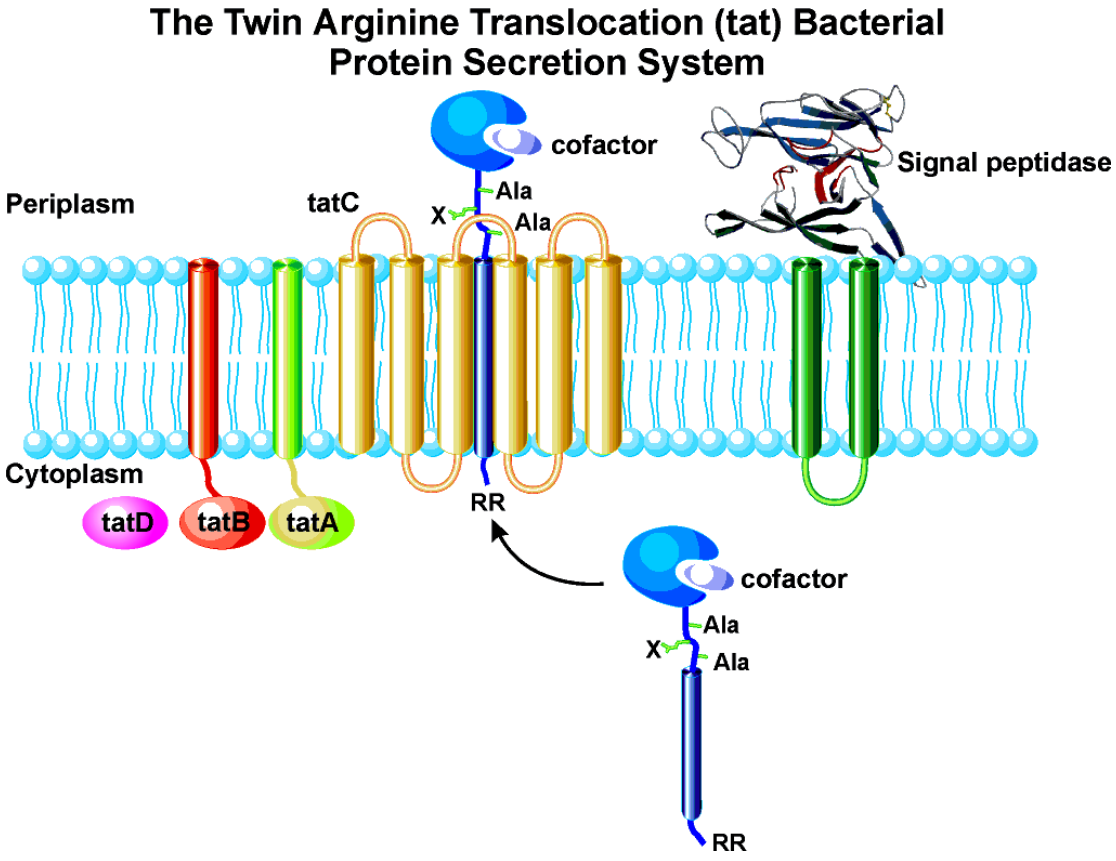

Figure 1.3 The twin-arginine translocation (Tat) pathway. Most (but not all) of the substrates for this pathway bind redox cofactors and function in cytoplasm. After synthesis in the cytoplasm, the substrates are folded and bound to cofactor, and transferred through the Tat apparatus in a folded form and finally undergo processing to the mature size (http://byron.biochem.ubc.ca/researchtat.tat.htm).

rendering the membrane freely permeable to proton and other ions. The substrates of this pathway are usually cofactor containing proteins (Berks, 1996; Santini *et al.*, 2001; Weiner *et al.*, 1998), however, some cofactor-less proteins may also be transported by Tat pathway probably because they either require cytoplasmic factor for folding or too rapidly or too tightly for Sec system to handle.

1.2.2.3 The Tat signal peptide

Signal sequences that target proteins to the Tat apparatus form the tripartite structure but have additional features that differ from those present on the Sec substrates. The most notable feature is a conserved S-R-R-x-F-L-K sequence motif at the n-region/h-region boundary, in which the consecutive arginine residues are almost invariant and the other amino acids are found with a frequency exceeding 50% (Berks, 1996; Wexler *et al.*, 2000) (Figuer 1.4). A number of several experimental studies using natural substrates of the Tat pathway have shown that both arginines are absolutely required to route a protein through the Tat pathway successfully (Grossman *et al.*, 1984; Halbig *et al.*, 1999) Further, bacterial Tat signal peptides are on average 14 amino acids longer than Sec signal peptides. The extended length of Tat signal peptides is due to an extended n-region (Cristobal *et al.*, 1999). Additionally, the h-region of Tat signal peptide is markedly less hydrophobic than that of Sec signal peptide (Cristobal *et al.*, 1999). The increasing of the hydrophobicity of the signal peptide h-region of chimeric precursor (fusion between Sec substrate and Tat signal peptide) resulted in a rerouting of export from Tat to the Sec translocon, even though the twin-arginine motif remained intact (Cristobal *et al.*, 1999). Finally, the c-region of Tat signal peptides frequently contains basic amino

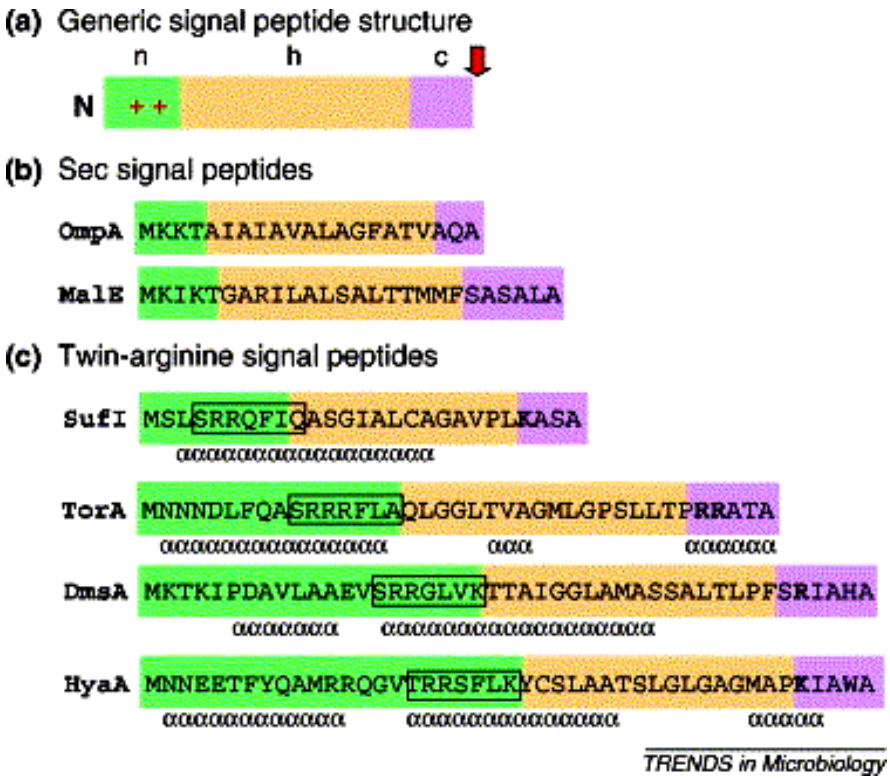

Figure 1.4 The tripartite structure of Sec and Tat signal peptides. (a) The generic tripartite structure of both Sec and Tat signal peptides (b) Two examples of extensively characterized Sec signal peptides from *E. coli*, the outer membrane protein (OmpA) and periplasmic maltose-binding protein (MalE). (c) Some representative twin-arginine signal peptides. The conserved residues that contribute to the twin-arginine motif are boxed in each case and always lie close to the n-region-c-region boundary. The "α" shows the regions predicted to be α-helic.

acids that are almost never found in Sec signals (Bruser *et al.*, 1998; Cristobal *et al.*, 1999) and are proposed to serve as Sec avoidance motif (Blaudeck *et al.*, 2003; Sargent *et al.*, 1998).

There are also some evidences that the Tat machinery from different bacteria may show specificity towards cognate signal peptides (Blaudeck *et al.*, 2001; Pop *et al.*, 2002). The glucose fructose oxidoreductase (GFOR) which is the natural Tat substrate of *Zymomonas mobilis* is not exported when heterologously expressed in *Escherichia coli*. However, when the native GFOR signal sequence is replaced by the *E. coli* Tat signal sequence, TorA, the hybrid protein is exported by Tat pathway of *E. coli* (Blaudeck *et al.*, 2001).

1.2.2.4 Substrates of the Tat pathway

The most prominent feature of the Tat system is the export of folded proteins and the currently identified Tat substrates are proteins containing certain types of cofactor. Therefore, the bacterial Tat substrates undergo a complex cytosolic association of cofactor like molybdopterin, FAD, $NADP^+$, iron-sulfur and iron-nickel cluster, and copper (Berks, 1996). However, some proteins without cofactor are also transported by the Tat pathway (Berks, 1996; Bruser *et al.*, 1998). Additionally, the heterologous proteins can be successfully targeted to the Tat machinery when fused to a Tat signal peptide. Examples include chloramphenicol acetyl transferase (Stanley *et al.*, 2002) and green fluorescent protein (GFP) which was not emitted green light in the periplasmic space of *Escherichia coli* when transported in an unfolded conformation as Sec substrate but only when exported as Tat substrate (Santini *et al.*, 2001; Thomas *et al.*, 2001). Other examples are the normally Sec-dependent proteins

β-lactamase, periplasmic P2 domain of signal peptidase and colicin V (Cristobal *et al.*, 1999; Niviere *et al.*, 1992; Stanley *et al.*, 2002). It is now clear that some substrates of the Tat system are integral membrane proteins (Hatzixanthis *et al.*, 2003; Summer *et al.*, 2000). The most apparent examples are Rieske iron-sulfer protein subunits of the cytochrome bc_1 and cytochrome b_6f complex.

1.2.2.5 Tat pathway components

In *Escherichia coli*, four genes have been shown to encode components of the Tat pathway (Sargent *et al.*, 1998). The *tatA*, *tatB*, and *tatC*, are tandem organized in *tatABCD* operon, whereas the fourth gene, *tatE*, is monocistronic. TatA, TatB and TatE are homologous proteins that are predicted to contain a membrane-spanning α-helix at the N-terminus, followed by an amphipathic helix located at the cytoplasmic side of the membrane and a carboxy terminal region of variable range (Palmer and Berks, 2003) (Figure 1.5). TatA and TatE exhibit greater than 50% sequence identity and share overlapping function in Tat translocation. In genetic study, *tatA* is transcribed and translated at a high level than *tatE*, thus a more severe export defect was observed in *tatA* mutant than *tatE* strain (Jack *et al.*, 2001). Deletion of either of these genes results in a decrease in the range of substrates, while deletions in both results in complete loss of Tat-dependent export (Sargent *et al.*, 1998). Although TatB has 20% sequence identity with TatA/TatE, it serves a distinct function in export. A deletion mutation of *tatB* alone is enough to completely abolish the translocation of some but not all Tat substrate. Additionally, deletion of *tatB* results in a rapid degradation of TatC, indicating that TatB might be important for its stabilization (Sargent *et al.*, 1999). TatC protein is predicted to have

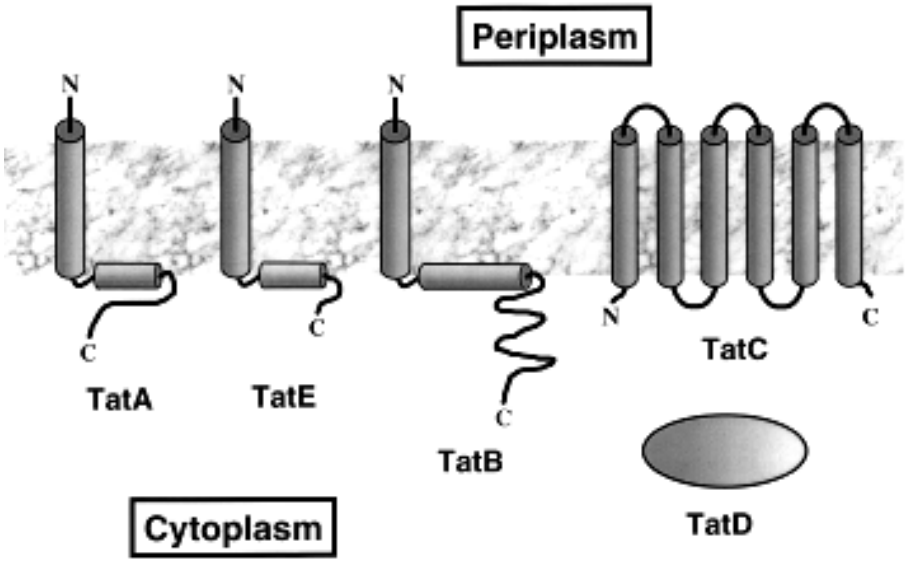

Figure 1.5 Predicted topological organization of the *E. coli* Tat components.

six transmembrane helices, with the amino and carboxy termini located at the cytoplasmic side of the membrane. All of the components of the Tat pathway, TatC shows the highest level of the amino acid conservation. The majority of these conserved residues locate at the cytoplasmic loop of the protein (Allen *et al.*, 2002; Buchanan *et al.*, 2002). TatC was recently shown to be both necessary and sufficient for specific membrane-targeting of Tat substrate *via* its twin arginine motif (Alami *et al.*, 2003). TatD is a water-soluble cytosolic protein which has no apparent function in Tat-dependent protein transport (Wexler *et al.*, 2000).

The studies in *Escherichia coli* indicate that TatA/TatB/TatC are produced at a ratio of 25 : 1 : 0.5 (Jack *et al.*, 2001). These data indicate that TatA is presented at a high molar excess over the other components of the Tat pathway. Two types of the high molecular weight complexes have been isolated from the membranes of *E. coli* overproducing the known Tat components. The first type contains TatB and TatC in an equimolar ratio together with a low level of TatA with a molecular weight of approximately 600 kDa (Bolhuis *et al.*, 2001). This complex specifically interacts with Tat signal peptides (de Leeuw *et al.*, 2002). The second type complex comprises TatA together with a very small proportion of TatB, which has a relative mass of 600 kDa (de Leeuw *et al.*, 2002; Sargent *et al.*, 2001).

Tat substrates are transported across the cytoplasmic membrane through an aqueous channel. Inspection of the structure of Tat substrate, the largest protein complex known to be transported by the Tat system is the 142 kDa FdnGH subcomplex of *E. coli* formate dehydrogenase-N (Berg *et al*, 1991). Even assuming that the subcomplex is spherical in shape, the diameter will not be above to 70 Å (Berks *et al.*, 2000). This consideration suggests that the protein translocating channel

of the Tat machinery has a maximum diameter of 70 Å. It is assumed that the number of α-helices requires to enclose an aqueous channel of this diameter is greater than 20 (Berks *et al.*, 2000). A number of evidences point towards TatA playing the major role in forming this because TatA presents in large molar excess over the other Tat components and apparently acts subsequent to signal-peptide recognition by TatBC. Additionally, negative-stain electron microscropy of an isolated TatAB complex which contains TatA in an approximate 20-fold molar excess over TatB (Sargent *et al.*, 2001), resulted in the visualization of annular structures with a central cavity of 70 Å in diameter.

1.2.2.6 Tat mechanism

In *Escherichia coli*, several models of the Tat transport operation have been proposed (Berks *et al.*, 2000; Bruser and Sanders, 2003; Musser and Theg, 2000). Recently, the *in vivo* and *in vitro* studies provided insight into the Tat transport mechanism (Palmer *et al.*, 2005). Tat substrate is synthesized with N-terminal signal peptide in the cytoplasm. The exposed twin-arginine signal peptide is recognized and bound by a proofreading chaperone. Binding of the chaperone conceals the twin-arginine motif and prevents targeting of the immature apoprotein to the TatBC complex. The proofreading chaperone also binds to the region of immature apoprotein maintaining it in a partly folded state that is competent for cofactor insertion. The integration of cofactor leads to release of proofreading chaperone and revealing the active Twin-arginine signal peptide. This quality control mechanism allows Tat system to discriminate against unfolded substrate (Palmer *et al.*, 2005) (Figure 1.6).

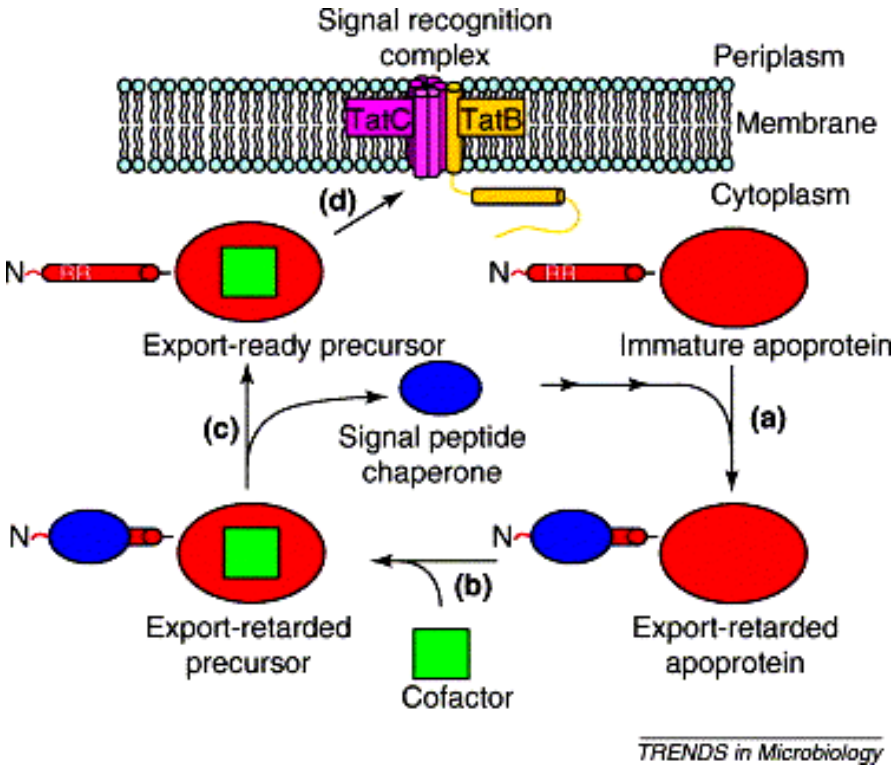

Figure 1.6 A model of quality control mediated by twin-arginine signal-peptide binding chaperones (Palmer T, 2005).

The export-ready preprotein binds to a site in TatC *via* Twin-arginine signal peptide (Alami *et al.*, 2003). The TatBC-substrate complex subsequently associates with TatA by proton motive force (Mori and Cline, 2002), allowing the movement of substrate across the membrane through TatA channel. After protein translocation has been completed the signal peptide is cleaved by the signal peptidase and the TatA and TatBC components subsequently dissociate and return to the resting state.

1.2.3 Protein folding in *E. coli*

Newly synthesized polypeptide chains must fold and assemble into unique three-dimensional structures in order to attain their biological function. In general, small peptides or single domain host proteins efficiently reach a native conformation owing to their fast folding kinetic, whereas large multidomain and overproducing recombinant proteins often require the assistance of folding modulators which include of molecular chaperones and folding catalysts. Molecular chaperones are a ubiquitous class of proteins that play an essential role in protein folding by helping other polypeptides reach a proper conformation without becoming part of the final structure. The chaperone proteins act in preventing off-pathway by shielding the hydrophobic amino acid residues and stabilizing nonnative polypeptides, whereas the folding catalysts accelerate specific rate-limiting steps in folding, such as isomerization of peptide bonds and rearrangement of disulfide bonds. Three functionally distinct classes of chaperones cooperate in the conformational quality control of the *E. coli* proteome. Folding chaperones (e.g., DnaK and GroEL) are the heart of chaperone network which uses conformational changes fueled by ATP

hydrolysis to promote the folding of bound substrates. Holding chaperones (e.g., IbpA and Hsp33) stabilize partially folded protein to await availability of folding chaperones upon stress abatement. Finally, disaggregating chaperones (e.g., ClpB) employ ATP-driven conformational changes to solubilize protein aggregates that a result of stress and transfer them to the folding chaperones for subsequent refolding.

1.2.3.1 Cytoplasmic protein folding

In the *E. coli* cytoplasm, three chaperones, trigger factor (TF), DnaK-DnaJ-GrpE and GroEL-GroES participate in the folding of newly synthesized protein. TF has a modular structure with an N-terminal domain that mediates ribosome binding, a central peptidyl-prolyl-*cis*/*trans* isomerase (PPIase) domain, and a C-terminal domain with unknown function (Hesterkamp *et al.*, 1997; Stoller *et al.*, 1995). TF associates with the large ribosomal subunit with the moderate affinity near the polypeptide exit site and crosslinks to the polypeptide chains (Kramer *et al.*, 2002). TF is thought to interact primarily with short nascent chains and its function overlaps with that of DnaK, which interacts with longer nascent chains downstream of TF. The chaperone action of DnaK is powered by ATP hydrolysis and is assisted by partner chaperones DnaJ and GrpE (Bukau and Horwich, 1998). DnaJ acts as ATPase activator and binds to an unfolded polypeptide and then presents it to the ATP-bound form of DnaK. The ATP hydrolysis causes the DnaK-peptide binding cleft to close and the polypeptide substrate is stably bound. GrpE then catalyzes the exchange of the bound ADP for ATP. This reaction reverts DnaK to its low substrate-affinity state and facilitates the release of the substrate. Once released, a newly synthesized protein may reach a native conformation, undergo additional cycles of interactions with DnaK, or

be transferred to the downstream GroEL-GroES system. GroEL is an oligomer organized as double-ring complexes, which make up the open-cylinder and expose a number of hydrophobic residues towards the ring cavity (Hartl and Hayer-Hartl, 2002). The folding cage, termed *cis*-cavity, is established from the binding of ATP and polypeptide substrate to the GroEL and following with the binding of GroES at the apical GroEL domain. After ATP hydrolysis in the *cis*-ring, the ATP binds to the *trans*-ring of GroEL and the GroES is released. Upon opening the folding cage, folded protein can exit, whereas the folding intermediates that still expose hydrophobic surface may be recaptured for a subsequent round of folding. In the environmental stress conditions, DnaK and GroEL mediate the refolding of host proteins by the assistance of holding chaperones (holdases), which stabilize partially folded or aggregated proteins without promoting their remodeling. Once stress abates, holdase-bound substrate is attracted by DnaK, and may be transferred to GroEL, for folding (Veinger *et al.*, 1998). The most extensively characterized holdases are IbpA and IbpB (Laskowska *et al.*, 1996; Mogk *et al.*, 2003). Hsp33 is also identified as holdase. Hsp33 is a redox-regulated chaperone in which the main function is to manage oxidative protein misfolding (Graf and Jakob, 2002). The redox sensor in Hsp33 is a cysteine center that coordinates zinc under reducing conditions. When the cytoplasm becomes more oxidizing from emergence of reactive oxygen species that accompanies heat shock, a zinc atom is released and Hsp33 forms two intramolecular disulfide bonds which exhibit the chaperone activity (Graf and Jakob, 2002).

The *E. coli* posses the third chaperone modulator (Figure 1.7), disaggregating chaperone, to defense the deleterious conditions associated with protein misfolding and aggregation. ClpB is described to disaggregate high molecular

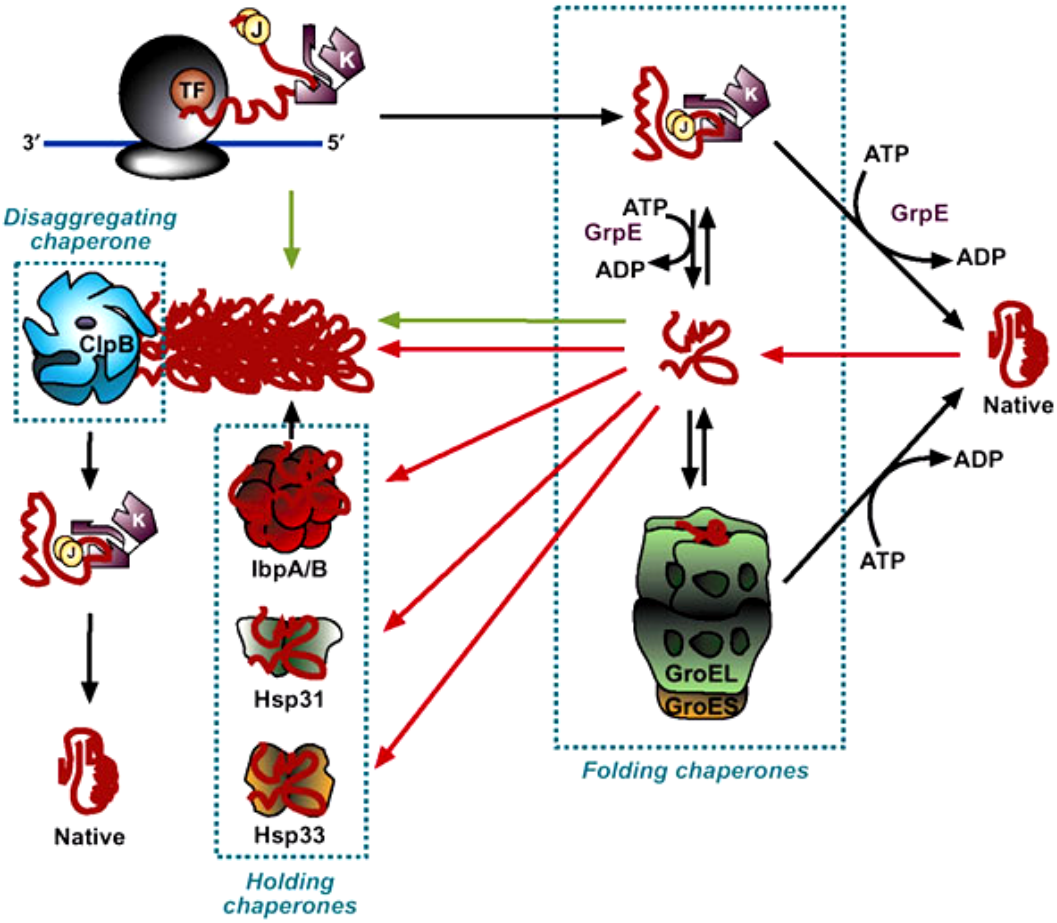

Figure 1.7 Chaperone-assisted protein folding in the cytoplasm of *E. coli*. Nascent polypeptides require the assistance of TF or DnaK-DnaJ. In time of stress (red arrow), thermolabile proteins unfold and aggregate. IbpB is required to serve partially folded proteins as unfolded intermediate until folding chaperone available. Proteins that miss an early interaction with TF or DnaK/DnaJ, that undergo multiple cycles of abortive interaction with folding chaperones, accumulate in inclusion bodies (green arrow).

weight aggregates into smaller aggregates by the motion of long surface domains which generate the mechanical force to pull apart of large aggregates (Lee *et al.*, 2003). ClpB also assists in the resolubilization of medium-size aggregates by threading substrate through the central cavity of ClpB (Lee *et al.*, 2003). The disaggregation of high molecular weight aggregates by ClpB creates the smaller and medium-size aggregates that may be exposed hydrophobic surfaces, which are recognized in turn by the DnaK chaperone system (Goloubinoff *et al.*, 1999)

1.2.3.2 Periplasmic protein folding

The proteins that translocated across the inner membrane exert their biological functions in the periplasmic or destine for insertion into outer membrane. These proteins must traverse the periplasm and the peptidoglycan layer to reach their destination. During they transit as nascent or partially folded polypeptide they require protection from misfolding and aggregation. The periplasm contains a lower group of molecular chaperones than does in the cytoplasmic compartment and all of them are ATP-independent chaperones. The periplasmic chaperones have been proposed to assist the folding and membrane insertion of outer membrane proteins. Skp is a 17 kDa protein that binds several outer membrane proteins including OmpA, OmpC, OmpF, PhoE, and LamB (Tamm *et al.*, 2004). The Skp central cavity can accommodate the substrates up to 20 kDa (Walton and Sousa, 2004). Skp chaperone activity is likely a holdase. It binds to unfolded substrate immediately after they are translocated across the inner membrane that keeps them in an unfolded form and prevents them from aggregation in the periplasm (Harms *et al.*, 2001). FkpA is the peptidyl-prolyl *cis/trans* isomerase, which catalyzes the interconversion between *cis*

and *trans* form of the peptide bond X-Pro, where X is any amino acid. FkpA is an important periplasmic chaperone with the generic folding activity (Missiakas *et al.*, 1996). It is believed to cradle partially folded substrates within the hydrophobic cleft formed at the dimerization interface, allowing the flexible C-terminal domains easy access to prolyl bonds requiring isomerization (Saul *et al.*, 2004). The other well characterized periplasmic protein folding modulator is SurA. SurA is a periplasmic peptidyl-prolyl isomerase that has been shown to assist the folding of several outer membrane proteins including OmpA, OmpF, and LamB. SurA is classified as specialized chaperone since it preferentially recognized an *Ar*-X-*Ar* motif (where *Ar* is an aromatic and X is any residue) that is common in outer membrane proteins but infrequent in other polypeptides (Bitto and McKay, 2003). It contains a 50 Å deep cleft within its core module that may be responsible for substrate binding (Bitto and McKay, 2002).

1.2.3.3 Disulfide bond formation

In addition to chaperones, which facilitate protein folding by binding to and stabilizing partially folded intermediates, cells contain enzymes that catalyze protein folding by breaking and reforming covalent bonds. Many proteins secreted into the periplasm form specific disulfide bonds that aid in both the folding and stabilization of the mature protein. Disulfide bonds are generally restricted to secreted proteins and some membrane proteins because the cytosol has a reducing environment that maintains cysteine residues in their reduced form. In contrast, the periplasm has an oxidizing condition that allows the formation of structural disulfide bonds. This difference appears to be the result of the particular enzymatic systems present in these

compartments, which are responsible for the oxidation, reduction, and isomerization of disulfide bonds in proteins. The cytoplasm has two systems that catalyzed the NADP-dependent reduction of disulfide bridges in target proteins: the thioredoxin/thioredoxin reductase system and glutaredoxin/glutathione system (Holmgren, 1989). The proper formation of the disulfide bonds in the periplasm is catalyzed by specific thiol-disulfide oxidoreductases. In both the periplasmic and cytoplasmic systems, the proteins responsible for catalyzing oxidation or reduction have a common thioredoxin active site motif, Cys-X-X-Cys (Martin *et al.*, 1993). Oxidation and reduction of disulfide bonds is mediated by thiol-disulfide exchange between the active site cysteines of the enzyme and cysteines in the target protein (Darby and Creighton, 1995; Frech *et al.*, 1996).

There are a number of membrane and soluble thiol-disulfide oxidoreductases that contribute to proper oxidation of structural disulfide bonds in periplasmic proteins (Figure 1.8). The first of these is DsbA, which is responsible for the formation of disulfide bonds in newly translocated proteins (Bardwell *et al.*, 1991). DsbA oxidizes its substrate by transferring the disulfide bond from its active site to the target protein, leaving its active site in the reduced state. The reduced DsbA is reoxidized in order for DsbA to catalyze another round of disulfide bond formation. Reoxidation of DsbA is performed by the integral membrane protein DsbB (Missiakas *et al.*, 1993). The genetic study suggests that DsbB is itself oxidized by passing electrons to the respiratory chain (Kobayashi *et al.*, 1997). Although DsbA can catalyze the formation of disulfide bonds, they are insufficient in catalyzing the rearrangement or isomerization of incorrectly formed disulfide bonds in substrate proteins with multiple cysteines. The isomerization of disulfide bonds is a function of

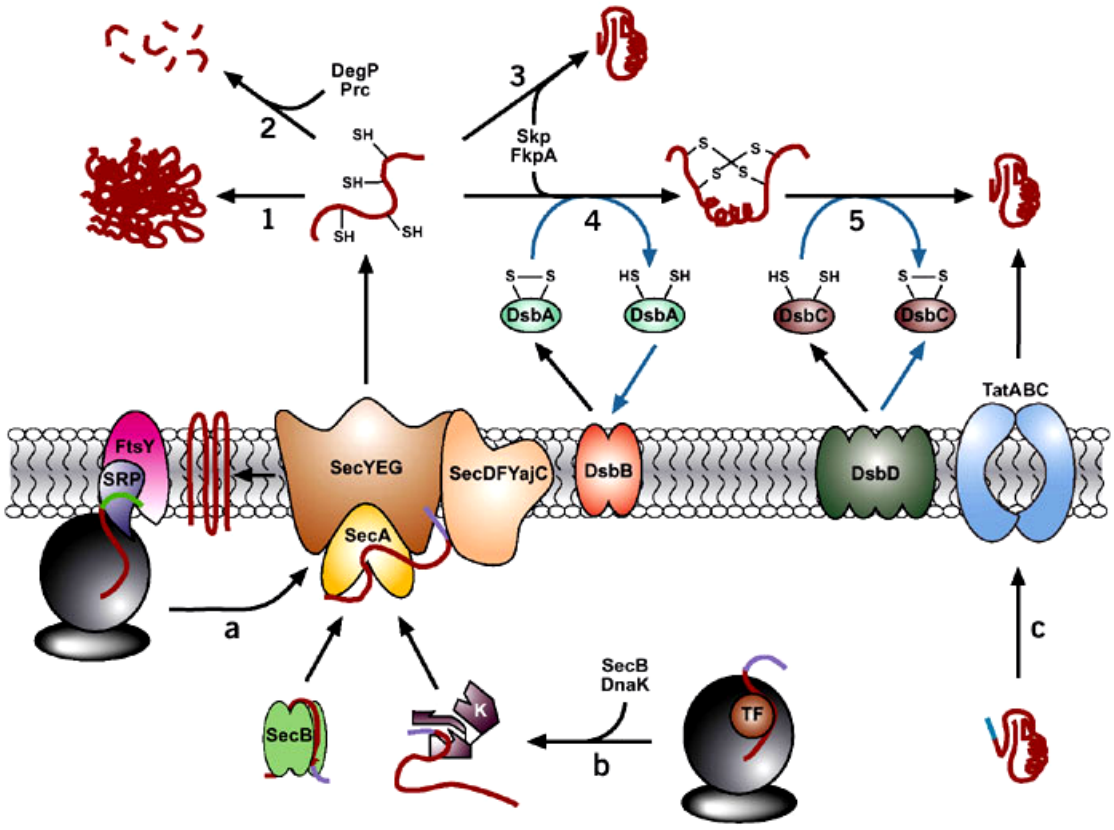

Figure 1.8 Disulfide bond formation in bacterial periplasm. Cysteine pairs in proteins containing disulfide bonds are oxidized by DsbA whereas incorrect disulfides are isomerized by DsbC. These oxidoreductasees are reactivated by DsbB and DsbD, respectively.

DscC, a periplasmic disulfide bond oxidorecudtase (Rietsch *et al.*, 1996). The Cys-X-X-Cys active site of DscC is usually found in the reduced state, making it competent for disulfide rearrangement. Maintaining DsbC in reduced form is the function of the cytoplasmic membrane protein, DsbD (Rietsch *et al.*, 1996). DsbD supports DsbC in the reduced state *via* an interaction with its own reduced active site motif. Evidence suggests that the reducing power of DsbD is acquired by electron transfer from cytoplasmic thioredoxin (Hiniker and Bardwell, 2003).

1.2.4 Protein misfolding and Inclusion body formation

Protein activity demands folding into precise three-dimensional structure. The proteins which fail to reach a native conformation or to interact with folding modulator have two possible consequences: form inclusion body and degradation. Stress situations including heat shock, starvation, exposure to toxic compounds, recombinant protein overexpression, and oxidative stress, impair protein folding and cause the formation of folding intermediates and protein misfolding. In recombinant bacteria, the over-expression of plasmid-encoded genes triggers transcription of heat-shock genes and other stress responses and often results in the aggregation of the encoded protein as inclusion bodies. The aggregation is formed by nonnative intermolecular hydrophobic interactions between protein folding intermediates, which have not yet buried their hydrophobic amino acid residues (Kiefhaber *et al.*, 1991).The factors contributing to the formation of inclusion body are the use of high inducer concentration, strong promoter and the inability of bacteria to support all post-translational that a protein requires in folding (Baneyx and Mujacic, 2004).

The Degradation by host protease is the fate of misfolded proteins to ensure that the abnormal polypeptides do not accumulate within the cell and allows amino acid recycling. In the cytoplasm, proteolytic degradation is initiated by five ATP-dependent heat shock proteases (Lon, ClpYQ/HslUV, ClpAP, ClpXP and FtsH) and completed by peptidases that hydrolyze sequences 2-5 residues in length. These proteases consist of a remodeling component that binds substrate proteins and couples ATP hydrolysis to unfolding and transfer of the polypeptide to an associated protease domain or proteolytic component. Since the periplasmic compartment has no ATP, the misfolded proteins are degraded by ATP-independent proteases. The generic periplasmic proteases are DegP, Tsp, protease III and OmpT and the most active of which are DegP and Tsp. DegP is a serine endopeptidase in which the proteolytic site is located within an inner cavity bounded by mobile side walls formed by PDZ domain. Tsp degrades nonpolar carboxy-terminal regions of protein with broad primary sequence specificity (Keiler *et al.*, 1995).

1.2.5 The stress response systems in *E. coli*

All cells appear to have systems that respond to stress situation. In *E. coli*, the stress response is compartmentalized into cytoplasmic and extracytoplasmic responses. The cytoplasmic response is operated by σ^{32}, the *rhoH* gene product, which responds to the accumulation of misfolded protein by directing the transcription of a well-characterized set of genes, including those encoding the GroEL/ES and DnaK/DnaJ chaperone and Lon protease (Bukau, 1993). These chaperones, in turn, are thought to down-regulate σ^{32} activity upon relief of cytoplasmic stress (Gamer *et al.*, 1996; Liberek *et al.*, 1992). In contrast, the extracytoplasmic is believed to be

controlled by at least two signal transduction systems, the σ^E-mediated system and the Cpx two-component system. The σ^E transcription factor protein appears to control the synthesis of several proteins, some of which are involved in protein folding or degradation in the periplasm. The σ^E regulates the synthesis of DegP protease and FkpA, as well as its own synthesis (Erickson and Gross, 1989; Raina *et al.*, 1995). The σ^E pathway is induced by the conditions that lead to misfolding of periplasmic protein. RseA and RseB appear to be negative regulator of σ^E (De Las Penas *et al.*, 1997). It has been proposed that when the periplasmic protein folding or degradation is impaired, misfolded proteins bind RseB, lowering the binding affinity of RseA for σ^E in the cytoplasm. This results in the release of σ^E to activate transcription of its own gene and of the gene for DegP protease and FkpA (Missiakas and Raina, 1997).

The second periplasmic stress response system is the two-component Cpx pathway, which composed of an inner membrane sensor kinase encoded by *cpxA* and a response regulator encoded by *cpxR*. This pathway appears to regulate the expression of gene encoding the DsbA thiol-disulfide oxidoreductase, CpxA periplasmic protein, DegP protenase and PpiD peptidyl-prolyl *cis/trans* isomerases (Danese and Silhavy, 1997; Pogliano *et al.*, 1997). The periplasmic protein misfolding and misfolded subunits of pili serve as potent activators of the Cpx signal transduction cascade (Jones *et al.*, 1997). In the absence of protein misfolding, CpxA is maintained in an inactive state by the CpxP periplasmic inhibitor. Envelope protein misfolding or pilus assembly are predicted to lead to relief of CpxP inhibition and activation of phosphotransfer between CpxA and CpxR. Phosphorylated CpxR upregulates expression of genes whose products are involved in envelope protein folding and degradation by binding to their promoter. There is another operon that appears to be

induced under stress conditions but the role of its product is not clearly proposed. Phage shock protein A (PspA), is a peripheral cytoplasmic membrane protein which is encoded by the first gene in the pspA-E operon. PspA expression can be induced by filamentous phage infection, heat shock and membrane-associated stress condition. Although the precise function of PspA is not clearly proposed, the studies demonstrated that the presence of PspA allows *E. coli* to survive in stationary phase at alkaline pH (Weiner and Model, 1994) as well as maintain the proton motive force under stress conditions (Kleerebezem *et al.*, 1996).

1.2.6 Filamentous bacteriophage

1.2.6.1 Structure of bacteriophage

The filamentous phages are a group of viruses that contain a circular single-stranded DNA genome encased in a long protein capsid cylinder. Unlike most bacterial viruses, filamentous phages are produced and secreted from infected bacteria without cell killing and lysis. The filamentous phage particle is approximately 6.5 mn in diameter and 930 nm in length (Specthrie *et al.*, 1992). Phage particles are composed of five coat proteins (Figure 1.9). The cylinder tube that surrounds the ssDNA is composed of approximately 2700 molecules of the 50-reidue major coat protein, also called gpVIII. The gpVIII molecules are packed tightly and overlapped like fish scales by the interaction between the hydrophobic midsection of adjacent subunit. The carboxy-terminal 10-13 residues form the inside wall of the cylinder. This region contains 4 positively charge lysine residues that interact with phosphates of the viral ssDNA (Greenwood *et al.*, 1991). The amino-terminal of gpVIII is presented on the outside of the particle. One end of the particle has approximately 5

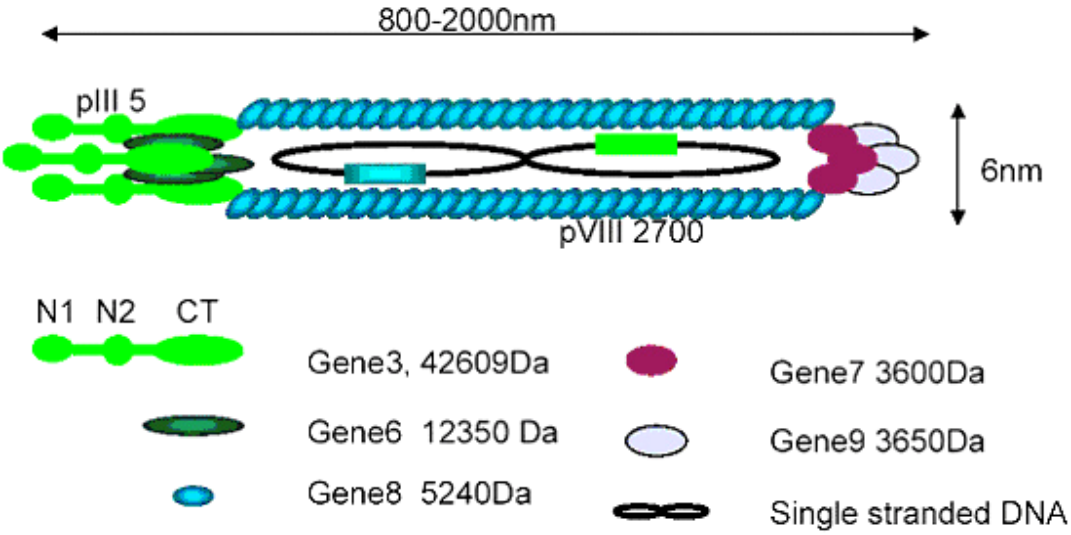

Figure 1.9 Structure of a filamentous bacteriophage. The body of phage particle is composed of the major coat protein. Five copies of gpIII and gpVI at the proximal end and five copies of gpVII and gpIX at the distal end cover the two ends of the phage particle.

molecules each of the small hydrophobic gpVII and gpIX protein. This end contains the packaging signal and is the first part of the phage to be assembled. It is not known how these two proteins are arranged at the end of the particle, however, the observation that antibodies to gpIX but not gpVII are able to interact with one end of the phage particle suggests that only gpIX is exposed at the surface and gpVII is buried close to the DNA (Endemann and Model, 1995). The other end of the particle contains approximately 5 molecules each of gpIII and gpVI. gpIII is the most commonly used coat protein for display. The N-terminal domain, which is necessary for phage infectivity, is surface exposed and forms small knob. gpIII is made up of three domains separated by glycine-rich regions (Stengele *et al.*, 1990). These three domains are designated N1, N2 and CT. The first domain, N1, is required during infection for the translocation of the DNA into the cytoplasm and the insertion of the coat proteins into the membrane. N2 is responsible for binding to the F pilus (Deng *et al.*, 1999). Both domains contain cysteine molecules that are involved in intramolecular disulfide bonds. The carboxy-terminal end makes up the third domain of gpIII (CT) that is essential for forming a stable phage particle. CT domain together with gpVI interacts with gpVIII to form the end of particle. The disposition of the gpVI in the virion is not known, but gpVI with fusions to the C-terminus can be incorporated into phage, this suggests that gpVI may be surfaced exposed.

1.2.6.2 The phage life cycle

Infection is the multistep process requiring interaction with the F conjugative pilus and the bacterial TolQ, R and A cytoplasmic membrane proteins. These three Tol proteins are absolutely required during phage infection for

translocation of the filamentous phage DNA into the cytoplasm and translocation of the phage coat proteins into the cytoplasmic membrane (Click and Webster, 1998; Russel *et al.*, 1988). Infection normally begins when the N2 domain of gpIII binds to the tip of a pilus. This binding releases N1 from N2, and allows N1 to interact with TolA. The pilus subsequently retracts, bringing the gpIII end of phage particle to the periplasm. The major capsid protein, gpVIII, gpVII and gpIX minor capsid proteins, disassemble into the cytoplasmic membrane as the phage DNA is translocated into the cytoplasm (Figure 1.10). Once in the membrane, the gpVIII protein of the infecting phage joins the pool of newly synthesized gpVIII, and both are assemble into newly form particle. Once the viral (+) strand DNA enters the cytoplasm, the complementary (-) strand is synthesized by the bacterial enzymes. The final product is a covalently closed, supercoiled, double-stranded DNA called the parental replicative form (RF) DNA. The (-) strand of this RF is the template for transcription and the resulting mRNAs are translated into all of the phage proteins. Of the 11 phage-coated proteins, three (gpII, gpX, gpV) are required to generate ssDNA, three (gpI, gpXI, gpIV) are required for phage assembly, and five (gpIII, gpVI, gpVIII, gpVIII, gpIX) are components of the phage particle. gpII nicks the (+) strand in the RF at a specific place in the intergenic region. The resulting 3'-hydroxyl acts as a primer for synthesis of a new viral strand. After one round, gpII circularizes the displaced viral (+) strand DNA, which then is converted to a covalently closed, supercoiled, double-stranded RF molecule by bacterial enzymes. gpV dimers bind cooperatively to newly generated (+) strand RF and prevent its conversion to RF DNA. The RF DNA synthesis continues until the amount of gpV reaches a critical concentration. gpX also is required for proper replication of phage DNA. Its exact role in replication is unclear,

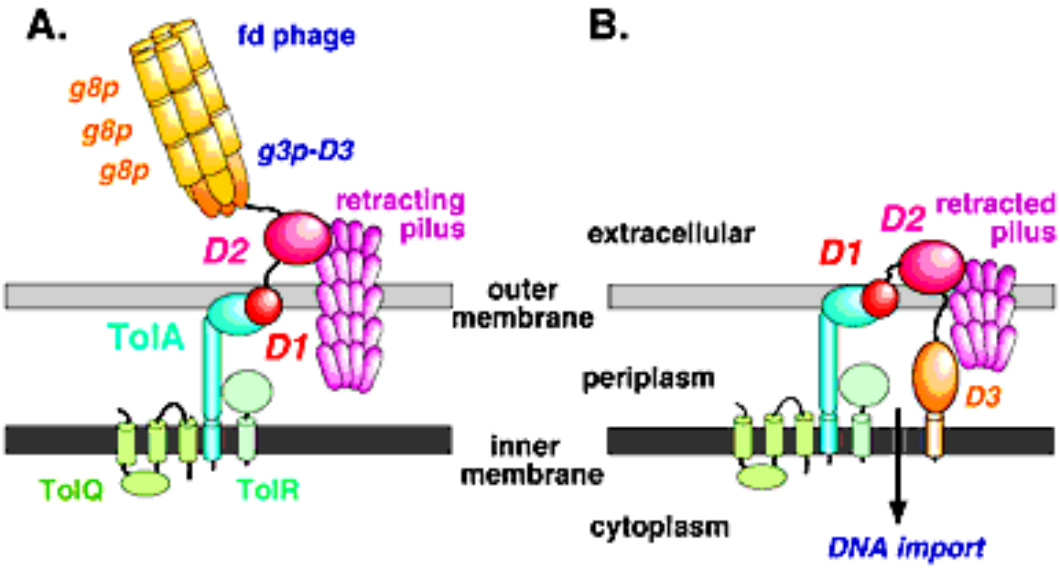

Figure 1.10 Model for Infection by filamentous phage. (www.ccmr-nmr.bioc.cam.ac.uk/~hij20/g3p.html).

but it appears to function as an inhibitor of gpII function (Fulford and Model, 1984). The result of this replicative process is a newly synthesized single-stranded DNA in a complex with many copies of gpV. The capsid proteins are all synthesized as integral membrane proteins that remain in the membrane until they are assembled around the DNA. Assembly occurs at specific site in the bacterial envelope where the cytoplasmic and outer membranes are in close contact by the interaction of gpI, gpXI and gpIV. During the assembly process, the viral DNA is extruded through the membrane-associated assembly site, where the gpV in gpV-DNA complex is removed and the capsid proteins are packaged around the DNA. This process continues until the end of the DNA is reached and the assembly is terminated by the addition of gpVI and gpIII.

1.2.7 Phage display Technique

In 1985, George Smith first showed that the linkage between phenotype and genotype could be established in the filamentous bacteriophage and gave birth to the new technology of Phage display (Smith, 1985). Phage display is a high potential technique for producing functional recombinant proteins. By this technique, DNA fragment of interest is inserted between C-terminal of the signal peptide and N-terminal coding region for phage coat protein. The recombinant proteins are synthesized in bacterial host together with other coat and accessory proteins of filamentous phage, and they will be incorporated into phage in the assembly process. The released phage particles expose the recombinant protein as a fusion product to one of the phage coat protein. By inserting different DNA fragments, a library of phage particles bearing different recombinant coat protein can be generated. Each

phage particle contains only one type of recombinant coat protein encoded by the corresponding gene fusion present inside the same phage particle. DNA sequencing allows for deduction of the amino acid sequence of recombinant coat protein. Individual can be rescued from libraries by interaction of the displayed protein with the cognate ligand, and the phage can be amplified by infection of bacteria. The recombinant polypeptides or proteins displayed on the surface of phage particles can be used for identifying and characterizing the interaction with their binding targets (Cesareni, 1992). Its potential applications include selection of high-affinity antibodies and investigation of potential ligands for orphan molecules.

1.2.7.1 Filamentous phage coat proteins used for phage display

All five capsid proteins have been used to display proteins or peptides. Generally, proteins or peptides are displayed on filamentous phage fused to either gpVIII or gpIII. There has even been a report of phage particle displaying proteins that are fused to the C-terminal portion of gpVI, although the efficiency of display appears to be lower (Jespers *et al.*, 1995). Recently, antibody heavy- and light-chain variable regions have been fused to the N-terminal of gpVII and gpIX and displayed on phage, showing that these two minor coat proteins can be used for display (Gao *et al.*, 1999).

The major coat protein, gpVIII, is a polypeptide of only 50 residues that makes up the body of phage. Generally, gpVIII presents approximately 2700 copies in an average phage particle of 6.5-kb genome. The DNA sequence encoding the peptide to be displayed on gpVIII is generally inserted between the sequence for the gpVIII signal peptide and the N-terminal coding region for the capsid protein.

This type of display on every copy of gpVIII is limited to peptide 6-8 amino acid long (Iannolo *et al.*, 1995; Malik *et al.*, 1996; Petrenko *et al.*, 1996), probably because of the size restriction of the gpIV channel through which phage pass during extrusion (Linderoth *et al.*, 1997; Marciano *et al.*, 1999). If longer insertion is required, the phagemid system can be used to provide enough wild-type gpVIII to make a hybrid phage capsid, containing both wild-type and peptide-fused gpVIII proteins which are encoded by helper phage and phagemid DNA.

The gpIII, a minor coat protein present only at the tip of the phage particle, contains 406 amino acids and forms a knob-like structure that mediates the attachment to the F pilus essential for subsequent infection. gpIII is the most used for display of proteins. The gene of interest is inserted between the signal sequence and the N-terminus of gpIII. The gpIII-base system has the advantage that the chimeric gpIII molecules containing large insert appear to package into phage reasonably well because the foreign proteins are resided at the very end of the package particle, it may create less steric hindrance when passing through the gpIV exit pore. However, its disadvantage is that only about 5 molecules can be displayed per phage particle and the large inserts tend to lower phage infectivity or make the phage noninfective (Smith, 1985). This problem can be overcome by making hybrid phage, in which only one of the gpIII molecules contains the displayed protein (Barbas *et al.*, 1991).

1.2.7.2 Phage-display vector

Protein can be displayed using vector based on the natural filamentous phage sequence (phage vectors) or using plasmid that contain only the fusion phage gene and no other phage genes (phagemid vectors) (Lowman, 1997). In phage vector,

the heterologous sequence for display is inserted into the region encoding the carboxyl terminus of the coat protein's leader sequence and the amino terminus of the mature coat protein. When introduced into *E. coli*, phage will be produced in which all copies of the coat protein display the recombinant protein (polyvalent display). In phagemid vector, the gene of interest is cloned into a small plasmid under the control of promoter. In addition to a plasmid origin of replication, the vector also has a phage origin of replication to allow production of single-stranded vector and subsequent encapsidation into phage particles. The infected cells express all the wild-type phage proteins from the helper phage genome, as well as small amount of the fusion protein encoded by the phagemid, so that phage particles are extruded by the cells that contain both proteins, usually with the wild-type in considerable excess. Because the helper phage genome is poorly packaged, nearly all the phage particles contain the phagemid genome, preserving the linkage between the displayed protein and its gene. A major advantage of phagemid vectors is their smaller size and ease of cloning, compared with the difficulties of cloning in phage vector without disrupting the structure of gene and promoter. In addition, the phagemid system must be used if monovalent display is desired in order to obtain selection based on true binding affinity.

1.2.7.2 The helper phage

In phagemid system, the helper phage provides all of the phage proteins and enzymes required for phage replication. The helper phage also provides the structural proteins that encapsulate both helper-phage and phagemid genome. Thus the helper phage is acting to help replicate and package the phagemid genome.

The M13K07 and VCSM13 are commonly used as helper phages. All have mutations that reduce packaging efficiency to ensure that phagemid genomes are preferentially packaged, and mutations that allow productive infection of bacteria harboring plasmids with phage origin of replication. M13K07 and VCSM13 carry a kanamycin resistant gene to allow antibiotic selection of helper-infected cells.

1.3 Objectives

1. To construct the new phagemid vector containing tat signal sequence for improving the folding of molecule when display by using phage display technique.
2. To study and compare the folding of CD147 epitope which produce by Sec and Tat pathway.
3. To produce phage displaying CD147 on gpVIII *via* Tat pathway.