# CHAPTER 1

# INTRODUCTION

## 1.1 Statement and significance of the problem

CD147 is a member of human leukocyte surface molecule which found on the surface of various cells. It was designated as cluster of differentiation (CD) system in 1997 at the Sixth International Workshop and Conference of Human Leukocyte Differentiation Antigens (HLDA workshop) (Stockinger *et al.*, 1997). The aliases of CD147 are basigin (Miyauchi *et al.*, 1991), M6 antigen (Kasinrerk *et al.*, 1992), extracellular matrix metalloproteinase inducer (EMMPRIN) (Biswas *et al.*, 1995). It is a glycoprotein of type 1 transmembrane protein of the Ig superfamily composing of 269 amino acids. The extracellular part of CD147 contains 2 domains, which are similar to the domain of human Ig.

CD147 expresses on surface of various cell types, especially on the surface of cancer cells and T cell after it was activated. Certain studies demonstrated the involvement of CD147 molecule in different cellular activations i.e. upregulated upon T cell activation (Kasinrek *et al.*, 1992), inhibited homotypic aggregation of the estrogen-dependent breast cancer cell line MCF-7 with CD147 mAbs (Schiavone *et al.*, 1997), homotypic cell aggregation of the U937 cell line with anti-CD147 mAbs (Kasinrerk *et al.*, 1999). However, the information of mechanisms which take part in signal transduction via CD147 is not clearly proposed since the surface ligand or receptor of CD147 has not yet been

identified. Thus, the suitable method to produce the precise molecule of CD147 in enough quantity is needed for discovering the cellular functions of CD147.

In addition, for studying of CD147 signaling several monoclonal antibodies had been previously used and shown different cellular activations (Kasinrerk *et al.*, 1992; Koch *et al.*, 1999 and Khunkeawla *et al.*, 2001). Hence, the map of epitopes recognized by those mAbs will provide a useful information about the bioactive domains of molecule. Several immunological techniques may be used for epitope characterization. Two methods have been reported for CD147. The epitope map of a soluble CD147-Fc fusion protein produced from transfected COS cells has been evaluated by BIAcore biosensor, which in principle is accurate but extremely expensive (Koch *et al.*, 1999). More recently, the epitope mapping of CD147 mAbs was analyzed by a fluorescence inhibition technique (Pengin, 2003). In addition, the competitive mAbs must be labeled with fluorescein dye, which is labour-intensive.

One possibility to produce a close to nature molecule of CD147 is by phage display technique. Basically, phage display is based on molecular biology and protein expression in bacteria. Thus it can reduce time consuming and needs no sophisticate techniques in comparison to mammalian expression system. Many reports previously described the success in displaying of proteins or polypeptides on the phage particles such as: phage display library of allergic protein (Appenzeller *et al.*, 2001) phage display CD99 (Tayapiwatana and Kasinrerk, 2002) and phage display scFv against gpIIa integrin (An *et al.*, 2002).

Regarding the advantages of phage display technique, we attempt to produce the extracellular domain of CD147 molecule. The recombinant phages

will be subjected to determine and characterize the CD147 molecule linking to pIII on phage particle using certain immunological methods.

Since, the molecular function of CD147 still unclear, thus, production of CD147 molecule is a key to trace the ligand-receptor of its. Phage display technique is applied to use to produce this molecule. By using *E. coli* as a host, the *E. coli* strain XL-1 Blue and TG-1 will be used. Moreover, the influence *E. coli* strain for folding of protein will be compared from both *E. coli* strains. The suitable folding molecule will be used to characterize the epitope using certain immunological methods.

## 1.2 Literature review

### 1.2.1 The human CD147

#### 1.2.1.1 Structure of CD147

Human CD147 molecule alias basigin, M6 antigen and extracellular matrix metalloproteinase inducer (EMMPRIN) is a molecule with a molecular weight of 50-60 kDa. It is an N-glycosylated protein consisting of 269 amino acids (Miyauchi *et al.*, 1991; Kasinrerk *et al.*, 1992; Biswas *et al.*, 1995). Human CD147 is the species homologue of mouse protein termed gp42 (Altruda *et al.*, 1989) and basigin (Miyauchi *et al.*, 1990), rat OX-47 (Fossum *et al.*, 1991) or CE9 (Nehme *et al.*, 1995) antigen, rabbit homologue (Schuster *et al.*, 1996) and the chicken molecule termed HT7 (Seulberger *et al.*, 1990), 5A11 antigen (Fadool *et al.*, 1993) or neurothelin (Schlosshauer *et al.*, 1990). CD147 has typical features of type 1 integral membrane protein (Figure 1.1.). The putative extracellular

domain consists of two Ig-like domains most probably of the domain V and C2 types as determined by comparison with other Ig domains. Endoglycosidase F treatment of immunoprecipitates resulted in a mobility shift from 54 kDa to 28 kDa demonstrating that the majority of the oligosaccharide chains are N-linked (Kasinrerk *et al.*, 1992). The 21 amino acids of the putative transmembrane region are 100% identical in the human, rat and chicken homologues and, with the exception of one amino acid, also in the mouse and rabbit forms (Kasinrerk *et al.*, 1992; Schuster *et al.*, 1996). The hydrophobic stretch of the transmembrane region is interrupted by a charged residue, a glutamic acid, and contains a leucine-zipper; both are potential protein-protein interaction motifs. This and the strong conservation suggest an important functional role for this region perhaps in interactions with other proteins within the plasma membrane. CD147 bears the high frequency Oka blood group antigen (Spring *et al.*, 1997). The CD147 gene has been mapped to band p13.3 of chromosome 19 (Kaname *et al.*, 1993), and the mouse gene has been found to consist of seven exons (Cheng *et al.*, 1994).

1.2.1.2 Cell distribution

The CD147 molecule is broadly expressed on both hemopoietic and non-hemopoietic cells. It is more strongly expressed on thymocytes than on mature peripheral blood T-cells (Kirsch *et al.*, 1997). Furthermore, its expression is upregulated on activated T-cells (Kasinrerk *et al.*, 1992). A high induction was also observed in neoplasms of the bladder, liver and lung (Rizzo *et al.*, 1997; Muraoka *et al.*, 1993).

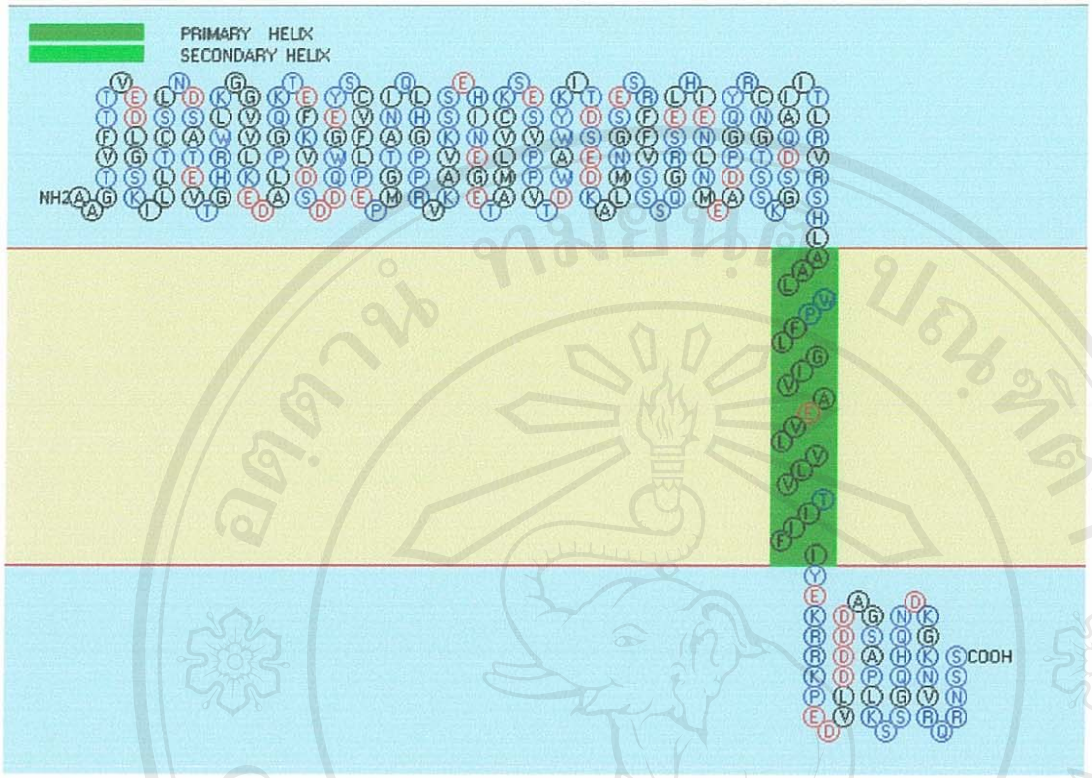


**Figure 1.1. The transmembrane prediction of CD147 molecule by using SOSUI web-base software. The amino acid sequence of CD147 from $Ala_{22}$ to $Ser_{206}$, $Leu_{207}$ to $Tyr_{227}$, and $Glu_{228}$ to $Ser_{269}$ represent extracellular, transmembrane and cytoplasmic regions, respectively.**

**(From http://sosui.proteome.bio.tuat.ac.jp/sosuiframe0.html)**



**1.2.1.3 Functions of CD147**

**In *in vitro* study, cell lysates and cell culture supernatants containing the CD147 antigen stimulated production of interstitial collagenase, gelatinase A, and stromelysin-1 by fibroblasts (Ellis *et al.*, 1989; Biswas *et al.*, 1995; DeCastro *et al.*, 1996; Guo *et al.*, 1997). CD147 seems also to directly bind interstitial collagenase (Guo *et al.*, 2000). These studies suggested an involvement**

of CD147 in regulating stromal matrix metalloproteases. Therefore, CD147 was termed EMMPRIN (Extracellular Matrix Metalloproteinase Inducer) by Chitra Biswas and colleagues (Biswas *et al.*, 1995).

The CD147 monoclonal antibodies (mAbs) AAA6 and UM-8D6 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 (Schiavone *et al.*, 1997). Monoclonal antibodies (mAbs) to the chicken homologue reduced retina cell aggregation (Fadool *et al.*, 1993). Furthermore, it was reported that some CD147 mAbs induced a CD11a/CD18-CD54 dependent aggregation of U937 cells (Kasinrerk *et al.*, 1999). CD147 co-immunoprecipitated with $\alpha_3\beta_1$and $\alpha_6\beta_1$ integrins and to co-localize with this integrins in area of cell-cell contact (Berditchevski *et al.*, 1997). Although not demonstrated yet, it is tempting to speculate that this physical association underlies a functional cooperation. Co-localization of CD147 with the monocarboxylate transporters MCT1 and MCT4 has also been reported (Juel *et al.*, 1999). Studies with a CD2-CD147 chimera implicate the transmembrane and cytoplasmic domains of CD147 in this interaction (Kirk *et al.*, 2000).

CD147 appears to involve also in regulation of T cell functions. The first indication for its T cell regulatory role was given by the finding that it is strongly up-regulated on T cells upon activation (Kasinrerk *et al.*, 1992; Koch *et al.*, 1999). Furthermore, one CD147 mAb, MEM-M6/6, was found to prevent human T cell proliferation in allogeneic mixed lymphocyte responses as well as in T cell receptor (TCR)/CD3 plus CD28 driven highly purified T cell cultures (Koch *et al.*, 1999)

Mice lacking the mouse homologue exhibit infertility of both sexes (Kuno *et al.*, 1998; Igakura *et al.*, 1998), seem to have an abnormality in reception of odor (Igakura *et al.*, 1996), and show a worse performance in short-term memory and latent learning (Naruhashi *et al.*, 1997). Moreover, lymphocytes of such mice give an enhanced mixed lymphocyte reaction (Igakura *et al.*, 1996). Together, these studies indicate a significant role of CD147 in reproduction, functioning of the neuronal system and regulation of the immune system.

1.2.1.4 The CD147 mAbs

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

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

### 1.2.2 Protein expression in *Escherichia coli*

#### 1.2.2.1 Expression of heterologous protein in *Escherichia coli.*

Genetic engineering plays an important role in protein expression works. The expressions in mammalian cells are difficulty and complexity. The expression levels are sometimes low and mammalian cells grow much more slowly than bacteria or yeast (Hinnen *et al.*, 1975). In addition, it is expensive to grow mammalian cells in large quantities. Mammalian cell transfections are generally less efficient, which contributes to lower overall expression levels in those systems. Gene manipulations are very difficult. Mammalian cells might contain oncogenes or viral DNA, thus recombinant protein products must be tested more extensively. Thus, the single cellular microorganism is a selected choice for heterologous protein expression because microorganism such as *E. coli* grows rapidly in simple and inexpensive media and it is easy to avoid contamination in culture step by antibiotic selecting media (Balbas and Bolivar, 1990).

*E. coli* was used as a model for studies relating to prokaryotic genetics, as these, the entire genome of *E. coli* was successfully sequenced. *E. coli* can express the protein of interest in the high level because its cells can grow in the simple and inexpensive media and the technology of fermentation is also well established. Because of it ease of genetic manipulation, *E. coli* has been used as a primary biopharmaceutical production system for many years (Schoner *et al.*, 1984; Tayapiwatna and Kasinrerk, 2002; Intasai *et al.*, inpress).

**1.2.2.2 Problem involving in production of recombinant protein production in *E. coli***

Two major problems are involved with the protein expression in *E. coli*, heterologous protein folding and heterologous protein toxicity (Suter-Crazzolara and Unsicker, 1995). The bacterial cytoplasm is usually not suitable for forming the correctly linked disulfide bonds of the heterologous proteins. Most of proteins need the high oxidizing environment for forming the correctly disulfide bonds (Hartl *et al.*, 1992). However, some proteins with disulfide structure can be correctly formed in bacterial cytoplasm (Blum *et al.*, 1992). For the secretory proteins, the structural genes of them are usually linked to the leader sequence (i.e. *omp*A, *pel*B and etc.). After translation, the signal peptide will convey the protein to periplasm (Halfmann *et al.*, 1993). Periplasm has high oxidizing environment for forming the disulfide bonds and locates among the bacterial cytoplasm and bacterial cell wall (Baker *et al.*, 1992; Walker *et al.*, 1994; Wall *et al.*, 1995). Some recombinant proteins are formed the insoluble inclusion bodies in the bacterial cytoplasm because of misfolding and aggregation of partially folded proteins (Valax *et al.*, 1993). By using the denaturing solvent such as urea and guanidine hydrochloride and reducing agent (i.e. 2-mercaptoethanol), the protein can be purified from the inclusion body (DeBernadez-Clark *et al.*, 1991). Unfortunately, the solubilized inclusion body protein is denatured and therefore, unfunctioned. For protein refolding, the denaturing solvent must be rapidly diluted or removed by slow dialysis. The yield of protein and its function depends upon the protocol which is used to enhance the protein refolding (Puri *et al.*, 1992).

Protein toxicity is the one major problem of protein expression in bacteria (Suter-Crazzolara and Unsicker, 1995). The presence of recombinant protein always effects to the metabolism of bacterial host cells. Protein toxicity can be observed by: bacterial growth rates are dropped down or the times of cell division are expanded, bacterial cells formed aggregation or formed a long chain. These factors cause the host cells lost their abilities to produce the recombinant product. Some recombinant proteins are extremely toxic to host because the presence of some protein may kill host cell or inhibit cell division (Miroux and Walker, 1996). Thus by its protein production should be repressed until later stage of cultivation without inducing any stress responses. In addition, cell cultivation designed according to metabolic pathway and gene expression analysis should be carefully carried out.

### 1.2.3 Protein folding in *Escherichia coli*

#### 1.2.3.1 Factor involving the correct folding of proteins

In prokaryotes, the chaperones are essential for the prevention of protein aggregation (inclusion body) in bacterial cytoplasm. DnaK chaperone of *E. coli* is known to bind some polypeptides and to release or dissociate from them following ATP hydrolysis. Its function is likely a 70 kDa of human heat shock protein (Liberek *et al.*, 1991). Two other crucial chaperones in *E. coli* are GroEL (chaperonin 60) and GroES (chaperonin 10). The crystal structure of GroEL had recently been determined and confirmed (Braig *et al.*, 1994). GroEL is always exposed to the hydrophobic surfaces that are likely to bind to hydrophobic patch on partially folded protein (Mendoza *et al.*, 1994).

Several hydrophobic compact folding intermediates have been observed to bind to GroEL *in vitro*. GroES binding and ATP hydrolysis promotes the acquisition of ordered tertiary structure at the surface of groEL. In addition, they are usually required to release the protein from GroEL in a native or native-like state (Martin *et al.*, 1991). By X-ray crystal structure and mutagenesis studies, the structure and function of GroEL are being elucidated (Braig *et al.*, 1994; Fenton *et al.*, 1994).

1.2.3.2 Protein exportation

Proteins expressed in prokaryote such as *E. coli* may be folded and maintained in cytoplasm or secreted into the periplasm (Figure 1.2.). There is a kinetic competition between folding and exportation because the partially folded protein is competent for exportation (Wickner *et al.*, 1991). By doing mutagenesis of leader sequence and the nascent protein demonstrated that the leader sequence was essential for maintaining the protein in an export competent state (Teschke *et al.*, 1991). SecB is an essential chaperonin which always found in the protein exportation in *E. coli*. It has high affinity for the partially folded cytoplasmic proteins and allows these proteins to be transported (Teschke *et al.*, 1991; Hardy *et al.*, 1991). Some proteins do not need SecB for transportation to periplasmic space. These proteins use the other chaperonin for exportation (Laminet *et al.*, 1991). Overall, it is difficult to achieve the secretion of properly folded protein from *E. coli*. Thus, the most common result of over expression in prokaryotes such as *E. coli* is the formation of inclusion bodies. Extensive work by Mitraki and King has demonstrated that these aggregates are

**generated from self-association of a hydrophobic folding intermediate (King *et al.*, 1986; Mitraki *et al.*, 1991).**

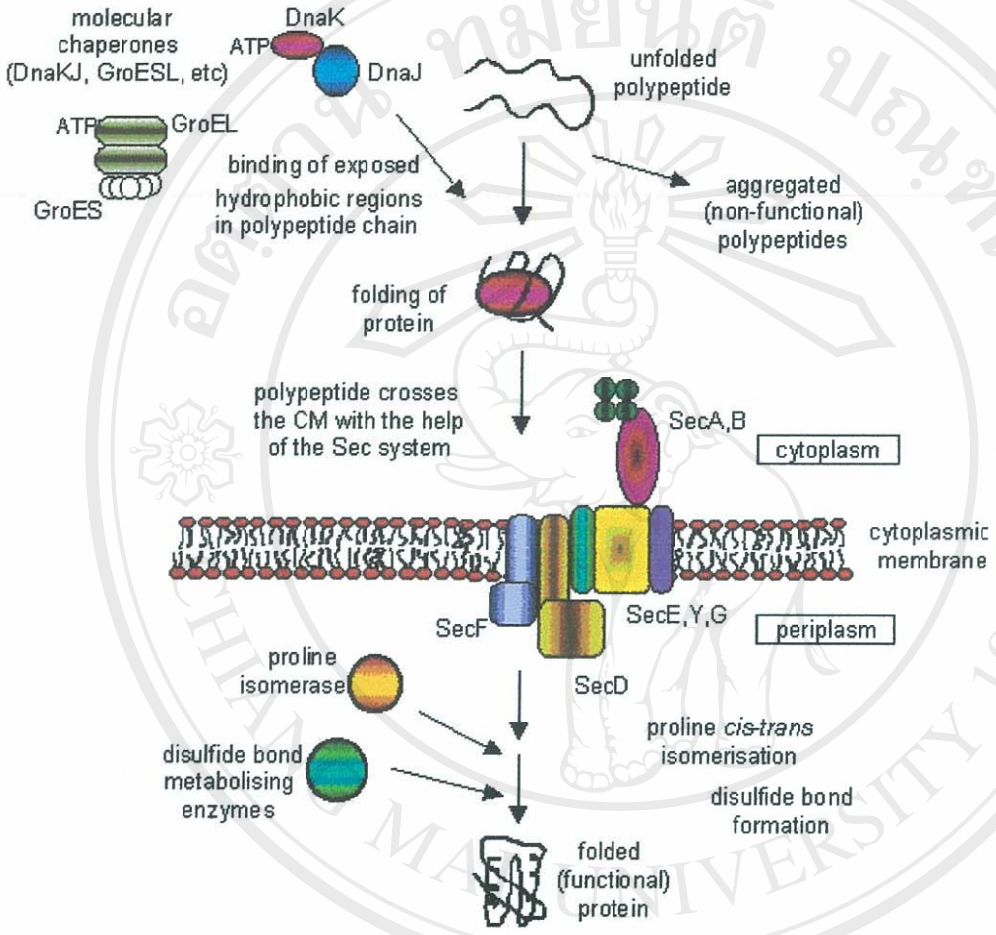


**Figure 1.2. A schematic view of protein folding and exportation. (From http://www.ul.ie/~ces/folding_Mar01.htm)**



### 1.2.4 Phage display technique

**Phage display, first introduced by G. Smith in 1985 (Smith, 1985), is a very effective way for producing large numbers of diverse peptides and proteins and molecule that perform specific functions (Barbas, 1993; Griffith,**

1993; Winter *et al.*, 1994; Burton, 1995; Ladner, 1995; Neri *et al.*, 1995; Hoogenboom, 1998; Ladner, 2000; Seigel, 2001). This technique can also be used to study protein-ligand interactions (Cesareni, 1992), receptor and antibody-binding site (Griffith, 1993; Winter *et al.*, 1994), and to improve or modify the affinity of proteins for their binding partners (Burton, 1995; Ladner, 1995; Neri *et al.*, 1995). Phage display involves the expression of proteins, including antibodies, or peptides on the surface of filamentous phage. DNA sequences of interest are inserted into a location in the genome of filamentous bacteriophage such that the encoded protein is expressed or "displayed" on the surface of filamentous phage as a fusion product to one of the phage coat proteins (Figure 1.3.). Therefore, instead of having to genetically engineer proteins or peptides variant one-by-one and then express, purify, and analyze each variant, phage display containing several billion variants can be constructed simultaneously. These libraries can then be easily used to select and purify a specific phage particle bearing a sequence with desired binding specificity from the non-binding variants.

Two key discoveries were essential for the development of antibody phage display technology. First, the demonstration that foreign DNA inserted into filamentous phage gene III (gIII) is expressed as a fusion protein and displayed on the surface of the phage (Smith, 1985). Second, the successful expression of functional antibody fragment in the periplasmic space of *E. coli* (Better *et al.*, 1988; Skerra *et al*, 1988).

A significant aspect of phage display lies in linking the phenotype of a bacteriophage-displayed peptide or protein with the genotype encoding that

molecule, packaged with in the same virion. This permits the selection and amplification of specific clone of phage representing desired binding sequences from pool of billion of phage clones. In case of filamentous phage, amplification is simply accomplished by infecting male *E. coli*. The genotype-phenotype linkage also permits the rapid determination of the amino acid sequence of the specific binding peptide or protein molecule by DNA sequencing of the specific insert in the phage genome.

Attempts were made to find alternative display methods such as display on bacterial surface (Fuchs *et al.*, 1991), Yeast surface (Boder *et al.*, 1997), or directly on the encoding plasmid DNA (Cull *et al.*, 1992). However, as all these systems still entail transformation of a cellular host, they have not succeeded in generating large diversity libraries.

### 1.2.5 Filamentous phage

#### 1.2.5.1 Structure of filamentous phage

Bacteriophages, or simply phages, are viruses that infect a variety of Gram-negative bacteria using pili as receptors. The Ff filamentous phage particles (strain M13, fl and fd), that infect *E. coli* via F pili, consist of a single-stranded (ss) DNA that enclosed in protein coat. The entire genome of the phage consists of 11 genes. A viable phage expresses about 2,700 copies of gene 8 protein (gpVIII or pVIII, a 50 amino acids residue protein that also known as the major capsid protein) and 3 to 5 copies of the gene III (g3)-encoded adsorption protein (gpIII or pIII, a 406 amino acids protein that is one of three minor coat protein of the filamentous phage) on its tip (Figure 1.3.) (Russel, 1991).

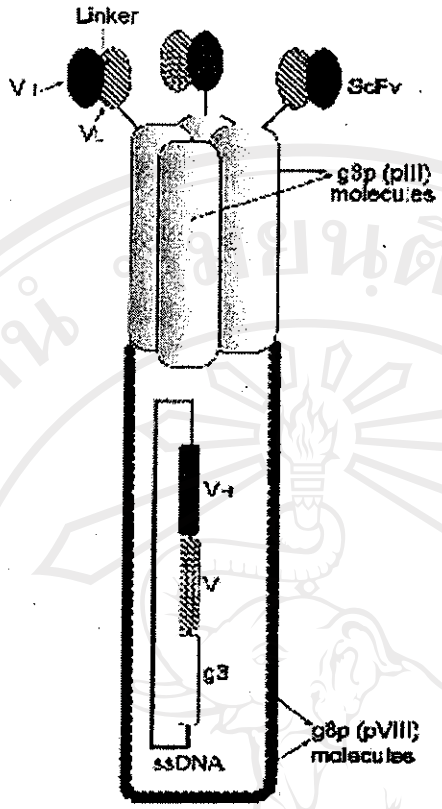


**Figure 1.3. Schematic diagram of a filamentous phage displaying scFv molecule. The phage consists of circular ssDNA surrounded by gpVIII (pVIII) proteins whereas gpIII (pIII) proteins are located at the tip of phage. The genes of scFv are linked to gene III in the phage genome. Consequently, the scFv is displayed as a fusion protein to not all gpIII protein at the tip of phage (Azzazy *et al.*, 2002).**



**1.2.5.2 Life cycle of filamentous phage**

**Filamentous phage does not produce a lytic infection in *E. coli*, but rather induces a state in which the infected bacteria produce and secrete phage particles without undergoing lysis. Infection is initiated by the attachment**

of the phage gpIII to the f pilus of a male *E. coli* (e.g., *E. coli* TG-1). Only the circular phage ssDNA enters the bacterium where it is converted by the host DNA replication machinery into the double-stranded plasmid like replicative form (RF). The RF undergoes rolling circle replication to make ssDNA and also serves as a template for expression of the phage protein gpIII and gpVIII. Phage progeny are assembled by packaging of ssDNA into protein coats and extruded through the bacterial membrane into the medium (Figure 1.4.).

The phage coat protein gpIII and gpVIII are involved in the cloning and detection of recombinant phage antibodies and peptides. Recombinant antibodies, and folded proteins, are typically expressed as gpIII fusion protein and are displayed at the tip of M13 phage. When these antibodies bind to the antigen bound phage is detected with an HRP-labeled antibody that recognizes the gpVIII coat protein. Since several thousand copies of gpVIII exist on the phage surface, it effectively amplifies the detection signal. On the other hand, peptide may be displayed as fusions to either gpIII or gpVIII. If peptides were fused to gpVIII, bound phage can be detected using rat monoclonal antibodies that recognize an epitope localized in the N-terminal portion of gpIII (Dente *et al.*, 1994; Intasai *et al.*, paper in press).

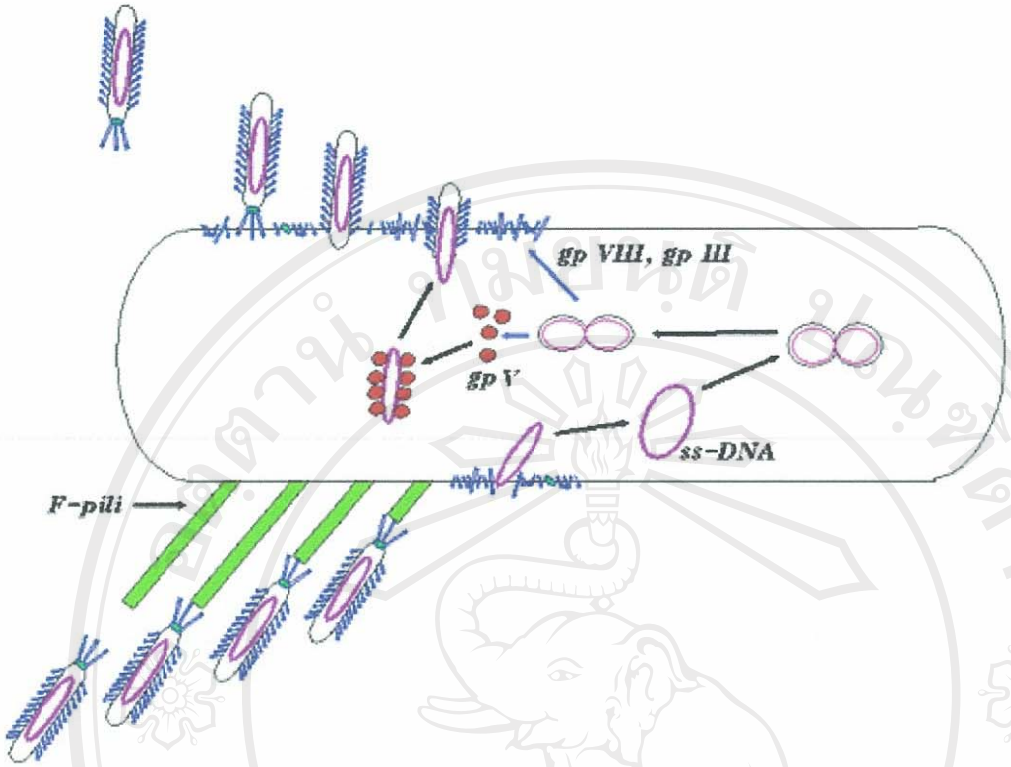


**Figure 1.4. The schematic demonstrates the life cycle of filamentous phage in *E. coli* host.**

**(From http:// www.ams.cmu.ac.th/depts/clinimm/p_chatchai.html.)**

**1.2.5.3 Phagemid cloning vectors**

**With the M13 phage, there are two forms of the phage DNA: ssDNA template that can be easily prepare from phage media and used for sequencing; and dsDNA (plasmid like RF) that can be isolated from the infected bacterial host and used for cloning a target fragment. Phagemids are always used for display. It is a hybrid of phage and plasmid vectors, such as pComb3HSS (Figure 1.5.). Phagemids are designed to contain the origin of replication for both in M13 phage and *E. coli* in addition to gene III, appropriate**

multiple cloning sites, and antibiotic-resistance gene (Mead *et al.*, 1988). However, phagemids lack all other structural and nonstructural gene products required for generating the complete virus. It can be grown as plasmids or alternatively packaged as recombinant M13 phage with the aid of helper phage that contains a slightly defective origin of replication (such as VCSM13 or M13KO7) and supplies all the structural protein for generating a complete phage. This process is called "phage rescue". The resulting phage particles may randomly incorporate either gpIII derived from the helper phage or the polypeptide-gpIII fusion protein, encoded by the phagemid. The ratio of polypeptide-gpIII fusion protein per wild type gpIII may range between 1: 9 to 1: 1,000 depending on the type of phagemid, growth conditions, the nature of polypeptide fused to gpIII and proteolytic cleavage of antibody-gpIII fusions.

The phagemid vector system enables coupling of affinity selection (based on the displayed repertoires of peptides or antibody fragments) to the recovery of the package gene encoding that peptide or antibody. Although this system imposes few limitations such as gene deletion and plasmid instability, it has been successfully used to isolate antibody fragments against a wide range of proteins, DNA, cell surface markers, viruses and parasites. Phagemid vector also allow either the conditional display of antibody on phage or the secretion of the antibody in the periplasmic space of *E. coli* in a form that can be easily detected, e.g., ELISA.

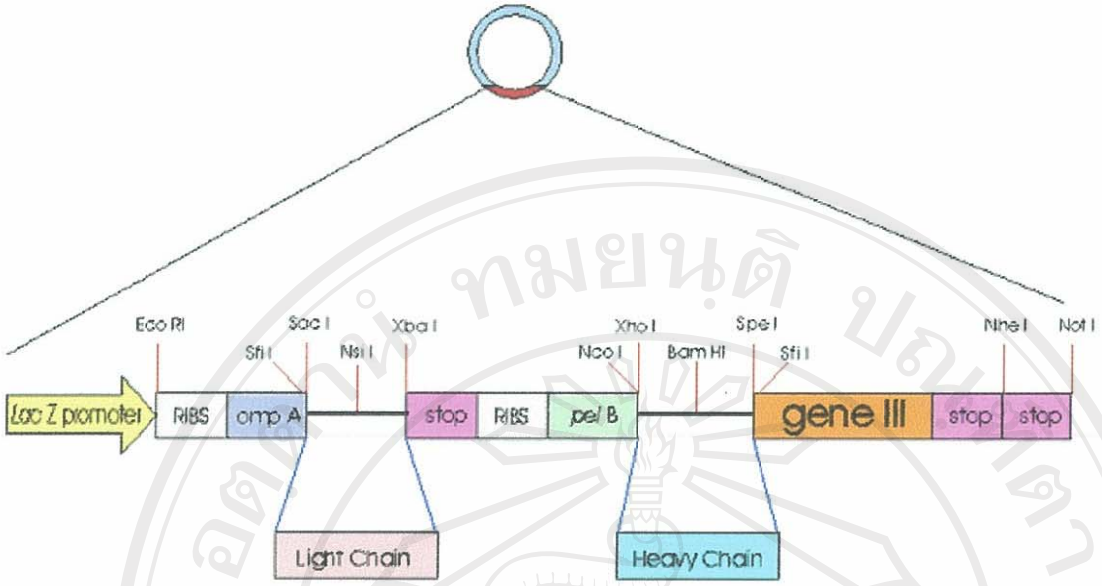


**Figure 1.5. Schematic diagram of pComb3HSS phagemid expressing vector.**

## 1.3 Objectives

**The research was base on 3 purposes:**

**1.3.1 Production of CD147 molecule using phage display technique.**

**1.3.2 Study of the influence of *E. coli* strains on the folding of CD147.**

**1.3.3 Characterization and mapping of the epitope recognized by CD147 mAbs on the molecule of CD147 which linked to phage particle.**

## 1.4 Scope of Study

**This study was divided into 4 parts as follows.**

**1.4.1 Construction of the recombinant phagemid vector containing the extracellular domain gene of human CD147.**

**1.4.2 Expression of recombinant extracellular domain of CD147 by phage display technique.**

**1.4.3 Detection of the phage-displayed CD147 by immunological techniques.**

**1.4.4 Epitope mapping of CD147 mAbs by competitive inhibition ELISA.**