CHAPTER 1

INTRODUCTION

1.1 Statement and significance of the problem

CD147 is a cell surface glycoprotein that belongs to the immunoglobulin superfamily. This surface molecule is expressed on all leukocytes, erythrocytes, thrombocytes and endothelial cells (Stockinger H, 1997). It plays an important role in mediating signal transduction induced by the binding of specific ligands and regulating central cellular processes such as cell proliferation (Wang et al., 2006), apoptosis (Intasai et al., 2006), or cell adhesion (Kasinrerk et al., 1999; Khunkeawla et al., 2001). In addition, CD147 also is involved in the invasion and metastasis processes of tumor cells in many types of cancers (Reimers et al., 2004; Tang et al., 2005; Nabeshima et al., 2006a). An increase in expression of CD147 in cancer cells has not only been linked to deregulation of epidermal growth factor receptor (EGFR) signaling (Menashi et al., 2003) but may also be a response to transforming growth factor-β stimulation (Gabison et al., 2005). CD147, also named extracellular matrix metalloproteinase inducer (EMMPRIN), has been identified as a cell surface inducer of matrix metalloproteinases (MMPs) both in tumor and stromal cells (Sun and Hemler, 2001; Gabison et al., 2005; Yan et al., 2005). Furthermore, CD147 is also defined as a lymphocyte activation-associated molecule (Kasinrerk et al., 1992; Stockinger, 1997; Koch et al., 1999). In T cells a negative regulatory signal

arises from cross-linking of CD147 molecules and T cell regulation has been demonstrated (Igakura et al., 1996; Koch et al., 1999; Staffler et al., 2003; Chiampanichayakul et al., 2006). Recently, two anti-CD147 mAbs, M6-1E9 and M6-1B9, which react with the membrane-distal Ig domain, inhibited OKT3-induced T cell proliferation (Chiampanichayakul et al., 2006). These mAbs inhibited cell proliferation by delivery of a negative signal through CD147 to suppress CD25 and IL-2 expression (Chiampanichayakul et al., 2006). Therefore, developing a tool for studying the function of CD147 would be valuable.

Advances in antibody engineering have allowed the manipulation of the antibody segments containing the antigen-binding regions and generation of small fragments that can be expressed in cells. These recombinant proteins are called intracellular antibodies or intrabodies. They have being successfully applied, mainly in the single chain Fv antibodies (scFv) format, to inhibit the function of intracellular target proteins in specific cellular compartments (Marasco et al., 1993; Lobato and Rabbitts, 2004). scFvs are the smallest fragment of an IgG molecule capable of maintaining the antigen-binding specificity of the parental antibody. They have the capacity to inhibit the translocation of cell surface molecules from the ER to the cell surface as ER-intrabodies (Richardson and Marasco, 1995; Steinberger et al., 2000b; Jendreyko et al., 2003). Intrabodies offer an effective alternative to gene-based knockout technologies to study the protein functions (Stocks, 2004). This technique has more advantages than RNA interference (RNAi) technology since intrabodies possess a much longer active half-life compared to RNA, and are also much more specific to their target molecules (Cao and Heng, 2005; Heng et al., 2005) and

generally do not disrupt target gene transcription. Moreover, these gene knock-out and silencing techniques cannot be used for domain structure and function analysis, including analysis of post-translationally modified protein functions. A recent study demonstrated that intrabody-mediated down-regulation of major histocompatibility complex (MHC) class I reduces the immunogenicity of rat aortic EC (RAEC) which may provide a suitable alternative supply for the lining of vascular prostheses (Doebis et al., 2006). In particular, intracellular use of antibody fragments can offer an effective alternative to gene-based knockout technologies. Therefore, intrabodies can be used as the most powerful tools to study the function of human leukocyte surface molecule especially CD147.

Normally, intrabodies are often introduced into cells by transfection of expression vectors encoding the intrabodies (Lobato and Rabbitts, 2004). Viral and nonviral gene transfer systems are available. Retroviral vectors as well as nonviral gene delivery methods such as calcium phosphate coprecipitation, electroporation and liposomal transfection target only a small fraction of cell population after extended selection periods, many of which, however, are not appropriate for transduction of human lymphocytes. Furthermore, expression levels are often low and insertion site-dependent silencing of the transgene expression is a frequent predicament (Doerfler et al., 1997; Baum et al., 2003; Hacein-Bey-Abina et al., 2003). In addition, the genome of retroviruses and lentiviruses is small, limiting the size of exogenous genes that can be packaged and transferred to target cells by the derivative vectors. Conversely, adenoviral vectors are an attractive alternative since they can efficiently transduce both dividing and non-dividing cells, and achieve transgene expression within hours

(Nevins et al., 1997). Moreover, adenoviruses do not integrate into the host genome, thus leaving the genetic package of targeted cells unmodified. This results in reproducible gene expression levels and also eliminates any undesirable effects related to the site of integration, allowing the specific analysis of the transgene effects (Kay et al., 2001). This characteristic, together with their relative ease of preparation and purification, has led to their extensive use as gene vectors.

From these data, we purposed to generate the recombinant scFv-M6-1B9 against CD147 as soluble form and as intrabody and study the biological activity of these recombinants. The suppression of CD147 expression on cancer cell surface may offer a novel approach for treatment of metastatic tumors in the near future.

1.2 Literature review

1.2.1 The human CD147

1.2.1.1 CD147 structure

The human cell surface molecule CD147 was designated at the Sixth International Workshop and Conference of Human Leukocyte Differentiation Antigens (Stockinger, 1997). It is a glycoprotein of 50–60 kDa having typical features of a type I integral membrane protein of the immunoglobulin superfamily as shown in **Figure 1.1**. CD147 is also known as M6 antigen (Kasinrerk et al., 1992), extracellular matrix metalloproteinase inducer (EMMPRIN) (Biswas et al., 1995) or human basigin (BSG) (Miyauchi et al., 1991). In various species glycoproteins homologous to CD147 have been identified, i.e. the rat protein OX-47/CE9 (Fossum et al., 1991; Nehme et al., 1993), the chicken blood–brain barrier-related molecule

HT7/neurothelin/5A11 (Schlosshauer and Herzog, 1990; Seulberger et al., 1990; Fadool and Linser, 1993), the mouse protein gp42/basigin (Altruda et al., 1989; Miyauchi et al., 1990) and the rabbit homologue (Schuster et al., 1996). The CD147 gene, designated BSG for basigin, is located on chromosome 19p13.3 (Muramatsu and Miyauchi, 2003) and encodes a 29 kDa protein, though migration on SDS-PAGE usually occurs between 35 and 65 kDa, depending on the degree of glycosylation (Kasinrerk et al., 1992; Kirsch et al., 1997; Kasinrerk et al., 1999). Treatment of the protein with endoglycosidase F demonstrated that almost half of CD147 molecular weight was due to the addition of N-linked carbohydrates (Kasinrerk et al., 1992). The structure of CD147 is composed of the extracellular region with two immunoglobulin domains, a single transmembrane domain and a short cytoplasmic domain (Miyauchi et al., 1991; Biswas et al., 1995; Muramatsu and Miyauchi, 2003). The protein is composed of an extracellular domain of 185 amino acid (aa.), a 24 aa residue transmembrane domain and a 39 aa cytoplasmic region as shown in Figure 1 (Kasinrerk et al., 1992; Biswas et al., 1995; Nabeshima et al., 2006b). The extracellular region contains three N-linked glycosylation sites, associated with 5–35 kDa glycosylation (glycan) content (Biswas et al., 1995; Tang et al., 2004a). The first extracellular Ig domain (EC I) is involved in matrix metalloproteinase (MMP) induction, binding to counter-receptors for itself (Sun and Hemler, 2001), carrying high-mannose-type L3 epitope (Heller et al., 2003) and association with integrins (Berditchevski et al., 1997). The second extracellular Ig domain (EC II) is required for association with caveolin-1 which leads to decreased self-association on the cell surface (Muramatsu and Miyauchi, 2003). The single transmembrane domain (TD)

sequence is completely conserved among human, mouse and chicken species. It may serve critical biological functions (Miyauchi et al., 1991). The presence of Pro211 and Glu218, in the middle of this domain are involved in association with cyclophilin 60 (Cyp60) and membrane targeting of CD147, respectively (Pushkarsky et al., 2005; Yurchenko et al., 2005). Similar to the leucine zipper motif, 3 leucines are repeated every seventh amino acid residue in the CD147 transmembrane domain (Seulberger et al., 1990; Fossum et al., 1991). This motif may be involved in the association between membrane proteins. By using fluorescence resonance energy transfer (FRET), Wilson *et al*. revealed that a cytoplasmic domain (CD) of CD147 binds to two molecules of monocarboxylate transporter 1 (MCT1) (Wilson et al., 2002).

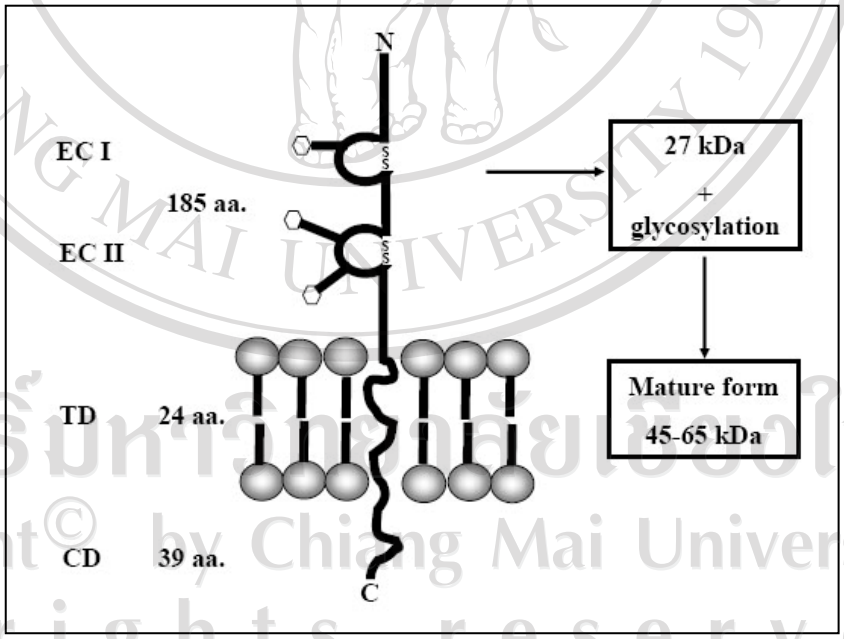


Figure 1.1 Schematic representation of the CD147 structure. EC I, first extracellular Ig domain; EC II, second extracellular Ig domain; TD, transmembrane domain; CD, cytoplasmic domain. Three N-linked oligosaccharides are shown by hexagons.

1.2.1.2 CD147 expression and their functions

The CD147 molecule is broadly expressed on human peripheral blood cells, endothelial cells, hemopoietic and non-hemopoietic cell lines (Stockinger, 1997). In T cells, its expression level is dependent on the differentiation state. Thymocytes are more strongly positive than mature T cells (Kirsch et al., 1997) and CD147 is up-regulated upon activation (Kasinrerk et al., 1992). Significant expression of CD147 has also been reported in malignant cells (Toole, 2003; Yan et al., 2005). The molecular function of neither CD147 nor any of its species homologues is fully understood. However, it has been suggested that CD147 is involved in signal transduction and cell adhesion functioning either directly as a signal transmitting adhesion molecule or as a regulator of adhesion.

A function of CD147 is also defined as a lymphocyte activation-associated molecule (Kasinrerk et al., 1992; Stockinger, 1997; Koch et al., 1999). In T cells a negative regulatory signal arises from cross-linking of CD147 molecules and T cell regulation has been demonstrated (Igakura et al., 1996; Koch et al., 1999; Staffler et al., 2003; Chiampanichayakul et al., 2006). Recently, anti-CD147 monoclonal antibodies (mAbs) *i.e.* M6-1B9 (IgG3), M6-2B1 (IgM), M6-1D4 (IgM), M6-1E9 (IgG2a), M6-1F3 (IgM) and M6-2F9 (IgM) have been generated (Kasinrerk et al., 1999). Some CD147 mAbs (M6-1D4, M6-1F3 and M6-2F9) induced homotypic aggregation of U937 cell line (Kasinrerk et al., 1999). Cell aggregation induced by the engagement of CD147 using mAbs to CD147 was described as a LFA-1/ICAM-1-dependent pathway (Kasinrerk et al., 1999). Furthermore, CD147 mAbs induced cell

aggregation depends upon the activation of protein kinases and a functional cytoskeleton (Khunkeawla et al., 2001). Two anti-CD147 mAbs, M6-1E9 and M6-1B9, which react with the membrane-distal Ig domain inhibited OKT3-induced T cell proliferation but did not induce cell aggregation (Chiampanichayakul et al., 2006). These mAbs inhibited cell proliferation by delivery of a negative signal through CD147 to suppress CD25 and IL-2 expression (Chiampanichayakul et al., 2006). In addition, CD147 was also demonstrated to induce apoptosis in U937 cells and that at least a portion of this cell death program involves a caspase-dependent pathway (Intasai et al., 2006). CD147 mAbs inhibited homotypic aggregation of the estrogen-dependent breast cancer cell line MCF-7, as well as MCF-7 cell adhesion to type IV collagen, fibronectin and laminin has been reported (Staffler and Stockinger, 2000) .

CD147 was also found to co-precipitate with α3β1 and α6β1 integrins and to co-localize with these integrins in areas of cell-cell contact (Berditchevski et al., 1997). Moreover, Tang *et al.* reported that using anti-integrin α3β1 antibody can decrease the enhancing effect of CD147 on adhesion, invasion capacities and secretion of matrix metalloproteinases (MMPs) in human 7721 hepatoma cells (Tang et al., 2008).

CD147 is enriched on the surface of tumor cells and stimulates adjacent stromal cells to produce several matrix metalloproteinases (MMPs), including interstitial collagenase (MMP-1), gelatinase A (MMP-2) and stromelysin-1 (MMP-3) (Majmudar et al., 1994; Biswas et al., 1995; Heppner et al., 1996; Guo et al., 1997; Lim et al., 1998). CD147 was termed as extracellular matrix

metalloproteinase inducer (EMMPRIN) (Sun and Hemler, 2001; Gabison et al., 2005; Yan et al., 2005) and is involved in the invasion and metastasis processes of tumor cells in many types of cancers (Reimers et al., 2004; Tang et al., 2005; Nabeshima et al., 2006a). An increase in expression of CD147 in cancer cells has not only been linked to deregulation of epidermal growth factor receptor (EGFR) signaling but may also be a response to transforming growth factor-β stimulation (Gabison et al., 2005). In addition, CD147 and P-glycoprotein (P-gp) is also overexpressed in multi-drug resistance (MDR) cancer cell lines (Yang et al., 2003; Yang et al., 2007; Jia et al., 2008). Increased production of MMP-2, MMP-9, and MMP-11 was observed only in MDR cancer cells (Li et al., 2007; Wang et al., 2008). Thus, CD147 and P-gp may play a crucial role in MDR cancer cells with the invasiveness and metastasis of tumor cells.

CD147 is reported to exist in both soluble and membrane bound forms (Sidhu et al., 2004; Tang et al., 2004b). Interestingly, several studies have provided evidence that microvesicular release of CD147 from tumor cells could play a role in tumor-stromal interactions through upregulation of MMP production (Sidhu et al., 2004). Shedding of membrane vesicles is observed in eukaryotic cells and suggested to be involved in several pathophysiological processes (Taraboletti et al., 2002). These shed vesicles are unstable and rapidly degraded over time, giving rise to increasing levels of CD147 in non-vesicle fractions (*i.e.* supernatant) of cell-conditioned medium. Recently, Millimaggi *et al.* demonstrated that tumor vesicle-associated CD147 cells could modulate the angiogenic capability of human umbilical vein endothelial cells (HUVECs) (Millimaggi et al., 2007).

Despite the interaction of mAbs and their targets mimicked the native ligand-receptor signaling. However, substrates or ligands for CD147 are still not identified. It is not clear whether CD147 is directly involved in signal transduction and cell adhesion as a signal transmitting adhesion molecule or as a regulator of adhesion in the interactions between epithelial cells and the extracellular matrix.

1.2.2 Recombinant protein expression in *Escherichia coli*

E. coli is one of the most widely used hosts for the heterologous protein production. This is due to its ability to grow rapidly and at high density on inexpensive substrates, its well-characterized genetics and the availability of an increasingly large number of cloning vectors and mutant host strains. The genetics of *E. coli* are far better characterized than those of any other microorganism. Recent progress in the fundamental understanding of transcription, translation, and protein folding in *E.* coli, together with serendipitous discoveries and the availability of improved genetic tools are making this bacterium more valuable than ever for the expression of complex eukaryotic proteins. Although there is no guarantee that a recombinant gene product will accumulate in *E. coli* at high levels in a full-length and biologically active form, a considerable amount of effort has been directed at improving the performance and versatility of this workhorse microorganism (Makrides, 1996; Baneyx and Mujacic, 2004).

1.2.2.1 Protein translocation across bacterial cytoplasmic membrane

The targeting and transport of proteins across biological membranes 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 cell. Nucleases, peptidases 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), followed by the c-region, which normally contains a recognition sequence for the enzyme signal peptidase. In bacteria, three major routes are used to achieve protein translocation across the cytoplasmic membrane *e.g.* the secretory (Sec), signal recognition particle (SRP) and the twin-arginine translocation (Tat) pathway as shown in **Figure 1.2**.

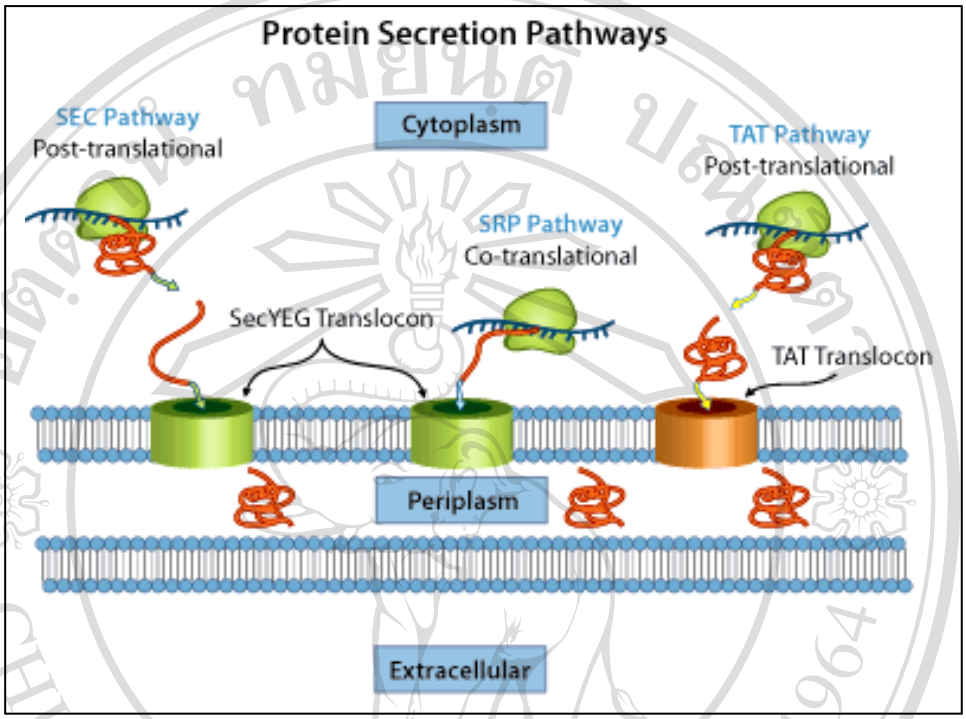


Figure 1.2 Diagram of the secretion pathways of *E. coli*.

(http://www.athenaes.com/tech_brief_ACESyebf.php)

1.2.2.1.1 The general secretory (Sec) pathway

The major route of protein translocation in bacteria is the so-called general secretion pathway (Sec pathway) (Choi and Lee, 2004). This transporter complex is build by multimeric proteins which are spanning through the inner membrane of *E. coli*. Secretion *via* this pathway involves 9 different components: Trigger factor (TF), SecA, SecB, SecY, SecE, SecG, SecD, SecF and YajC. The core of the translocase consists of a proteinaceous channel formed by the protein complex of SecYEG and the peripheral adenosine triphosphatase (ATPase) SecA as molecular motor. Ribosome-associated nascent chains of secreted proteins bind TF (**Figure 1.3**), which is bound to the ribosomes (Maier et al., 2003). This association is maintained until the preprotein leaves the ribosome, thus preventing cotranslational binding of the nascent chain to SRP components (Mergulhao et al., 2005).

Secreted proteins targeted to the SecB-dependent pathway contain an amino-terminal signal peptide that functions as a targeting and recognition signal. These signal peptides are usually 18–30 amino acid residues long and are composed of a positively charged amino terminus (n-region), a central hydrophobic core (h-region), and a polar cleavage region (c-region) (Fekkes and Driessen, 1999; Choi and Lee, 2004). The n-region is believed to be involved in targeting the preprotein to the translocase and binding to the negatively charged surface of the membrane lipid bilayer. Increasing the positive charge in this region has been shown to enhance translocation rates, probably by increasing the interaction of the preprotein with SecA (Fekkes and Driessen, 1999). The h-region varies in length from 7 to 15 amino acids.

Translocation efficiency increases with the length and hydrophobicity of the h-region, and a minimum hydrophobicity is required for function (Wang et al., 2000).

Secreted proteins are kept in a translocation-competent state by the chaperone SecB (de Gier and Luirink, 2001), which interacts with the mature region of the preprotein to prevent premature folding (Khokhlova and Nesmeianova, 2003) and targets it to SecA (Figure 3). In the presence of preprotein, SecB binds SecA (Fekkes et al., 1998; Woodbury et al., 2000), thus releasing the precursor protein that is transferred to SecA (Fekkes and Driessen, 1999). SecA binding to the preprotein is facilitated by the signal peptide, which it recognizes specifically. At this point SecA is bound to the SecY subunit of the SecYEG complex. SecYEG constitutes a pathway ('channel') for polypeptide movement. Binding of ATP at one of the two ATP-binding sites on SecA causes the release of SecB from the membrane (Mergulhao et al., 2005). There is no consensus on how the Sec components form a functional translocon, and monomeric, dimeric and oligomeric translocons have been proposed (Mergulhao et al., 2005). Binding of the preprotein to membrane-bound SecA results in the translocation of approximately 20 amino acids, and subsequent binding of ATP to SecA promotes SecA membrane insertion and translocation of additional 15-20 amino acids. ATP hydrolysis releases the preprotein from SecA into the translocation channel (Driessen et al., 1998). ADP is then released and SecA deinserts from the membrane where it can be exchanged with cytosolic SecA. Multiple rounds of SecA insertion and deinsertion promote protein translocation through the channel (de Keyzer et al., 2003). Proton-motive force (PMF) can complete translocation when the preprotein is halfway through the translocase, even in the absence of SecA

(Nishiyama et al., 1999). The mechanism by which PMF drives translocation is unknown but it has been suggested that PMF assists in the initiation phase of protein translocation (Mori and Ito, 2003) and that it accelerates SecA membrane deinsertion (Nishiyama et al., 1999). Finally, the folded substrate protein was released into the periplasmic space. SecD, F and YajC are accessory proteins that aid in translocation into the periplasm, where proteins are folded into their final confirmation.

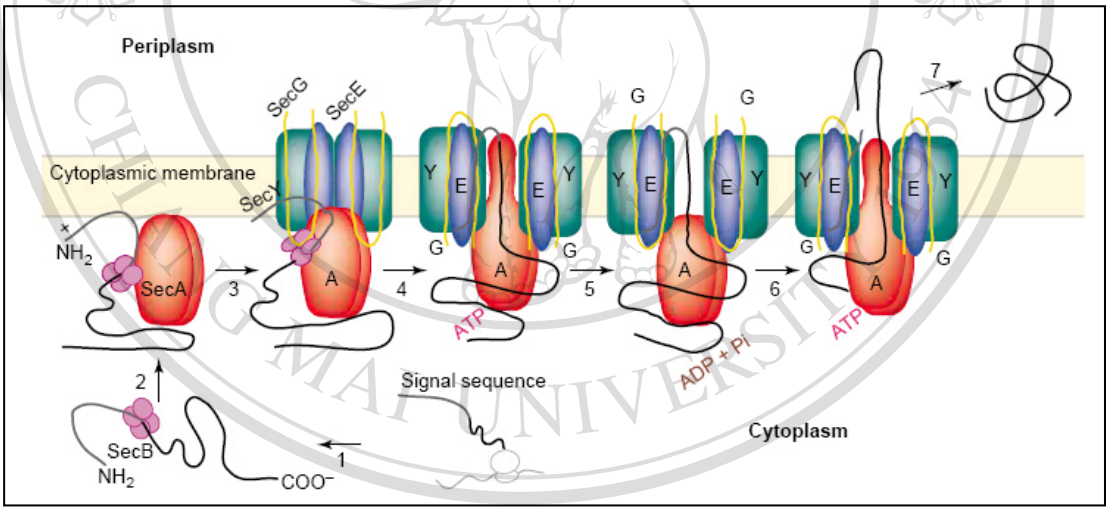
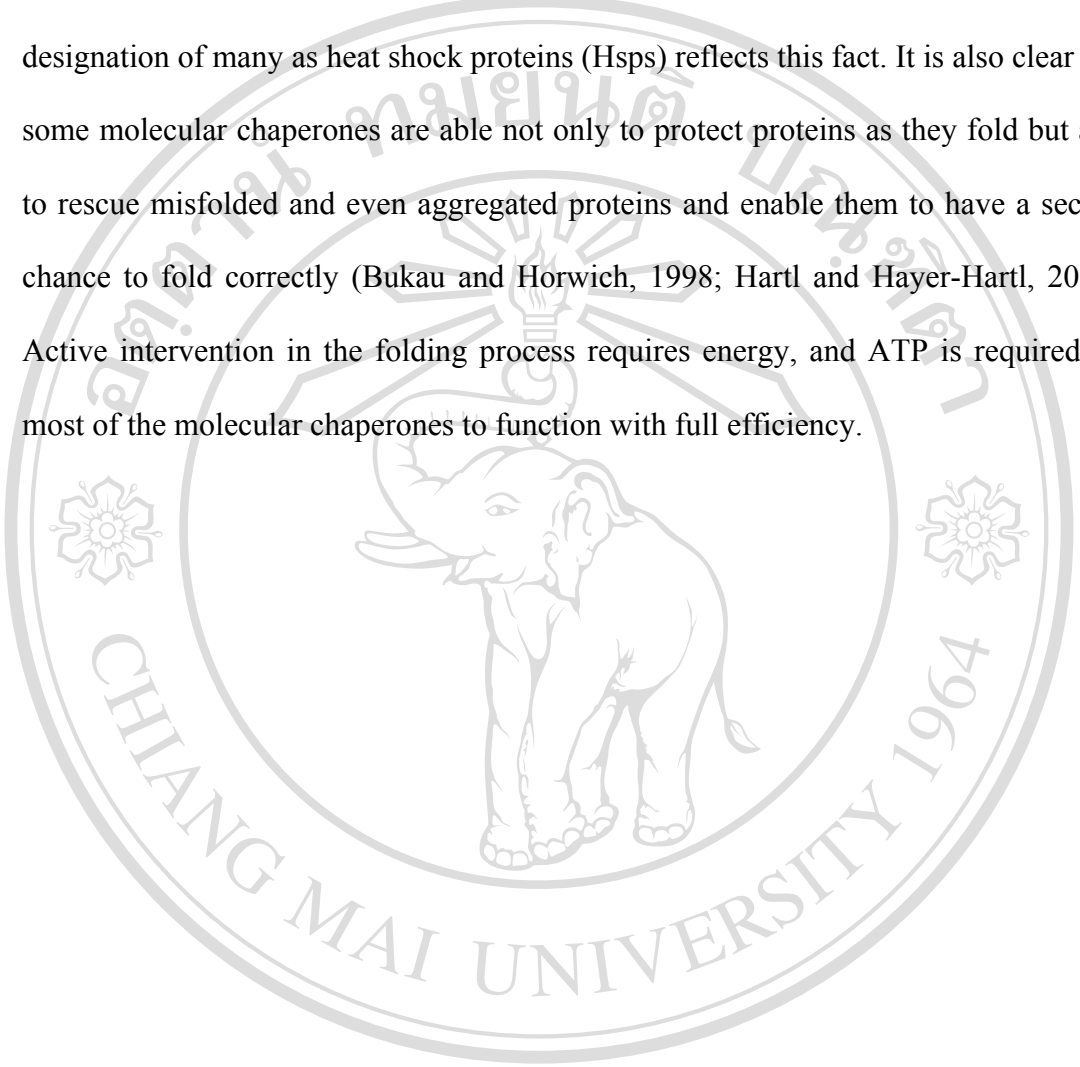


Figure 1.3 The general Sec pathway in bacteria (Mori and Ito, 2001).

1.2.2.2 Protein folding in *E. coli*

Newly synthesized polypeptide chains must fold and assemble into unique three-dimentional 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 overproduced recombinant proteins often require the assistance of folding modulators which include molecular chaperones and folding catalysts. Molecular chaperones are an 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 protein acts 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. Molecular chaperones can be divided into three functional subclasses based on their mechanism of action (**Figure 1.4**). Folding chaperones (e.g. DnaK and GroEL) are the core of the chaperone network which uses conformational changes fueled by ATP hydrolysis to promote the folding of bound substrates. Holding chaperones (e.g. IbpA, IbpB, Hsp31 and Hsp33) stabilize partially folded proteins on their surface to await availability of folding chaperones upon stress abatement. Finally, the disaggregating chaperone ClpB promotes the solubilization of proteins that have become aggregated as a result of stress and transfer them to the folding chaperones for subsequent refolding (Baneyx and Mujacic, 2004). Clear evidence that molecular chaperones are needed to prevent misfolding and its consequences come from the fact that the concentrations of many

of these species are substantially increased during cellular stress; indeed, the designation of many as heat shock proteins (Hsps) reflects this fact. It is also clear that some molecular chaperones are able not only to protect proteins as they fold but also to rescue misfolded and even aggregated proteins and enable them to have a second chance to fold correctly (Bukau and Horwich, 1998; Hartl and Hayer-Hartl, 2002). Active intervention in the folding process requires energy, and ATP is required for most of the molecular chaperones to function with full efficiency.



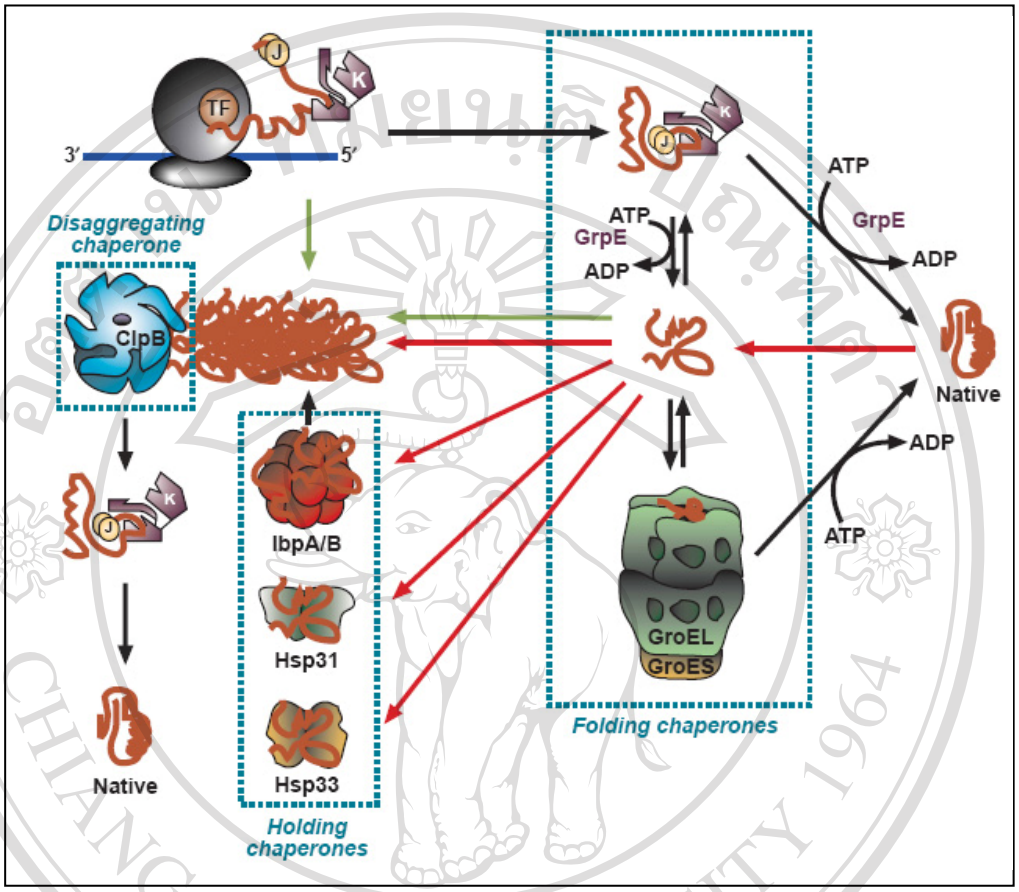


Figure 1.4 Chaperone-assisted protein folding in the cytoplasm of *E. coli*.

Nascent polypeptides requiring the assistance of molecular chaperones first encounter trigger factor (TF) or DnaK-DnaJ. In times of stress (red arrows), thermobile proteins unfold and aggregate. IbpB is required to serve partially folded proteins on its surface to serve as a reservoir of unfolded intermediates until folding chaperone available. Recombinant proteins that miss an early interaction with TF or DnaK-DnaJ, that undergo multiple cycles of abortive interactions with folding chaperones, or titrate them out, accumulate in inclusion bodies (green arrows) (Baneyx and Mujacic, 2004).

1.2.2.2.1 Periplasmic protein folding

The proteins that translocate across the inner membrane exert their biological functions in the periplasmic or destine for insertion into outer membrane. These proteins must transverse the periplasm and the peptidoglycan layer to reach their destination. During their 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 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).

The *E. coli* periplasm contains a series of enzymes such as disulfide-binding proteins (DsbA, DsbB, DsbC, and DsbD) and petidyl-prolyl isomerases (SurA, RotA, FklB, and FkpA) that promote the appropriate folding of thiol-containing proteins (Shokri et al., 2003). FkpA is the peptidyl-prolyl-*cis*/*trans* isomerase (PPIase), which catalyses 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 PPIase 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 recognizes an Ar-X-Ar motif (where Ar is an aromatic and X is any amino acid 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).

1.2.2.3 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 help in both the folding and stabilization of the mature protein. Disulfide bonds are generally restricted to secreted proteins and some membrane protein 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 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 glutaredoxinn/glutathione system (Holmgren, 1989). The proper formation of disulfide bonds in the periplasm is catalyzed by specific thiol-disulfide oxidoreductase. In both the periplasmic and cytoplasmic systems, the proteins responsible for catalizing 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 1.5**). 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 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 DsbC, a periplasmic disulfide bond oxidoreductase (Rietsch et al., 1996) .The Cys-X-X-Cys active site of DsbC 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).

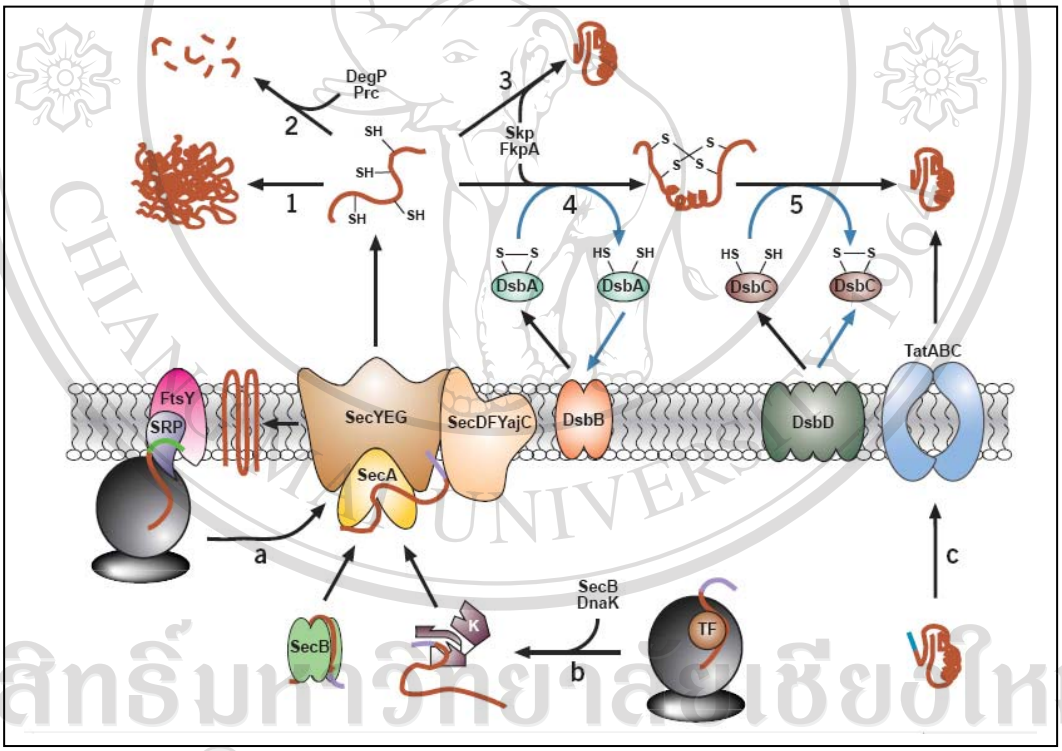


Figure 1.5 Disulfide bond formations in bacterial periplasm. Cysteine pairs in proteins containing disulfide bonds are oxidized by DsbA whereas incorrect disulfides are isomerized by DsbC. These oxidoreductases are reactivated by DsbB and DsbD, respectively (Baneyx and Mujacic, 2004).

1.2.2.4 Protein misfolding and inclusion body formation

Protein activity demands its folding into precise three-dimentional 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 overexpression 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 non-native intermolecular hydrophobic interactions between protein folding intermediates, which have not yet buried their hydrophobic amino acid residues (Kiefhaber et al., 1991). 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 steps 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 to allow 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).

1.2.2.5 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 *rho*H 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 (Liberek et al., 1992; Gamer et al., 1996). In contrast, the extracytoplasmic is believed to be controlled by at least two signal transduction system, 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 is composed of a sensory histidine kinase encoded by *cpxA* and a response regulator encoded by *cpxR* (Danese et al., 1995; Snyder et al., 1995). This pathway appears to regulate the expression of genes encoding the DsbA thiol-disulfide oxidoreductase, CpxA periplasmic protein, DegP proteinase and PpiD peptidyl-prolyl-*cis*/*trans* isomerases (Danese and Silhavy, 1997; Pogliano et al., 1997). The periplasmic 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 is predicted to lead to relief of CpxP inhibition and activation of phosphotransfer between CpxA and cpxR. Phosphorylated CpxR upregulates expression of genes whose products 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).

1.2.3 Protein transporting in Eukaryotes

In contrast to most prokaryotic cells, eukaryotic cells contain, in addition to the plasma membrane, internal membranes. These internal membranes are structural components of organelles and vesicles. Most proteins are synthesized by cytosolic ribosomes and must pass a membrane to reach their final destination. Plasma membrane proteins and extracellular proteins are synthesized and processed at the rough endoplasmic reticulum. Therefore the initial step in secretion of most eukaryotic proteins is their transport into the lumen of the endoplasmic reticulum (ER) as shown in **Figure 1.6**. Transport of presecretory proteins into the ER involves cleavable signal peptides at the amino terminus of the precursor proteins and a transport machinery which operates co- or post-translationally. Typically this transport occurs as a sequence of three consecutive steps: i) specific membrane association of the precursor protein, ii) membrane insertion, and/or iii) completion of translocation (**Figure 1.7**).

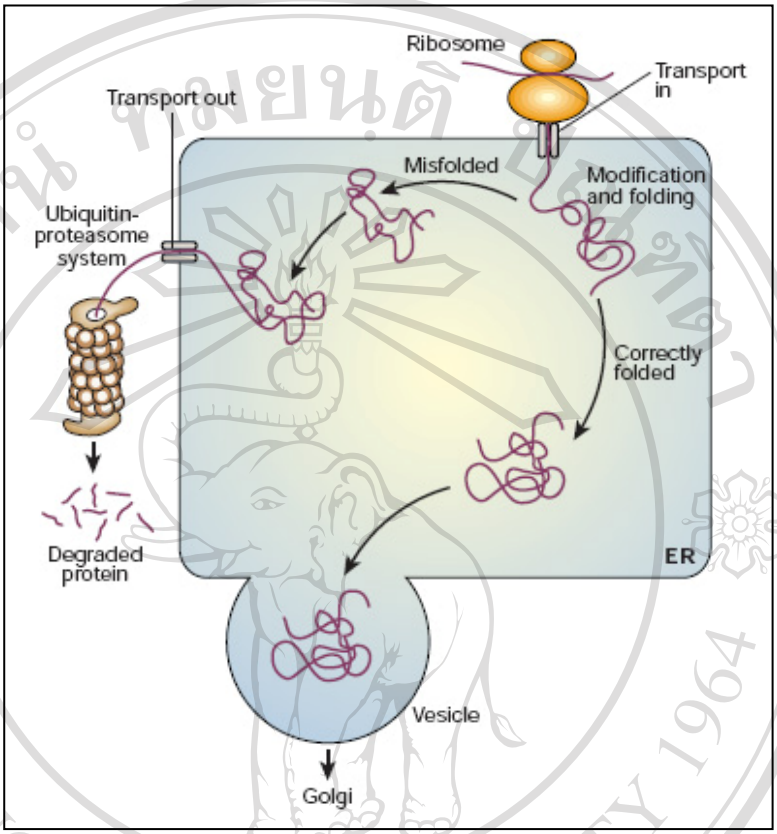


Figure 1.6 Regulation of protein folding in the ER (Dobson, 2003).

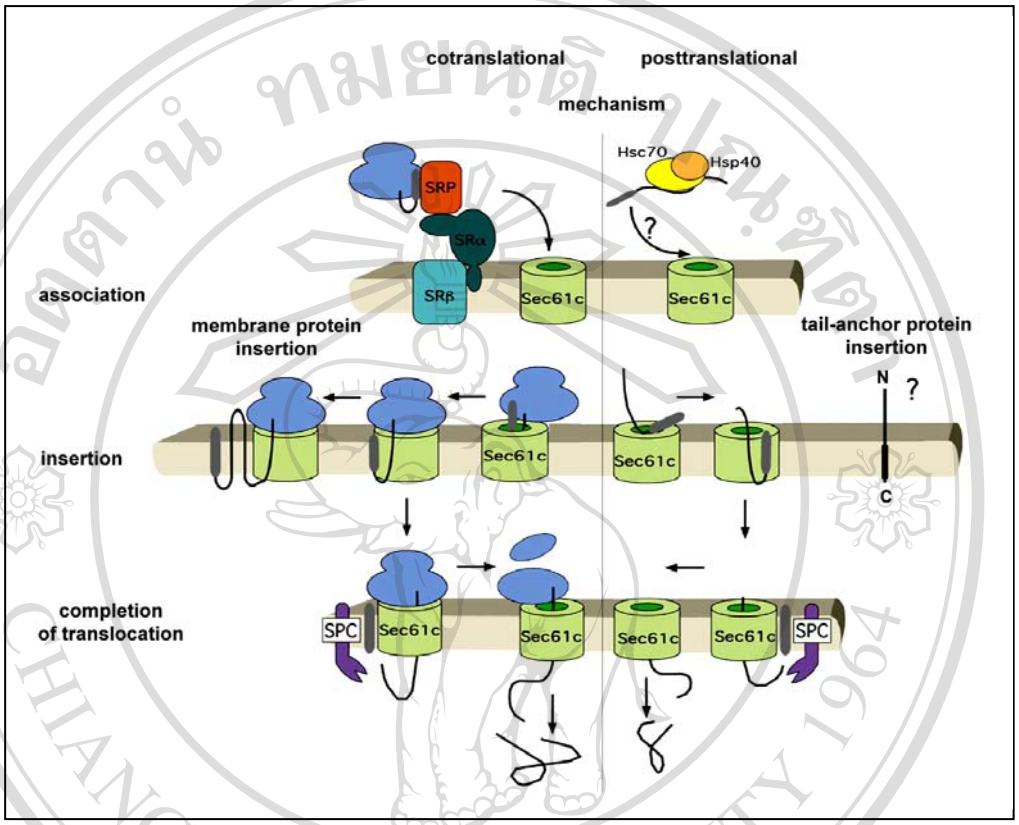


Figure 1.7 Protein transport- and insertion-pathways at the ER-membrane.

(http://wwwalt.med-rz.uni-sb.de/med_fak/biochemie/MartinJ/Eninformation.html)

It is well established that proteins destined for the secretory pathway begin their nascent polypeptide chains on the ribosomes and that SRP (signal recognition particle) is required to bring these chain bearing ribosomes to the ER membrane for translocation. There are several proteineaous factors located at the ribosomal polypeptide exit site which interact or bind directly to the emerging polypeptide chain, for example the nascent polypeptide associated complex (NAC) and cytosolic molecular chaperones (Hsc70, Hsp40). Therefore SRP-dependent cotranslational transport needs a specific signal sequence in the nascent polypeptide chain for successful targeting to the ER-membrane. The ER signal peptides have features similar to those of their bacterial counterparts. The eukaryotic signal sequences typically contain three regions: a short and positively charged hydrophilic amino-terminal segment, a central hydrophobic part (7-15 residues) and a more polar carboxy-terminal region. This signal sequence is recognized by SRP, a ribonucleoprotein complex consisting of a 7S RNA and six different polypeptide subunits. SRP interacts through its 54 kDa subunit with the signal sequence of nascent polypeptide chains and directs the entire complex, consisting of the ribosome, the nascent chain and the SRP to the ER membrane. Elongation of the polypeptide chain is delayed or even arrested until SRP is bound to the SRP receptor, an integral protein complex of the ER membrane. Once bound, SRP is released from the signal sequence in a GTP-dependent manner, and the ribosome together with the nascent chain is then passed onto the protein translocase. The continuing translation inserts the polypeptide into the lumen of the ER. Cleavable signal sequences are cotranslationally processed by the signal peptidase complex (SPC) at the lumenal side of the ER membrane

(Paetzel et al., 2002). Co- or post-translational proteins are translocated to the Sec61p complexes (**Figure 1.8**). The Sec61p complex seems to be the main structural component of the protein conducting channel. In mammals, the Sec61p complex contains a Sec61α-subunit with ten membrane-spanning domains, and the Sec61β- and Sec61γ-subunit, which belong to the class of tail-anchored proteins. In addition to these heterotrimeric Sec61p complexes, ATP-binding proteins of the mammalian ER lumen are part of protein translocase which facilitates insertion of presecretory proteins into the Sec61p complex as well as completion of translocation. The ATP-binding proteins (BiP, Grp170) were identified as ER resident members of the Hsp70 protein family. These Hsp70 protein family members of the mammalian ER may be recruited to the Sec61p complex by either one or both of the two membrane integrated Hsp40 protein family members, Sec63p and Mtj1p.

The ER contains a wide range of molecular chaperones and folding catalysts, and in addition the proteins that fold here must satisfy a quality-control check before being exported (Hammond and Helenius, 1995; Kaufman, 2002). Such a process is particularly important because there seem to be few molecular chaperones outside the cell, although one (clusterin), at least, has recently been discovered (Wilson and Easterbrook-Smith, 2000). This quality-control mechanism involves a remarkable series of glycosylation and deglycosylation reactions that enables correctly folded proteins to be distinguished from misfolded ones (Hammond and Helenius, 1995). Correctly folded proteins with the help of a series of molecular chaperones and folding catalysts are then transported to the Golgi complex and

proceed to their final destinations *via* the secretory pathway (Lewin et al., 2007). However, incorrectly folded proteins are detected by a quality-control mechanism and sent along another pathway (the unfolded protein response) in which they are ubiquitinated and then degraded in the cytoplasm by proteasomes (Schubert et al., 2000).

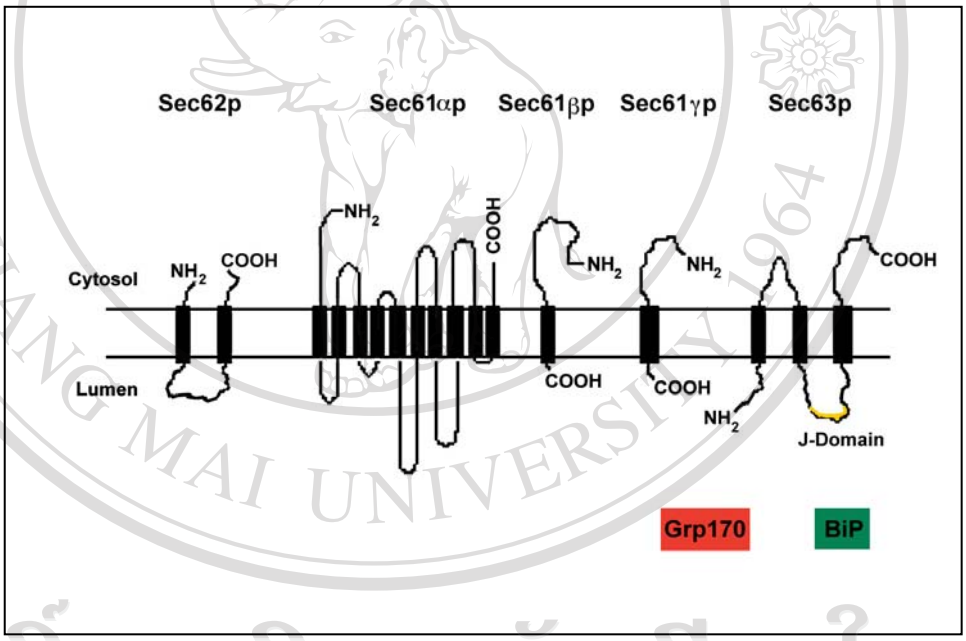


Figure 1.8 Components of the protein translocase of the ER-membrane.

(http://wwwalt.med-rz.uni-sb.de/med_fak/biochemie/MartinJ/Eninformation.html)

1.2.4 Phage display technology

1.2.4.1 Introduction

Phage display was first developed with the *E. coli* specific bacteriophage M13 (Smith, 1985), and the success of M13 phage display has prompted the development of numerous alternative display systems (Ren and Black, 1998; Santini et al., 1998). Phage display is a powerful method for selecting and engineering polypeptides with desired binding specificities. This technology can be applied in the field of immunology, cell biology and pharmaceutical biotechnology (Sidhu, 2000). Displaying of peptides and gene fragments enables the analysis of protein-protein interactions such as structural mapping of epitopes (Tayapiwatana et al., 2003; Abbasova et al., 2007), characterization of receptor and ligand interaction (Jager et al., 2007; Mohrluder et al., 2007; Casey et al., 2008), functional analysis (Intasai et al., 2006; Yang et al., 2006) and immunodiagnosis (Robles et al., 2005; Hell et al., 2009). This method is accomplished by inserting the gene fragments encoding the protein of interested into a phagemid genome as a fusion with M13 coat protein genes. These fusion genes can be incorporated in bacteriophage particles that also display the heterologous proteins on their surfaces. In this way, a physical linkage is established between phenotype and genotype of the expressed protein.

1.2.4.2 Biology and structure of M13 filamentous bacteriophage (Smith and Petrenko, 1997; Barbas CF. III, 2001; Barbas, 2001; Sidhu, 2001)

The filamentous phage constitutes a large family of bacterial viruses that infect a variety of Gram-negative bacteria, using pili as receptors. The best characterized are the very similar phages M13, fd and f1, that infect *E. coli via* F pili. The filamentous bacteriophages are a group of viruses that contain a circular single-stranded DNA genome encased in a long protein capsid cylinder. The Ff phage particles are flexible rods about 1 μm long and less than 10 nm in diameter. The mass of particle is approximately 16.3 MD, of which 87% is contributed by protein. The particle consists of a single-stranded DNA core surrounded by a proteinaceous coat. The length of the cylinder consists of approximately 2,700 molecules of 50 amino acid major coat protein, called gene VIII protein (gpVIII). The four minor coat proteins are present at about 5 copies per particle; protein-VII and protein-IX (gpVII and gpIX) cap one end of the particle while protein-III and protein-VI (gpIII and gpVI) cap the other end (**Figure 1.9** and **Table 1.1**). Each of the five different coat proteins has been successfully used as a platform for the functional display of heterologous polypeptides as either N- or C-terminal fusions. Protein VI and gpIII are crucial for host recognition and phage infectivity, whereas gpVII and gpIX are required for phage assembly (Gailus et al., 1994). The gpIII is the most commonly used coat protein for display. This protein is made up of three domains separated by glycine-rich regions. These three domains have been designated N1 or D1, N2 or D2 and CT or D3 by different groups. The first domain, N1, is required during infection

for the translocation of the DNA into the cytoplasm and insertion of the coat proteins into membrane. N2 is responsible for binding to F pilus (Deng et al., 1999). The carboxy-terminal end makes up the CT domain and is essential for forming a stable phage particle. The DNA is oriented within the virion such that a 78 nucleotide hairpin region called the packaging signal (PS) is always located at the end of the particle containing the gpVII and gpIX proteins.

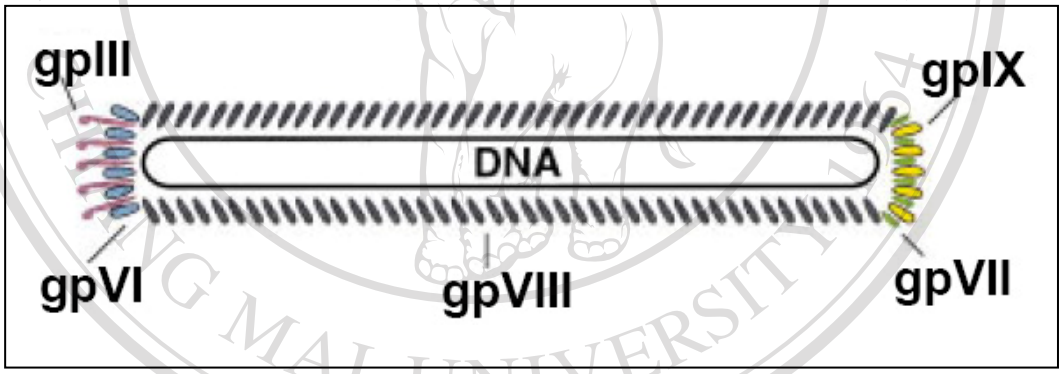


Figure 1.9 Structure of a filamentous bacteriophage. A diagram of the bacteriophage particle represents the single-stranded DNA core surrounded by a proteinaceous coat (Sidhu, 2001).

Table 1.1 Phage coat proteins.

Protein	Number of amino acids	Molecular weight	Copies per phage	Function
gpIII	406	42,500	~5	Minor capsid protein
gpVI	112	12,300	~5	Minor capsid protein
gpVII	33	3,600	~5	Minor capsid protein
gpVIII	50	5,200	~2,700	Major capsid protein
gpIX	32	3,600	~5	Minor capsid protein

1.2.4.3 M13 filamentous bacteriophage life cycle

Infection of *E. coli* by a filamentous phage is a multistep process requiring interactions of the phage gpIII protein with the F conjugative pilus and the bacterial TolQ, R and A cytoplasmic membrane proteins. These three Tol proteins are bacterial proteins that appear necessary for maintaining the integrity of the bacterial outer membrane, to avoid a leak of periplasmic proteins into the culture medium (Lazzaroni et al., 1999). The TolQRA proteins are required during phage infection for translocation of the Ff phage DNA into the cytoplasm and translocation of the phage coat proteins into the cytoplasmic membrane (Russel et al., 1988; Russel et al., 1997). Infection is initiated by attachment of the N2 domain of gpIII to the tip of the pilus; this is the end of the particle that enters the cell first (Holliger and Riechmann, 1997; Witty et al., 2002; Karlsson et al., 2003). This binding releases N1 from N2, and allows N1 to interact with TolA. As the process continues, the coat proteins dissolve into the periplasm. The major coat protein, gpVIII, gpVII and gpIX minor capsid proteins, and the uncoated ssDNA concomitantly enters the cytoplasm (**Figure 1.10**). Once the viral (+) strand DNA enters the cytoplasm, the complementary DNA (-) strand is synthesized by bacterial enzymes, resulting in a covalently closed, supercoiled and double-stranded replicative form (RF). The (-) strand of this RF is a template for transcription and consequencing mRNAs are translated into all of 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, gpVII, gpVIII, gpIX) are components of the phage particle. After one round, gpII circularizes the displace viral (+) strand DNA, which then is converted to

a covalently closed, supercoiled and double-stranded RF molecule by bacterial enzyme. The 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. The gpX is crucial for the proper replication of the phage DNA and functions as an inhibitor of gpII (Fulford and Model, 1984). The RF replicates to make progeny RFs and is also the template for transcription of phage genes and synthesis of progeny ssDNAs. Assembly occurs at specific sites in the bacterial envelope where the cytoplasmic and outer membranes are in close contact by the interaction of gpI, gpVI and gpXI and form a gated pore complex that spans the inner and outer membranes. Phage assembly is initiated by the incorporation of gpVII and gpIX at one end of the particle. This process continues until the end of the DNA is reached and the assembly is terminated by the incorporation of gpVI and gpIII (**Figure 1.11**). Progeny virions are secreted continuously without lysis of the *E.coli* host; chronically infected cells continue to divide, though at a slower rate than uninfected cells.

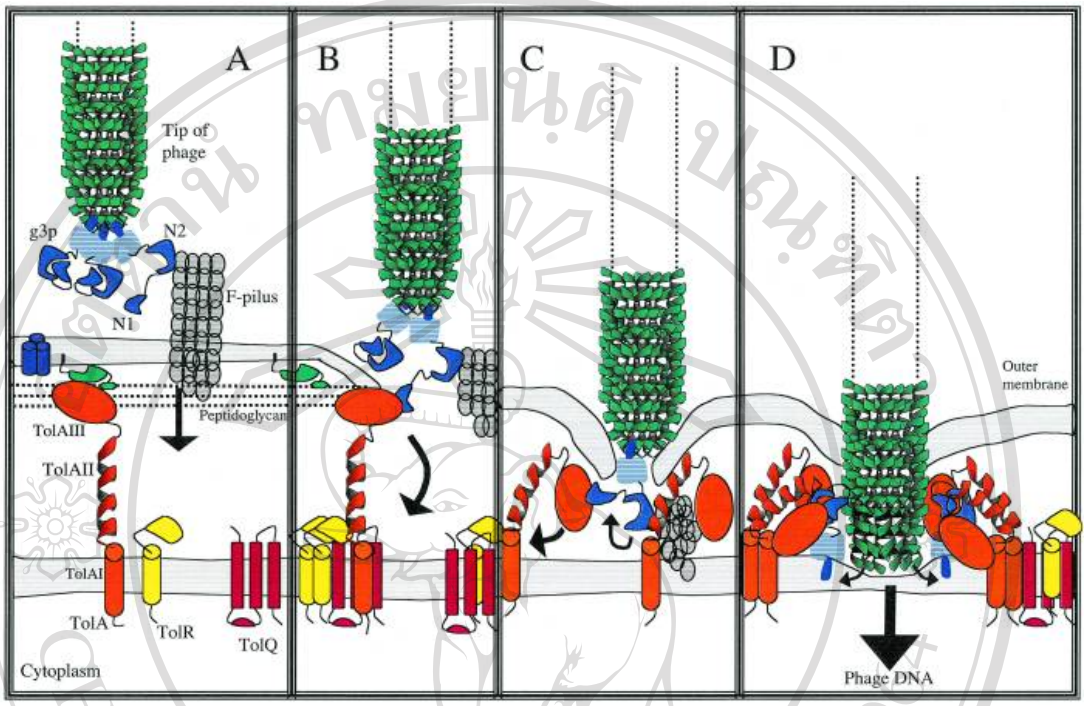


Figure 1.10 Model of filamentous phage infection. A) The phage N2 domain of pIII protein interacts with the F-pilus on the outside of the bacteria. The outer membrane proteins OmpF (blue cylinders) and Pal lipoprotein (greenish) are also included, as there have been reports of TolA interacting with these proteins prior to infection (Lazdunski et al., 1998; Cascales et al., 2000). **B)** After F-pilus retraction, N1 domain of gpIII binds to the C-terminal domain of bacterial TolA domain III (TolAIII). **C)** The retracting pilus brings phage gpIII domains in closer contact with TolA domains. As a consequence, TolA can assume a more compact state of assembly, thus bringing the outer and inner membranes of the bacteria closer together. At this stage, the central domain of TolA (TolAII) has the possibility to interact with the N2 domain of gpIII. **D)** The phage gpIII is inserted into the inner membrane, and the cap of the phage head is opened to allow phage DNA to enter the bacteria (Karlsson et al., 2003).

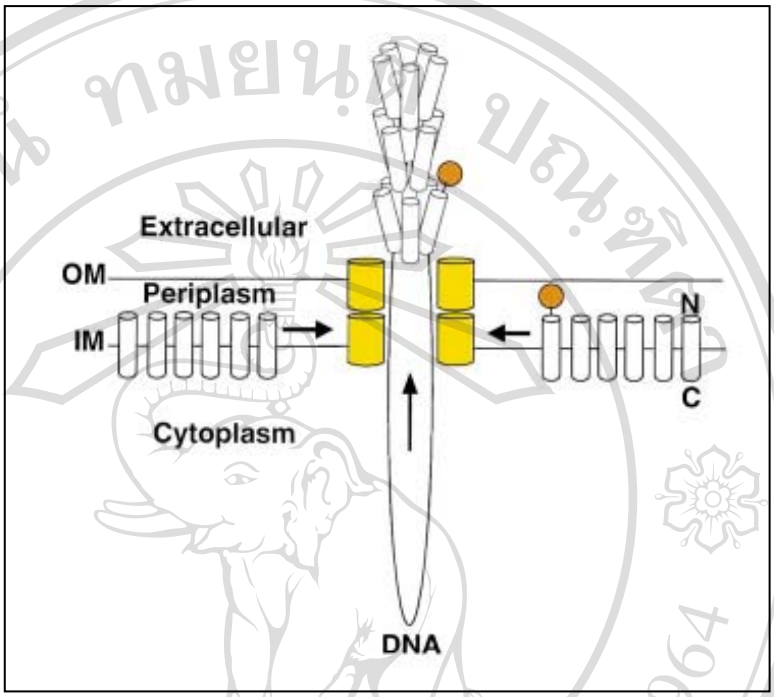


Figure 1.11 M13 bacteriophage assembly. Newly-synthesized coat proteins (white cylinders) are embedded into the inner membrane (IM) with their N termini in the periplasm and their C termini in the cytoplasm. Single-stranded viral DNA is extruded through a pore complex (yellow cylinders) that spans the inner membrane and the outer membrane (OM). Coat proteins also interact with the pore complex, where they surround the DNA and are thus transferred from the bacterial membrane into the assembling phage coat. The assembled phage particle is extruded to the extracellular environment. A heterologous protein (red circle) will be displayed on the phage surface if it is fused to a coat protein that can successfully incorporate into the phage coat (Sidhu, 2001).

1.2.4.4 Phagemid vector

Protein can be displayed using vectors based on the natural filamentous phage sequence (phage vector) or using plasmids that contain only the fusion phage gene and no other phage genes (phagemid vectors) (Lowman, 1997). Phagemids, a more popular vector for display, are designed to contain the origins of replications for both the M13 phage and *E. coli*, appropriate multiple cloning sites for insertion the gene of interest, and an antibiotic-resistance gene for selection and propagation as with typical phagemids (Barbas, 2001). Phagemids can be grown as plasmids or alternatively packaged as recombinant M13 phage with the aid of a helper phage that contains a slightly defective origin of replication (such as M13KO7 or VCSM13) and supplies, *in trans,* all the structural proteins and enzymes required for generating a complete phage. They also carry a kanamycin resistance gene to allow antibiotic selection of helper-infected cells. Thus almost the phage particles may incorporate either gpIII derived from the helper phage or the polypeptide-gpIII fusion protein, encoded by the phagemid. A major advantage of phagemid vectors is their small size and ease of cloning, compared with the difficulties of cloning in phage vector without disruption of the structure of gene and promoter. In addition, the large DNA inserts are more readily maintained by phagemid genomes than phage genomes. Two-gene display systems (Type 3+3 and 8+8 phagemid and Type 33 and 88 phage systems) allow modulation of the valency (i.e., the number of copies per phage particle, Table 2) of the displayed fusion protein. However, the ratios of polypeptide-gpIII fusion protein: wild type gpIII may range between 1:9 and 1:1000 depending on the type of

phagemid, growth conditions, the nature of the polypeptide fused to gpIII, and proteolytic cleavage of antibody-gpIII fusions (Azzazy and Highsmith, 2002).

Table 1.2 Classification of phage-display vectors.

Vector type	Coat protein used for display	Display on all or some copies of coat protein	Number of coat protein genes	Fusion encoded on genome of
3	pIII	All	1	phage
8	pVIII	All	1	phage
33	pIII	Some	2	phage
88	pVIII	Some	2	phage
3+3	pIII	Some	2	phagemid
8+8	pVIII	Some	2	phagemid

1.2.4.5 Applications of phage display

Since the phage display reported in 1985 (Smith, 1985), this technology has rapidly evolved into an efficient tool used by structural biologists for the discovery and characterization of diverse ligand-receptor-binding interactions. By this technique, the gene of interest is inserted between the C-terminal of the signal peptide and N-terminal coding region of phage coat protein. The recombinant proteins are synthesized in *E. coli* 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 proteins encoded by the corresponding gene fusion present inside the same phage particle. Individual phages can be rescued from libraries by an interaction of the displayed protein with the cognate ligand by a panning step, which allows to select of the hundreds of millions of clones, those few phages that display a peptide that binds the target molecule. These phages can be amplified by infection of bacteria. The recombinant polypeptides or proteins displayed on the phage surface can be used for identifying and characterizing the interaction with their binding targets (Cesareni, 1992).

Phage display of a functional protein has now become a standard first step of proof of principle prior to the application of combinatorial strategies using the cloned DNA template to evaluate or remold functional activity (Bass et al., 1990; Roberts et al., 1992). Even though a natural functional domain can represent the end

product of a highly directed evolutionary process, phage-display approaches can create variations of the domain with altered binding affinity or fine specificity, or with structural refinements that greatly enhance stability (Lowman et al., 1991; Lowman and Wells, 1993; Hao et al., 2008; Hoyer et al., 2008; Kwong et al., 2008; Berntzen et al., 2009; Hertveldt et al., 2009).

Advances in phage display and antibody engineering have led to the development of phage-displayed antibody technology (McCafferty et al., 1990; Sidhu, 2000). This technology allows one to isolate antibodies directly from diverse repertoires of antibody genes, generating high affinity binding sites and unique specificity (McCafferty et al., 1990; Winter et al., 1994; Neri et al., 1995; Hoogenboom et al., 1998). Phage antibody genes can be easily sequenced, mutated, and screened to improve antigen binding. The advantage of this technology can also be amplified immunoglobulin variable (V) genes from hybridomas or B lymphocytes using polymerase chain reaction (PCR) technology, cloned into phagemid vectors and rescued the monoclonal antibodies from genetically unstable hybridomas. Finally, soluble recombinant antibodies (not displayed on phage) can be produced rapidly and economically and can be used as *in vitro* diagnostic reagents. Various formats of antigen binding fragments, including Fab and scFv have been cloned and displayed on the surface of M13 viral particles with no apparent loss of the antibody's specificity and affinity as shown in **Figure 1.12** (Nissim et al., 1994; Azzazy and Highsmith, 2002). These antibodies have become important tools in several fields, including molecular biology (Kwong et al., 2008; Kato-Takagaki et al., 2009), pharmaceutical

and medical research (Sidhu, 2000), as well as in the treatment of diseases such as cancer (Schrama et al., 2006) and infectious diseases (Mullen et al., 2006).

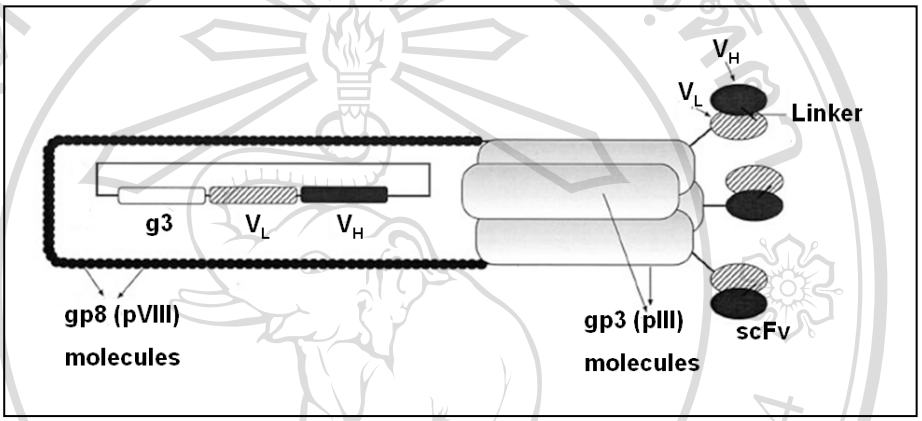


Figure 1.12 Schematic diagram of a filamentous phage displaying single chain variable fragment (scFv) molecules. The phage consists of circular ssDNA surrounded by a coat protein. The genes encoding the variable domains of the scFv and a linker are fused to gene III (g3) in the genome of the filamentous phage. Consequently, the scFv is displayed as a fusion to gp3 (gpIII) protein at the tip of the phage (Azzazy and Highsmith, 2002).

1.2.5 Intracellular antibody (Intrabody)

1.2.5.1 Introduction

Intrabodies are defined as antibody molecules which are expressed intracellularly and directed to defined subcellular compartments. Although mammalian cells are the most commonly used target cells (Williams and Zhu, 2006), expression of intrabodies is not restricted to these cells and a variety of other cell types, including plant cells (Tavladoraki et al., 1993; Tavladoraki et al., 1999), fungal cells (Carlson, 1988; Reinman et al., 2003), and even bacteria (Tavladoraki et al., 1999), have been used. The therapeutic concept of using intrabodies is based on the induction of a phenotypic knockout of a relevant target molecule either by directly inhibiting the function of the antigen or by diverting it from its normal intracellular location. In some cases, intrabodies have also been used to restore the function of a target antigen and thus rescuing a phenotype. Thus, intrabody therapy combines the specificity of antibodies with a gene-therapeutic strategy to selectively affect an intracellular target protein. In contrast to the direct administration of a therapeutic drug, this approach engages the cellular machinery to produce the therapeutic agent. As the intrabodies are produced only inside the cells, this strategy has advantages regarding safety and efficacy (Kontermann, 2004).

1.2.5.2 Antibody structure and format

The basic structure of antibodies (also called immunoglobulins; Igs) consists of two identical heavy and two identical light polypeptide chains. The chains are held together by disulfide bonds leading to a 'Y' shaped protein molecule (**Figure 1.13**). The most common immunoglobulin class is IgG. The amino-terminal protein

domains contain regions of highest sequence variability which mostly contribute to the antigen binding site. These regions, called complementary-determining regions (CDRs) enable the antibody to be specific for a particular target. The rest of the IgG is composed of constant domains that only vary between Ig classes.

Recombinant antibodies are finding an ever increasing number of applications in biotechnology and medicine. A variety of antibody formats have been employed, which reflect differences in the production method, the need for mono- or multivalency as well as the intended use (Worn and Pluckthun, 2001). Antibody formats smaller than the full IgG can be created by genetic engineering as shown in Figure 13. In particular, intrabodies are in the form of single-chain Fv (scFv) antibody fragment contains only the heavy variable domain (V_H) and the light variable domain (V_L) of a full immunoglobulin are fused *via* a peptide linker, generally (Gly4Ser)3, to create a single polypeptide (Williams and Zhu, 2006). This is the smallest antibody fragment which retains the binding specificity of the parental molecule as shown in **Figure 1.14**.

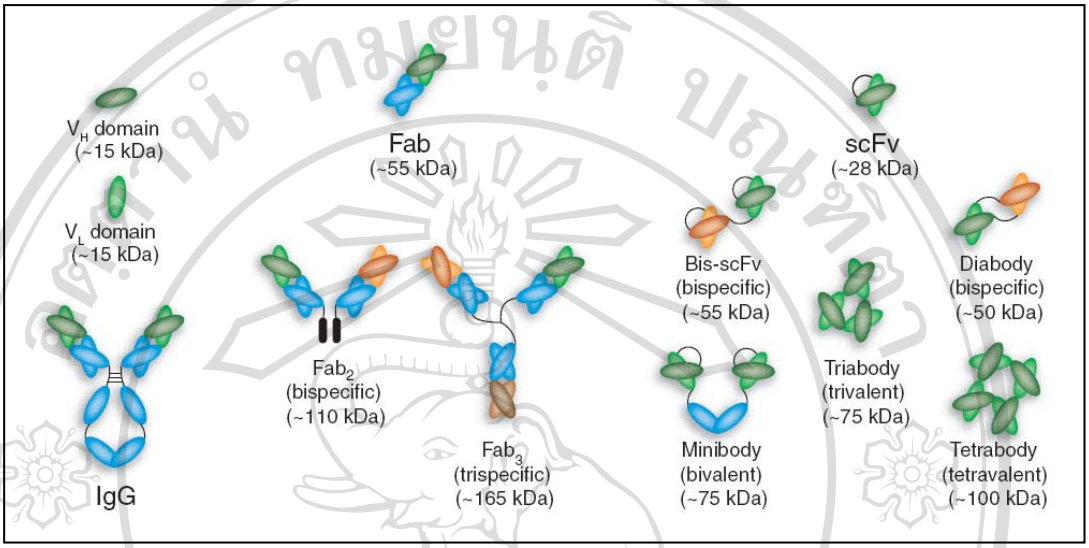


Figure 1.13 Schematic representation of different antibody formats, showing intact immunoglobulin G (IgG). A variety of antibody fragments are depicted, including Fab, scFv, single-domin V_H and multimeric formats, such as minibodies, bis-scFv, diabodies, triabodies, tetrabodies and chemically conjugated Fab' multimers. Sizes are approximately given in kilodaltons (kDa) (Holliger and Hudson, 2005).

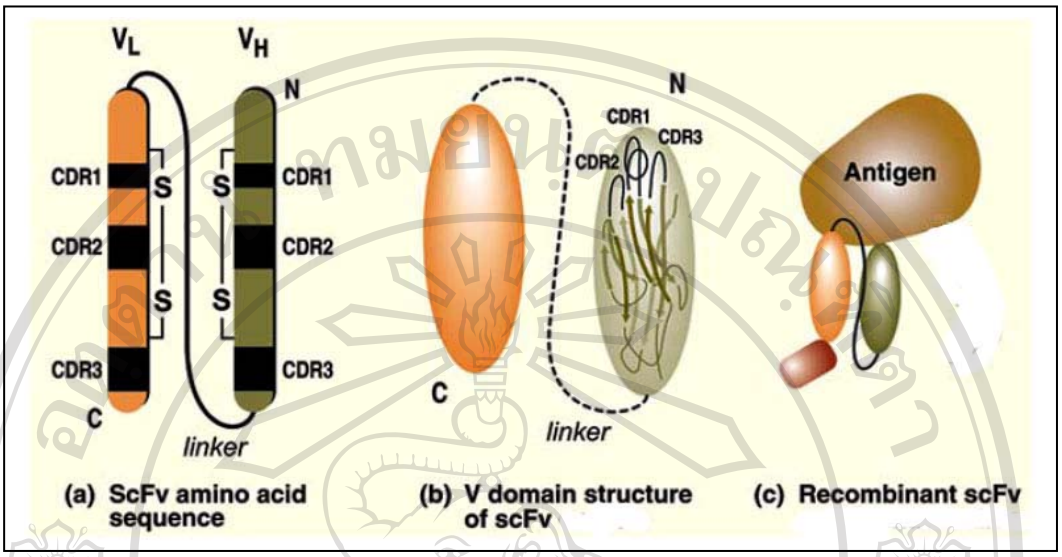


Figure 1.14 scFv fragment format. a) There are three hypervariable regions, the complementary determining regions (CDR), which change in size and sequence among different antibodies, determining the specificity of the antigen-antibody interactions. To facilitate expression of both V_H and V_L domains as single antibody fragments, a short linker peptide (Gly4Ser)3 is added between the carboxy-terminal (C) ends of V_H domain and the amino-terminal (N) of V_L domain. The locations of the cysteines implicated on the formation of the intra-domain disulfide bridge (-S-S-) are also indicated. **b)** The V domain structure of scFv. The structure of a scFv (with linker as dotted line) shows the outline structural representation of a single domain (in this case VH) indicating the three CDRs that determine the antigen specificity occurring in the loop regions**. c**) Schematic recombinant scFv intrabodies binding their antigen. The lighter ends of the V segments represent antigen-binding sites, and the various tags at the carboxy-terminal end, indicate additional protein domains or targeting signals which may add cellular localization signals (nuclear, cytoplasmic, endoplasmic reticulum) (Lobato and Rabbitts, 2004).

1.2.5.3 Mechanisms of action

Intrabodies, in their various forms, function by utilizing the antibody binding site coupled with a molecular tag to direct the antibody-antigen interaction to a specific cellular compartment, or by directly interfering with or neutralizing the function of a target as shown in **Figure 1.15**. Compartmental targeting is achieved through the use of N- or C-terminal tags that encode specific signal sequences used (Williams and Zhu, 2006). These signals allow the intrabody to enter cellular compartments it would not normally enter. As a result, the mechanisms through which intrabodies can achieve their therapeutic effect are expanded.

The most commonly used intrabodies can be targeted to the ER through the use of a leader sequence and an ER retention signal, such as the KDEL peptide, ensuring that they are retained within the ER–Golgi complex (Munro and Pelham, 1987). The purpose of these antibodies is usually to prevent secretion of target proteins, or the maturation of target proteins that would normally be expressed on the cell surface. Under most circumstances, protein secretion occurs *via* a forward pathway. Proteins are synthesized in the ER, transported to the cis-Golgi network (CGN) apparatus and secreted from the secretory vesicles that bud from the trans-Golgi network (TGN). However, there is also a retrograde pathway from the Golgi to the ER. Proteins that have the ER trapping signals can bind to ERD2.1 or ERD2.2 (ERD) and return to the ER. ERD exists in a nonbinding state in the forward pathway, but undergoes a conformational change to allow ligation of the KDEL sequence in the retrograde pathway. Retained proteins are degraded either within the ER by the non-lysosomal degradative pathway or by the cytoplasmic proteasome. The intrabody can

be targeted to the nucleus with a nuclear localization sequence (NLS) or to the mitochondria with appropriate signal sequences or to the cytosol by deleting the signal leader sequence. The intrabody may neutralize protein function, such as that of enzymes, or function primarily by trapping, rerouting, or enhancing the degradation of the target protein (Wheeler et al., 2003a; Boldicke, 2007).

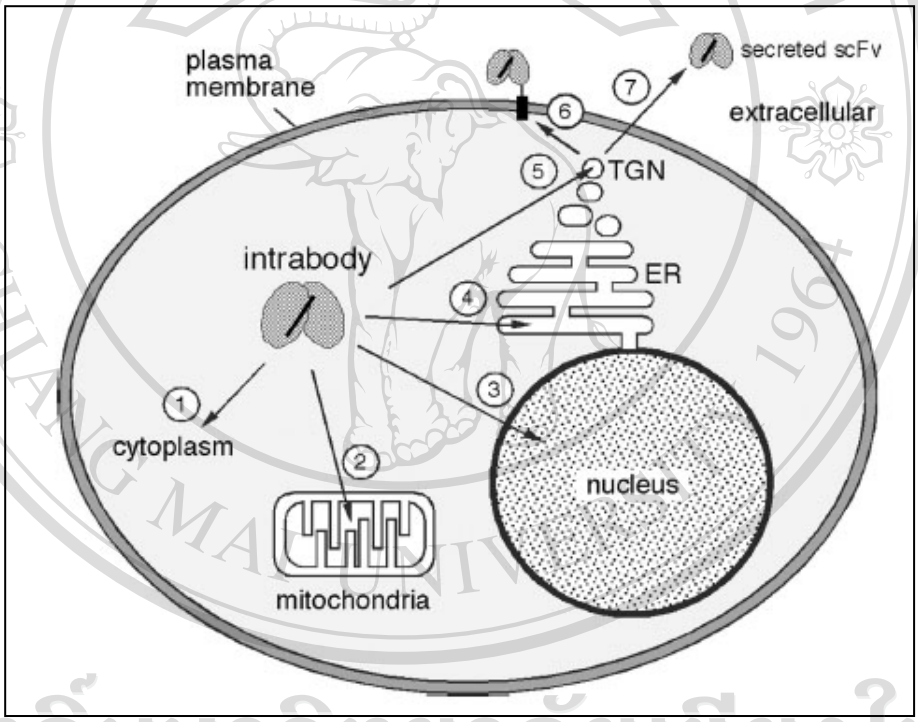


Figure 1.15 Subcellular location of intrabodies. Intrabodies have been expressed intracellularly and directed to the cytoplasm (1), mitochondria (2), the nucleus (3), the endoplasmic reticulum (4), the *trans*-Golgi network (TGN) (5), the plasma membrane (6), or secreted into the extracellular space (7) (Kontermann, 2004).

1.2.5.4 Intrabodies and their application

The use of an intrabody against an intracellular target was first described in 1988. Neutralization of yeast alcohol dehydrogenase I (ADH I) enzyme activity *in vivo* was observed using heavy and light-chain cDNAs expressed in *Saccharomyces cerevisiae* (Carlson, 1988). Recent studies have demonstrated that an intrabody strategy can successfully produce both phenotypic and functional knockouts of target molecules (Alvarez et al., 2000; Steinberger et al., 2000a; Wheeler et al., 2003b; Liu et al., 2005; Peng et al., 2007). This approach depends on a receptor-mediated system for retention of certain proteins within the endoplasmic reticulum. Intrabodies could be a useful tool not only for clinical applications such as neurologic disorders (Messer and McLear, 2006; Emadi et al., 2007; Lynch et al., 2008) and cancer therapy (Popkov et al., 2005; Griffin et al., 2006; Tanaka et al., 2007) but also for functional analysis of proteins inside the cells (Wheeler et al., 2003b; Boldicke et al., 2005; Doebis et al., 2006; Goenaga et al., 2007).

Intrabodies against several different HIV-1 proteins have been tested in cultured and primary human cell lines. These intrabodies have targeted the HIV-1 envelope protein gp120 by localizing the intrabody to the endoplasmic reticulum, the transcriptional regulatory proteins tat and rev by localizing the intrabodies to the cytoplasm or the nucleus, and the structural protein gp17. Notably, scFvs have been shown to reduce HIV viral production and the infectivity of released particles (Marasco et al., 1993; Mhashilkar et al., 1995; Marasco et al., 1998). Similarly, intrabodies have also been used to down-regulate the expression of CCR5 which is known as a co-receptor for HIV infection into macrophages. This CCR5-intrabody

efficiently blocked surface expression of human and rhesus CCR5 and thus prevented cellular interactions with CCR5-dependent HIV-1 and simian immunodeficiency virus envelope glycoprotein. Intrabody-expressing cells were shown to be highly refractory to challenge with R5 HIV-1 viruses or infected cells. These results suggest that deletion of the functional receptor or reduced expression of CCR5 should be beneficial in the treatment of HIV-1 disease (Steinberger et al., 2000a; Steinberger et al., 2000b).

Intrabodies are being developed to bind to, neutralize, or modify the function or localization of cancer-related targets and thereby affect the malignant phenotype. This has resulted in a promising new tool for the study and treatment of cancer. Their small size facilitates expression and assembly of functional molecules (Hudson, 1998). Growth factor receptors like erbB-2, transferrin receptor (TfR), vascular endothelial growth factor receptor 2 (VEGFR2) or epidermal growth factor receptor (EGFR) are the molecular targets thus far studied with the use of scFv (Lobato and Rabbitts, 2004). An anti-erbB-2 scFv construct containing an ER-directed leader sequence was transiently expressed in the human ovarian carcinoma cell line SKOV3 using the adenovirus-polylysine vector. This strategy can knock-out erbB-2 expression and induce a significant anti-neoplastic effect in ovarian cancer cells overexpressing this growth factor receptor. In addition, the ability to accomplish selective abrogation of erbB-2 expression in animal models and to transfect and eradicate primary ovarian cancer cells justifies further investigation of this novel strategy in ovarian cancer patients (Deshane et al., 1995a; Deshane et al., 1995b; Alvarez et al., 2000). Intrabodies have been also generated to a number of growth

factor receptors, such as VEGFR2 (Afanasieva et al., 2003; Boldicke et al., 2005; Jendreyko et al., 2005) and EGFR (Beerli et al., 1994; Jannot et al., 1996). They have been shown to markedly decrease the cell surface expression of the targeted receptor, leading to cell growth inhibition both *in vitro* and *in vivo* (Beerli et al., 1994; Jannot et al., 1996; Afanasieva et al., 2003; Boldicke et al., 2005). Likewise, intrabodies have been used to inhibit the proliferation of tumor cells. Peng et al. have been developed an anti-TfR scFv-intrabody (scFv-HAK) as a growth inhibitor of TfR overexpressing tumors. This intrabody was able to block surface expression of TfR in tumor cells MCF-7. Furthermore, expression of scFv-HAK can dramatically induce cell cycle G1 phase arrest and apoptosis of tumor cells, and consequently significantly suppress proliferation of tumor cells MCF-7 (Peng et al., 2007).

The application of intrabodies has more advantages than RNA interference (RNAi) since they possess a much longer active half-life compared to RNAi, and are also much more specific and affinity to their target molecules as shown in **Table 1.3** (Cao and Heng, 2005; Heng et al., 2005; Williams and Zhu, 2006). Furthermore it is possible to analyze the function of specific domains of a protein in its native environment without destroying the structure of the target protein. Binding of the antibody to an intracellular molecule directly has a potential to block, suppress, alter or even enhance the process mediated by that molecule. In particular, intracellular use of antibody fragments can offer an effective alternative to gene-based knockout technologies (Stocks, 2004). Therefore expression of intrabodies becomes a broadly applicable technology for probing the biological function of interested

protein. Therefore, this technique will be applied, in this study, for functional study of CD147 surface molecule.

Table 1.3 The properties of intrabodies and RNAi (Boldicke, 2007).

Intrabodies	**RNAi**
Prerequisite is a specific antibody	Prerequisite is the sequence of the mRNA or promoter of the target
Time consuming technology	Much less technical challenge
Very high specificity to the target	Non-specific effects
Long active half-life	Relatively short active half-life
Targeting of specific protein domains	Loss of multiple functions of the target
Inhibition of post-translational modifications	Not possible

1.2.5.5 Intrabody and gene delivery

A major goal is the development of vectors to selectively deliver the genetic material encoding the intrabody to a specific target cell type. There are two methods to transfer the intrabody genes into living cells *e.g.* non-viral or viral transfer systems. Non-viral vectors are often composed of cationic peptides, polymers, and lipids that interact with the negatively charged backbone of DNA to form nanometer sized particles that spontaneously transfect cells in culture (Bloomfield, 1996). At present, the main disadvantage is the low transduction efficiency and transient transgene expression, no specific cell targeting and difficult *in vivo* applications (Boldicke, 2007). Since, the intracellular dissociation and subsequent metabolism of a cationic peptide polymer leads to the release and premature metabolism of plasmid DNA by DNase leading to a low level and short duration of expression (Liu and Knapp, 2001). The enzymes responsible for the metabolism of peptide-mediated non-viral delivery systems are believed to be in the lysosomes (Wiethoff and Middaugh, 2003). However, recent evidence has implicated the involvement of the proteasome (Kim et al., 2005).

Traditionally, gene therapy approaches utilize viral infection to deliver the genetic material to the target cells. An advantage of this method can yield continuous expression of an intrabody to knock-out protein functions inside a target cell. Recombinant adenoviruses are attractive vectors for gene delivery. This DNA virus can infect a wide variety of cell types including dividing and quiescent cells and can be easily produced at high titer. The expression of the transgene is transient and the viral genome does not normally integrate into the host genome with no risk for

insertional mutagenesis (Robbins et al., 1998; Boldicke, 2007). Recombinant adenovirus have been successfully employed for *in vitro* transduction of ER intrabodies into cell lines (Wright et al., 1997; Mhashilkar et al., 2002) and primary cells (Wheeler et al., 2003b; Beyer et al., 2004). Conversely, retroviral vectors are not able to transduce non-dividing cells. Further disadvantages are low vector titer and low transfection efficiency and particle instability. Retroviral vectors are suitable for *ex vivo* gene therapy, and despite the disadvantages, retroviral gene delivery systems have been used already in a number of clinical trials (Rainov and Ren, 2003).

1.2.6 Adenovirus (Ad)

1.2.6.1 Biology of adenovirus

Adenovirus is an infectious agent that had been derived from human adenoids in 1953 by Rowe. Adenoviruses are widespread in many species and can be isolated from both sick and healthy individuals. Adenoviruses infections are mostly asymptomatic but may be associated with diseases of the respiratory, ocular and gastrointestinal systems. In 1962, Trentin reported that human adenovirus type 12 can cause tumors in newborn hamsters. However, later studies could not demonstrate any association with malignant disease in humans (Berk, 2007).

Adenoviruses are classified in two genera based on the basis of the presence of a genus-specific antigen: the mastadenoviruses that infect mammals and the aviadenoviruses with infecting birds. There are 51 immunologically distinct human adenovirus serotypes, further classified into 6 subgroups (A-F) (**Table 1.4**)

that can cause human infections ranging from respiratory disease, conjunctivitis, pharyngoconjunctival fever and gastroenteritis, but do not show oncogenic potential in humans. Adenoviruses are primarily spread *via* respiratory droplets, however they can also be spread by fecal routes (Russell, 2000; Majhen and Ambriovic-Ristov, 2006).

Table 1.4 Subgrouping of human adenoviruses: association with diseases

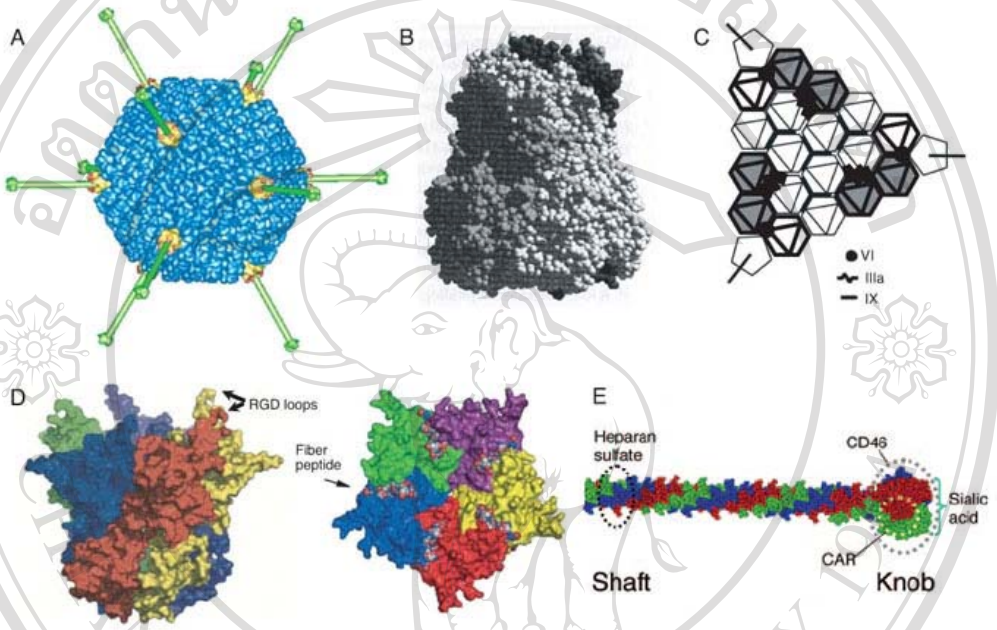
Subgroup	Serotypes	Syndrome
A	12, 18, 31	Gastroenteritis
B	3, 7, 11, 14, 16, 21, 34, 35, 50	Upper respiratory illness, conjunctivitis, cystitis
C	1, 2, 5, 6	Pharyngoconjunctival fever
D	8, 9, 10, 13, 15, 17, 19, 20, 22-30, 32, 33, 36-39, 42-49, 51	Epidemic keratoconjunctivitis, Immunocompromised host disease
E	4	Respiratory disease
F	40, 41	Gastroenteritis

The serotypic origin of the E1A gene determines the oncogenic phenotype of adenovirus-transformed cells. Viruses belonging to subgroup A (such as adenovirus 12, Ad12) induce tumors with high frequency and short latency, while viruses from subgroup B (such as Ad3 and Ad7) are weakly oncogenic. In harmless Adenoviruses from subgroup C (which includes the well studied serotypes Ad2 and Ad5), D, E and F are non-oncogenic (Russell, 2000).

1.2.6.2 Morphology and structure of adenovirus

The human adenoviruses are non-enveloped icosahedral particles approximately 90 nm in diameter, with fibers projecting from the vertrices of the icosahedron. Virions consist of a protein shell (capsid) surrounding a DNA-containing core. There are three major capsid proteins that make up the viral particle. Hexon is the most abundant structural component and constitutes the bulk of the mature virion. Five subunits of the penton base are found at each of the twelve vertices of the capsid and form the platform for the twelve fiber homo-trimers that protrude from the virion. At the distal tip of each linear fiber is a globular knob domain (**Figure 1.16A**) (Glasgow et al., 2004). The fiber is composed of three domains, an N-terminal domain that binds to the penton base (**Figure 1.16D**), central shaft with slight flexibility important for infection and a globular C-terminal knob that binds the primary receptor on host cells. The length of shaft varies among serotypes from six repeating units in Ad3 to 22 in Ad2 and Ad5. Adenoviral particles have no membranes or lipids and are therefore stable in chemical or physical agents and

adverse pH condition, allowing for prolonged survival outside of the body and water (Berk, 2007).



.**Figure 1.16 Adenovirus structure. A)** Model of the Ad2 virion from computer reconstruction of cryo-electron microscopic images. Each of the 20 triangular faces of the capsid is composed of 12 copies of the hexon trimer (blue). At each fivefold vertex, a fiber (green) emerges from the pentameric penton base (yellow). **B)** Space-filling model of the Ad5 hexon trimer. Each subunit is shown in a different shade. **C)** Schematic of the current model for the locations of polypeptides in one facet of the icosahedron. Hexons belonging to adjacent facets are shaded gray. Pentagon and bars represent penton bases and fibers. The minor coat proteins are depicted as shown in the key. **D)** Penton base viewed from the side (left) and top (right) and showing flexible RGD loops projecting from the surface. The insertion sites for fiber are also shown. **E)** Fiber structure and receptor binding sites. The fiber is a trimer whose monomers are indicated in red, blue, and green; the shaft is a tightly wound triple spiral; the knob is a more bulbous trefoil (Zhang and Bergelson, 2005).

1.2.6.3 Adenovirus genome

Adenoviruses are nonenveloped viruses possessing a double-stranded DNA genome. A virus-coded terminal protein is covalently linked to the 5'end of each DNA strand. The genome of the most commonly used human adenovirus (serotype 5) has 36-kb, and contains inverted terminal repetitions (ITRs) of 103 bp. In general, Ad composed of multiple copies of 11 different structural proteins, 7 of which form the icosahedral capsid (II, III, IIIa, IV, VI, VIII, IX) and 4 of which are packaged with the linear double-stranded DNA in the core of the particle (V, VII, μ and terminal protein) as shown in **Table 1.5**.

Table 1.5 Type of proteins in adenovirus capsid.

Name	Location	Known Functions
II	Hexon monomer	Structural
III	Penton base	Penetration
IIIa	Associated with penton base	Penetration
IV	Fiber	Receptor binding haemagglutination
V	Core associated with DNA and penton base	Histone-like; packaging?
VI	Hexon minor polypeptide	Stabilization/assembly of particle?
VII	Core	Histone-like
VIII	Hexon minor polypeptide	Stabilization/assembly of particle?
IX	Hexon minor polypeptide	Stabilization/assembly of particle?
TP	Terminal protein	Genome replication

As shown in **Figure 1.17**, Hexon bases are packed together to form a protein shell with 12 pentons at the apices of the icosahedral capsid. The positions of hexons (II), penton bases (III), fibers (IV) and protein (IX) are well established; 12 copies of polypeptides IX are found between 9 hexons in the center of each facet. The positions of proteins IIIa, VI and VIII are tentatively assigned. Two monomers of IIa appear to penetrate the hexon capsid at the edge of each facet. Multiple copies of VI form a ring underneath the peripentonal hexons. The 12 penton bases are formed by the interaction of 5 polypeptides (III) and are less tightly associated with the neighboring (peripentonal) hexons. Each of the vertex pentons carries 1 or 2 fibers, each consisting of 3 polypeptides (IV) that interact to form a shaft of characteristic length and a distal knob. Polypeptide VIII has been assigned to the inner surface of the capsid. Other polypeptides (monomers of IIIa, trimers of IX and multimers of VI) seem to interact with hexons to stabilize the capsid. The core consists of the DNA genome complexed with 4 polypeptides (V, VII, μ, TP).

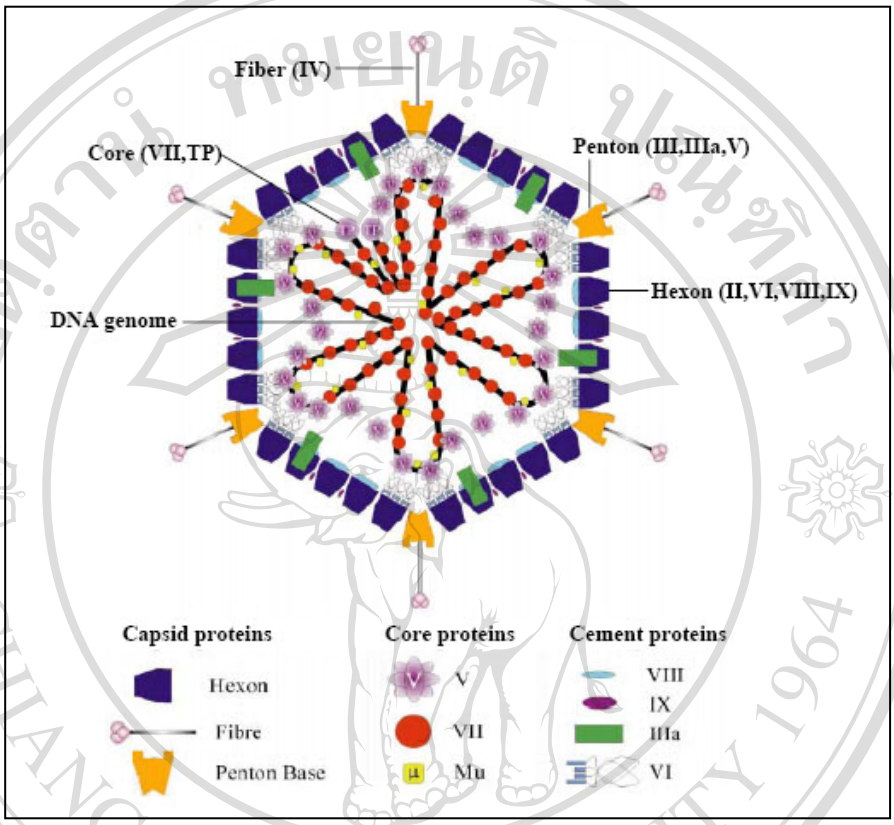


Figure 1.17 Schematic of adenovirus virion summarizing the current model for locations of polypeptides in the capsid and core of the particle.

1.2.6.4 Adenovirus entry

Adenovirus entry into cells, as defined by experiments with cultured cells, generally involves attachment to a primary receptor, followed by interaction with a secondary receptor responsible for internalization. The route of intracellular trafficking is influenced by the fiber knob and, thus, by interaction with a specific primary receptor. Many experiments demonstrate that the 46-kDa coxsackievirus B and adenovirus receptor (CAR), a member of the immunoglobulin superfamily, is the receptor for most Ads from subgroups A, D, E, and F, but not for subgroup B or the short fiber of subgroup F (Bergelson et al., 1997). Both human and mouse CARs have been identified. CAR is expressed on the surface of many cell types and is present within the tight junctions between polarized epithelial cells (Cohen et al., 2001). In addition, other receptors have been described such as α2 domain of the class I major histocompatibility complex (MHC-I) (Hong et al., 1997), heparan sulfate glycosaminoglycans (HS-GAGs) (Dechecchi et al., 2001), vascular cell adhesion molecule 1 (VCAM-1) (Li et al., 1993). In some cases, virus attachment may depend on direct interaction between the penton base and a cell surface integrin, without the need for a primary fiber receptor. Fiber-deficient Ad2 virions can infect CAR-negative monocytic cells by a mechanism that involves a primary attachment to integrins αMβ2 and αLβ2, followed by an interaction with αv integrins that is needed for internalization (Huang et al., 1996). After knob-CAR binding, receptor mediated endocytosis of the virion is affected by interaction of penton base Arg-Gly-Asp (RGD) motifs with cellular integrins $\alpha_V\beta_3$, $\alpha_V\beta_5$ or other integrins as shown in **Figure**

1.18 (Wickham et al., 1993; Wickham, 2000; Li et al., 2001). Once endocytosed, acidification of the endosome triggers a conformational change in the viral capsid, the virion is then released into the cytoplasm and translocated to the nucleus. The viral genome then enters the nucleus and from its episomal location undergoes transcription and then replication. Viral gene products are then produced in the cytoplasm following translation and capsid proteins localize to the nucleus where virus assembly occurs. Virus can then be released from the cell following lysis.

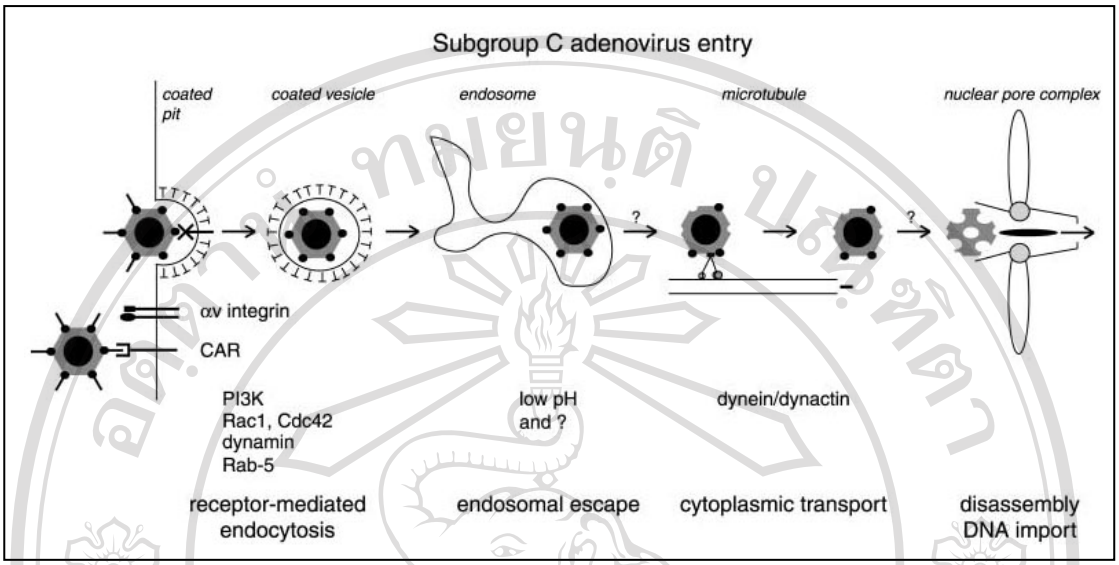


Figure 1.18 Illustration of subgroup C adenovirus entry. Ad-2 or Ad-5 bind to the coxsackievirus-adenovirus receptor (CAR) and the secondary receptor αv integrin and enter cells by receptor-mediated endocytosis *via* coated pits and coated vesicles. Early entry steps require PI3K, Rac1 and Cdc42, and also dynamin. Virus then escapes from endosomes by an unknown mechanism, somehow assisted by low pH. Cytosolic virus particles are transported by the dynein/dynactin motor complex along microtubules in the minus-end direction towards the nucleus. Nuclear pore complex-docked virus particles are dismantled, and the DNA genome is released into the nucleus (Greber, 2002).

1.2.6.5 Transcription and replication of adenovirus

Adenovirus transcription can be defined largely as a two-phase event, early and late, respectively occurring before and after virus DNA replication. Transcription is accompanied by a complex series of splicing events, with four early cassettes of gene transcription termed E1, E2, E3 and E4. The different transcripton units for the wild-type Ad5 are indicated by arrows (**Figure 1.19**). Inverted-terminal-repeat (ITR) sequences function as replication origins, and the ψ sequence is required for packaging of the viral genome. The viral genome is transcribed from both DNA strands by host cell enzymes (Russell, 2000).

The E1 gene products can be subdivided further into E1A and E1B. E1A itself has two major components sharing substantial stretches of sequence that are termed 289R (or 13S) and 243R (or 12S), based on the number of amino acid residues. These E1A proteins are primarily concerned with modulating cellular metabolism to make the cell more susceptible to virus replication. E1A proteins interfere with the processes of cell division and with the regulation of NF-κB and p53, and do this by a great variety of strategies involving both direct and indirect interaction with cellular proteins. They can also modulate transcription patterns in favour of virus transcription. Moreover, other virus gene products can modulate these cellular interactions significantly. For instance, the E4 gene products can co-operate with E1A to promote cell cycle-independent adenovirus growth (Goodrum and Ornelles, 1999). The E1B gene product 19K also seems to function cooperatively with E1A and p53 in promoting oncogenesis and transformation (Kannabiran et al., 1999),

mainly by ensuring that the downstream consequences of cell cycle release do not induce apoptosis.

The E1B 19K gene product is analogous to that from the cellular Bcl-2 gene. This gene product is concerned with prolonging cell survival by interacting and ablating members of the Bax family (whose transcription can be promoted by p53), which induce apoptosis and necrosis (Han et al., 1996).

The E2 gene products are subdivided into E2A (DBP) and E2B (pTP and Pol). These provide the machinery for replication of virus DNA (Hay et al., 1995) and the ensuing transcription of late genes, and this is mediated by interaction with a number of cellular factors. The E3 genes, which are dispensable for the replication of virus in tissue culture, provide the essential of proteins that subverts the host defense mechanisms. One of these E3 gene products has been termed the adenovirus death protein (ADP), since it facilitates late cytolysis of the infected cell and thereby releases progeny virus more efficiently (Tollefson et al., 1996). The gene products derived from the E4 cassette (termed orfs 1-6/7) mainly facilitate virus messenger RNA metabolism (sometimes in association with E1B gene products) (Goodrum and Ornelles, 1999) and provide functions to promote virus DNA replication and shut-off of host protein synthesis (Halbert et al., 1985). They are also associated with resistance to lysis by CTLs (Kaplan et al., 1999).

Adenoviruses also transcribe a set of RNAs that are not translated, termed the VA RNAs, and these play a role in combating cellular defense mechanisms. DNA replication begins from both DNA termini and requires sequences within the ITRs as origins of replication (Hay et al., 1995). Thereafter, late

transcription ensues, with five cassettes of transcripts (termed L1 to L5) resulting from a complex series of splicing events. These lead to the production of the virus structural components and the encapsidation and maturation of virus particles in the nucleus. A key player in the control of transcription is the major late promoter (MLP), which is attenuated during transcription of the early genes. However, it should be noted that there is a low basal level of late transcription occurring early in infection, even before the MLP comes into play. After the onset of virus DNA replication, the IVa2 and IX genes are expressed at high levels and transcription *via* the MLP is fully functional by specific activation. This is accomplished *via* the IX and IVa2 gene products (Lutz et al., 1997) and is also influenced by effective competition for the limiting transcription factors (Fessler and Young, 1998). The encapsidation process is governed by the presence in the virus DNA of a packaging signal at the conventional left end, which consists of a series of AT-rich sequences (Hearing et al., 1987). These events are accompanied by major changes in the nuclear infrastructure and the permeabilization of the nuclear membrane (Rao et al., 1996). At the very late stages of infection, when the cellular nucleus is packed full of virions, adenovirus synthesizes an ADP, which promotes cell lysis and thereby allows adenovirus to be released from cells and infect other cells.

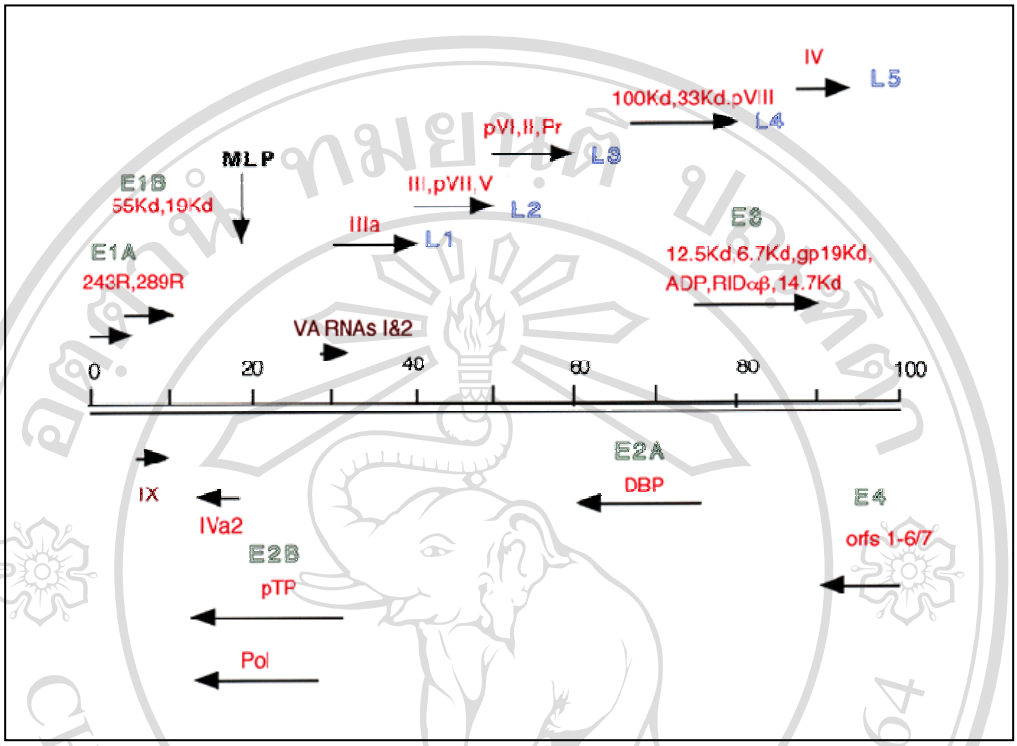


Figure 1.19 Transcription of the adenovirus genome. The early transcripts are outlined in green, the late in blue. Arrows indicate the direction of transcription. The gene locations of the VA RNAs are denoted in brown. MLP, Major late promoter (Russell, 2000).

1.2.6.6 Adenoviral vector for gene delivery

Gene delivery consists in introducing DNA and RNA into cells, tissues, or organisms, in order to study regulation and function of genes and proteins. The biggest hurdle that gene delivery technologies have to overcome is the cell membrane, which is impermeable to negatively charged macromolecules such as DNA and RNA. Several gene delivery systems are available. Retroviral vectors as well as nonviral gene transfer methods such as calcium phosphate coprecipitation, electroporation and liposomal transfection, target only a small fraction of cell population after extended selection periods, many of which, however, are not appropriate for transduction of human lymphocytes. Furthermore, expression levels are often low and insertion site-dependent silencing of the transgene expression is a frequent predicament (Doerfler et al., 1997; Baum et al., 2003; Hacein-Bey-Abina et al., 2003). In addition, the genome of retroviruses and lentiviruses is small, limiting the size of exogenous genes that can be packaged and transferred to target cells by the derivative vectors. Conversely, adenoviral vectors are an attractive alternative since they can efficiently transduce both dividing and non-dividing cells, and achieve transgene expression within hours (Nevins et al., 1997). Moreover, adenoviruses do not integrate into the host genome, thus leaving the genetic package of targeted cells unmodified. This results in reproducible gene expression levels and also eliminates any undesirable effects related to the site of integration, allowing the specific analysis of the transgene effects (Kay et al., 2001). This characteristic, together with their relative ease of preparation and purification, has led to their extensive use as gene transfer vectors.

The adenovirus group C (Ad2 and Ad5) is most commonly used as gene transfer vectors by deleting the E1 and/or E3 gene cassettes, allowing the introduction of up to 6.5 kb of foreign DNA, often under the control of a heterologous promoter. In the case of the E1 deletions, care was taken to ensure the retention of the ITR and the packaging sequences (ψ). The E1 deletion renders the viruses defective for replication and incapable of producing infectious viral particles in target cells. The El-, replication-defective virus can be propagated in a cell line that provides the El polypeptides *in trans*, such as the human embryonic kidney cell line, 293. A gene of interest (GOI) can be inserted by recombination in place of the El gene; expression is driven from either the El promoter or a heterologous promoter. In the case of the E3-deleted vectors, there were similar sequelae as a result of the elimination of the E3 gene-mediated defenses against host responses (Poller et al., 1996).

1.2.6.7 Adenoviral vector and their application

Recombinant adenovirus technology is largely used in gene therapy which aims at treating both genetic diseases such as cancer (Korn et al., 2004; Mizuguchi and Hayakawa, 2004; Tuve et al., 2007), cystic fibrosis (Driskell and Engelhardt, 2003; Griesenbach et al., 2006) and infectious diseases e.g. AIDS (Wu and Nemerow, 2004; Barouch and Nabel, 2005; Ura et al., 2008) by introducing new genetic material into selected cells. This technology is used to overexpress proteins of interest and to subsequently study their functions. In contrast to prokaryotic or insect cell-based systems, the use of human cells permits the complex post-translational

protein modifications required to ensure the proper folding and post-translational modifications of the protein.

Adenovirus type 2 and 5 (Ad2 and Ad5) are most commonly used as gene transfer vectors. The primary receptor responsible for attachment of all Ad serotypes except those from group B is the CAR and following knob-CAR binding, receptor mediated endocytosis of the virion is affected by interaction of penton base Arg-Gly-Asp (RGD) motifs with cellular integrins $\alpha_V\beta_3$, $\alpha_V\beta_5$ and/or other integrins. Unfortunately, Ad5 has been ineffective at infecting hematopoietic progenitor cells (HPC). Human hematopoietic cells including lymphocytes express low levels of CAR and $\alpha_V\beta_3$ integrins, which results in poor adenoviral transduction efficiencies (Neering et al., 1996; Rebel et al., 2000). In order to overcome the refractoriness of CAR-negative cells to conventional Ad vectors, several groups have developed several types of fiber-mutant Ad vectors (Mizuguchi et al., 2001; Mizuguchi and Hayakawa, 2002) and Ad vectors derived from other serotypes (Sakurai et al., 2003) which are able to transduce cells *via* CAR-independent pathways, leading to high transduction efficiencies even in CAR-negative cells.

Recently, it has been demonstrated that CD46 also acts as a receptor for the majority of subgroup B adenoviruses (Ads), including Ad serotypes 11 (Ad11) and 35 (Ad35) (Gaggar et al., 2003; Segerman et al., 2003). The fiber knob domain of Ad11 or Ad35 binds to short consensus repeats (SCRs) 1 and/or 2 in CD46 for infection (Fleischli et al., 2005; Gaggar et al., 2005; Sakurai et al., 2006). CD46 is a membrane glycoprotein that protects cells from complement damage. CD46 is also a

receptor for measles virus laboratory strains, for human herpes virus 6, and for certain pathogenic bacteria (Cattaneo, 2004). In humans, CD46 is expressed on all nucleated cells at a low level. Adenovirus serotype 5 (Ad5)-based vectors can be retargeted with fiber receptor specificity of serotype 35 adenovirus (Ad5/F35) and thereby bypass the insufficiency of the CAR on hematopoietic cells by utilizing CD46 as cellular receptor (Gaggar et al., 2003; Nilsson et al., 2004). This chimeric Ad5/F35 vector efficiently transduced human hematopoietic progenitor cells (Shayakhmetov et al., 2000), dendritic cells (Rea et al., 2001), primary human T lymphocytes and NK (Schroers et al., 2004), normal human B lymphocytes (Jung et al., 2005), primary chronic myeloid leukemia cells, chronic lymphocytic leukemia B cells (Nilsson et al., 2004), and B cell acute lymphoblastic leukemia cells (Yotnda et al., 2001).

1.3 Objectives

The ultimate goal of this study is to down-regulate the cell surface CD147 expression using intrabody technology. The studies provide insight into a better understanding of CD147 molecule. The specific aims of this study were as follows:

1. To generate the soluble form of scFv-M6-1B9 against CD147
2. To generate recombinant adenoviruses expressing scFv-M61B9
3. To assess the biological activity of scFv-M6-1B9 against CD147 as soluble form and intrabody