CHAPTER I

INTRODUCTION

1.1 Statement and significance of the problem

The role of highly active antiretroviral therapy (HAART) in controlling HIV-1 is successful. However, HAART is not totally effective in all cases especially in the multi-drug resistant HIV-1, and has problematic side effects. Several strategies have been emerged as novel therapeutic approaches (Check, 2005; Luque et al., 2005; Trkola et al., 2005; von Laer et al., 2006; Berkhout, 2009). Focusing on gene therapy, the most therapeutic genes to be transferred into the cells for evaluating its function were the derivative of immunoglobulin molecules. However, the complex structure of these molecules may limit their intracellular functions.

Since the rising knowledge about protein-protein interactions, superior understanding of protein engineering and the further development of advance selection technologies, novel binding molecules based on protein frameworks 'scaffolds' have been explored to mimic the binding principle of immunoglobulins with virtually target proteins. Several protein-binding scaffolds have been generated to overcome the limitation of antibody and its derivative molecules in terms of stability, ease of modification, robustness and cost-efficient production. Nowadays, different types of scaffold proteins have been successfully used to generate the artificial library and to isolate the specific binding molecules with various targets

(Forrer et al., 2003; Nygren and Skerra, 2004; Binz and Pluckthun, 2005; Hey et al., 2005; Hosse et al., 2006). However, the binding pocket of different protein scaffolds is unique. It might be more or less suited with some targets depending on the nature of the target molecules or their epitopes to which the binding protein should be directed (Nygren and Skerra, 2004).

Focusing on repeat proteins, the repeating structural units (repeats) stack together to form an elongated protein domain (repeat domains). The binding surface of this scaffold is variable in size as the number of repeats can vary (Groves and Barford, 1999; Sedgwick and Smerdon, 1999; Montgomery et al., 2000). One group of protein presenting in this architecture, the ankyrin repeat proteins, is an attractive scaffold to generate the specific binding molecules. These proteins mediate many important protein-protein interactions in virtually all species and are found in all cellular compartments indicating that these proteins can be adapted to function in different environments. Moreover, this scaffold has been used to generate large libraries from which several specific binding molecules were successfully isolated. Therefore, this scaffold appears as a very promising candidate to generate "intrabody" for the desired intracellular function.

Considering the main advantages of ankyrin repeat proteins in term of solubility and stability in the cytoplasm, we aimed to apply these molecules as intracellular interference of HIV-1 assembly and maturation. An artificial ankyrin repeat protein library was constructed by designing special mixed bases corresponding to natural sequences and the special polymerization technique. Specific binding molecules to HIV-1 matrix and capsid domain (MACA) were isolated using the phage display technique. These binders were verified for their binding activity *in*

vitro using various methods. Finally, stable cell lines expressing ankyrin binder in two cellular compartments: as a membrane-bound and cytoplasmic form were generated. These stable lines were subsequently evaluated for their intracellular functions. It was expected that these selected molecules could be able to serve as novel agents for HIV therapy in the future.

1.2 Literature review

1.2.1 Gene therapy for the HIV/AIDS treatment

1.2.1.1 HIV/AIDS treatment

The care of the human immunodeficiency virus (HIV)-infected patient has changed dramatically in the last few years. The approach known as Highly Active Antiretroviral Therapy (HAART) consists of using combination of ARV includes both nucleoside and non-nucleoside inhibitors of the viral enzyme reverse transcriptase (NRTIs and NNRTIs) and inhibitors of the viral protease (PIs) and integrase enzymes. Recommendations for antiretroviral therapy and monitoring are evolving constantly due to the rapid progress in the development of active compounds and new insights into HIV pathogenesis (Arora et al., 2010). The success of HAART in controlling HIV-1 offers hope, but the difficulties of this therapy and the ability of HIV-1 to mutate into drug-resistant variants necessitate continuous development of new therapies. In addition, this regimen is not totally effective and has problematic side effects. Drug-resistant HIV-1 emerges frequently (Wensing and Boucher, 2003; Tang and Pillay, 2004; Lee et al., 2009). Investigation into additional therapeutic approaches should therefore be continued. Gene therapy offers the promise of

preventing progressive HIV infection by sustained interference with viral replication in the absence of chronic chemotherapy. Gene therapy represents one such treatment and several strategies are currently under development. However, this approach will probably not replace pharmacotherapy, but may rather play an important supporting role in the future.

1.2.1.2 Gene therapy for the HIV infection and AIDS

The goals of anti-HIV-1 gene therapy are to deliver transgenes: (a) to hematopoietic progenitor cells (HSC) to protect their differentiated progeny from HIV-1; (b) directly to HIV-1-susceptible cells, to render them resistant to HIV-1 infection or inhibit HIV-1 replication in them; (c) to immunize against HIV-1 antigens; and (d) to inhibit HIV-1 in discrete organ target sites (e.g., central nervous system) (Strayer et al., 2005). Over the past 15 years several different anti–HIV-1 gene therapy approaches have been tested in hematopoietic cells. These approaches can be classified into two categories (**Figure 1.1**): (i) RNA-based agents (including antisense, ribozymes, aptamers and RNA interference (RNAi)); and (ii) protein-based agents (including dominant-negative proteins, intrabodies, intrakines, fusion inhibitors and zinc-finger nucleases). Very recent works suggest that intracellular HIV-1-specific single-chain variable fragment antibodies (ScFv) can target and redirect essential HIV-1 proteins away from required subcellular compartments and block the function or processing of such essential proteins as HIV-1 gp120 (Marasco et al., 1993), Rev (Vercruysse et al., 2010), Gag (Levin et al., 1997), reverse transcriptase (RT) (Bagasra, 1998), and integrase (IN) (Levy-Mintz et al., 1996).
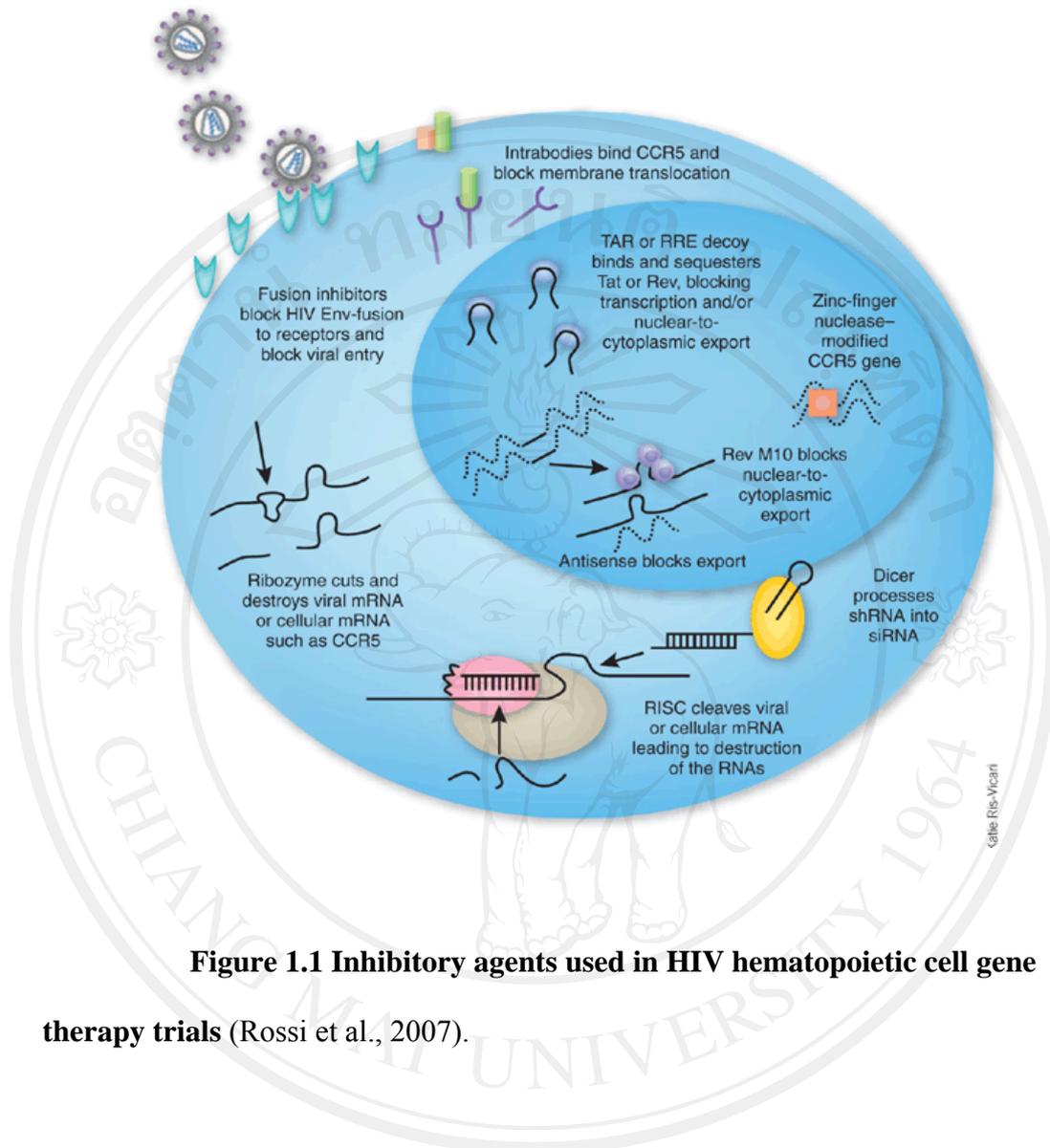
Figure 1.1 Inhibitory agents used in HIV hematopoietic cell gene therapy trials (Rossi et al., 2007).

1.2.2 Role of Gag polyprotein on HIV assembly

The structural proteins of HIV-1 are derived from the Gag polyprotein, which is myristoylated co-translationally. This molecule is composed of three folded polypeptides (Matrix (MA), Capsid (CA), and Nucleocapsid (NC)) and three smaller peptides (SP1, SP2, and p6) as shown in **Figure 1.2A, and B.** Gag molecules assemble at the plasma membrane, and immature virions acquire a lipid envelope as they bud **(Figure 1.2C, and E)**. The viral protease is activated during assembly, and it cleaves Gag to generate a set of mature proteins. These newly processed proteins then reassemble to form the distinct layers of the mature virion **(Figure 1.2D, and F)**. Assembly of retrovirus including HIV-1 is driven by the Gag polyprotein. This process consists of multiple steps starting from the Gag targets to the site of assembly and interacts with lipid bilayer membrane. HIV-MA located at the N-terminus of Gag is a helical domain that displays a conserved patch of basic residues on the same face as the N-terminal myristoyl modification. Membrane binding is mediated by insertion of the myristoyl group into the lipid bilayer and by the basic patch, which binds phosphoinositol (4,5) bisphosphage (PI(4,5)P2, a phosphoinosotide that is concentrated in the raft domain. The multimerization of Gag then occur following by the encapsidation of genomic RNA and the incorporation of Env into virus particles. Finally, nascent viral particles detach from the host cells (**Figure 1.3)** (Ono, 2010; Simons and Gerl, 2010).
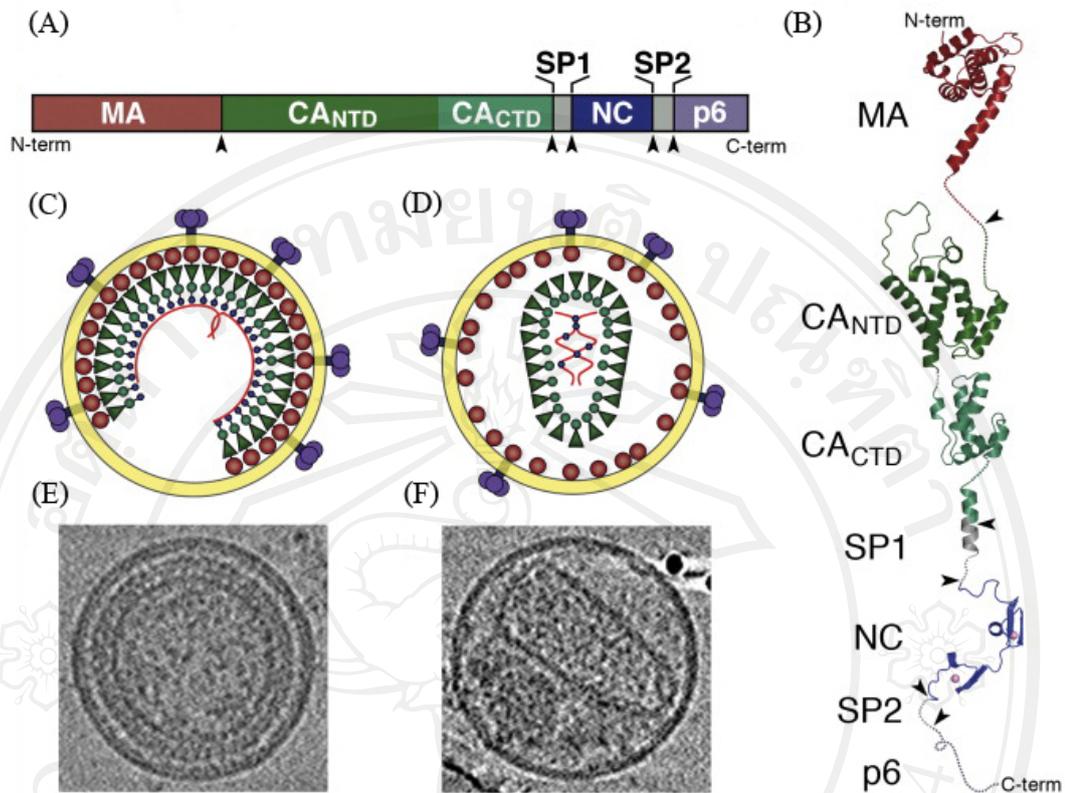
Figure 1.2 The organization of Gag polyprotein. **(A)** HIV-1 Gag polyprotein domain structure, showing the locations of MA, CA_{NTD}, CA_{CTD}, SP1, NC, SP2, and p6. **(B)** Structural model of the extended Gag polypeptide, derived from high-resolution structures and models of isolated domains. Unstructured and linker regions are represented by dashed lines. PR cleavage sites are indicated by the arrowheads. Schematic models represent the immature **(C)** and mature **(D)** HIV-1 virions. Electron cryotomography showed central slices through tomograms of immature **(E)** and mature **(F)** HIV-1 particles. The spherical virions are approximately 130 nm in diameter (Ganser-Pornillos et al., 2008).
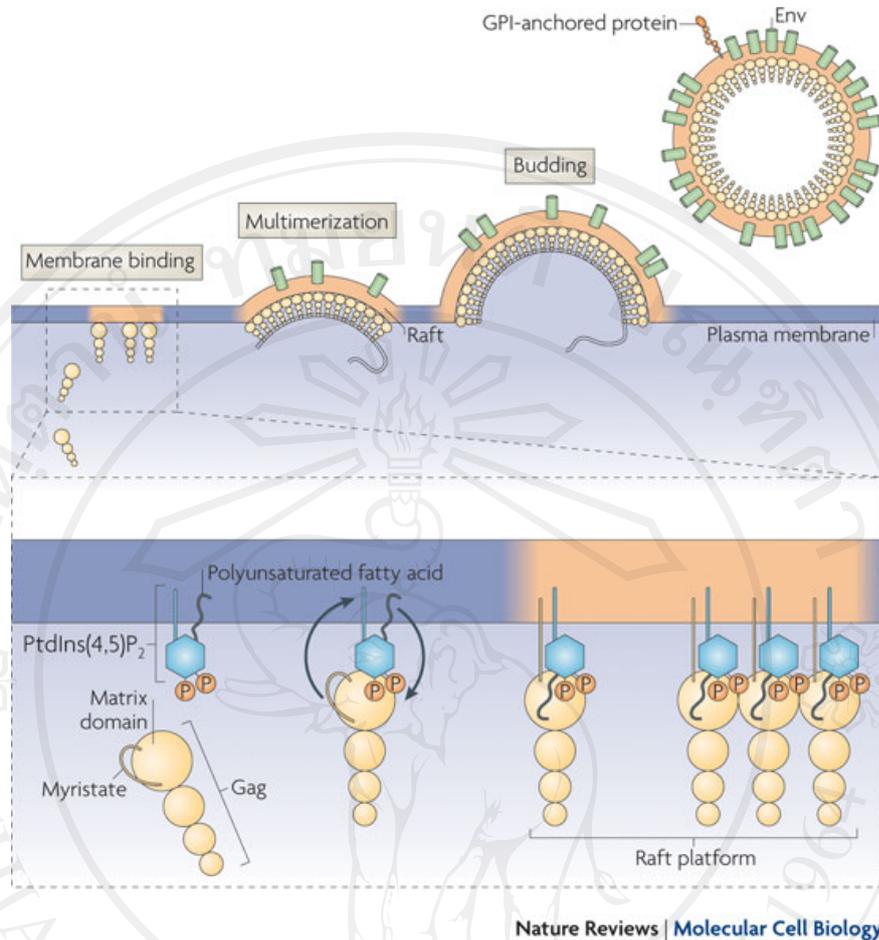
Figure 1.3 HIV assembly and release. Gag binding to phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)P$_2$) through its matrix domain releases the bound myristate group, which inserts into the plasma membrane. The polyunsaturated fatty acid of PtdIns(4,5)P$_2$ has been proposed to flip into the hydrophobic cleft of the matrix domain, and multimerization of the Gag protein stabilizes the raft platform. Gag binding induces the formation of raft platforms at the plasma membrane of the host cell and Gag multimerization drives the assembly of the virus particle. The viral Env protein is incorporated during the budding process and the virus detaches from the plasma membrane (Simons and Gerl, 2010).

1.2.3 Alternative non-antibody scaffolds for novel binding functions

1.2.3.1 Alternative non-antibody scaffolds

Over the past three decades, monoclonal antibodies have been used in a wide range of applications in research, diagnostics and therapy. However, it has become obvious that the whole molecule of antibodies suffers from fundamental disadvantages for example in some therapeutic applications; their Fc regions are not really required. The constant Fc region mediates immunological effector functions and often leads to undesired interactions and rather their large molecules limited tissue penetration (Blakey, 1992; Dearden, 2007). The engineered antibodies in the form of ScFv, Fab and multivalent fragments have been obtained from synthetic libraries or B cells (Dantas-Barbosa et al., 2005; Jia et al., 2008; Bostrom and Fuh, 2009). They are made of two different polypeptides containing the variable part of light (V_L) and heavy (V_H) chains. However, the complex architecture of their antigen-binding site, which is formed by six hypervariable loops, is difficult to manipulate and also requires complicated cloning steps for the recombinant expression and the generation of synthetic libraries. Moreover, the stability of these molecules relies on disulfide bonds which can occur in the appropriate condition in the cells (McAuley et al., 2008).

These limitations and the better understanding of other natural binding proteins inspired scientists to transfer the concept of a universal binding site from antibody structure to alternative protein frameworks ('scaffolds'). The term 'scaffold' in the context of protein engineeting is defined as a polypeptidic framework with a high tolerance of its fold for modifications such as multiple insertions, deletions or

substitutions. This conformational stability enables the directed randomization and drastic change within a defined region of the protein to gain the certain novel properties whereas its structure and original physicochemical behavior remains conserved (Skerra, 2000). The generation of novel binding molecules based on protein frameworks is a concept that has been strongly promoted during the past decades. At present, the different types of protein scaffolds that have been successfully exploited for the construction of artificial binding proteins can be classified into the following groups as shown in **Figure 1.4**. In addition to the specific binding, these artificial binding proteins provide favorable characteristics such as robustness, ease of modification and cost-efficient production.
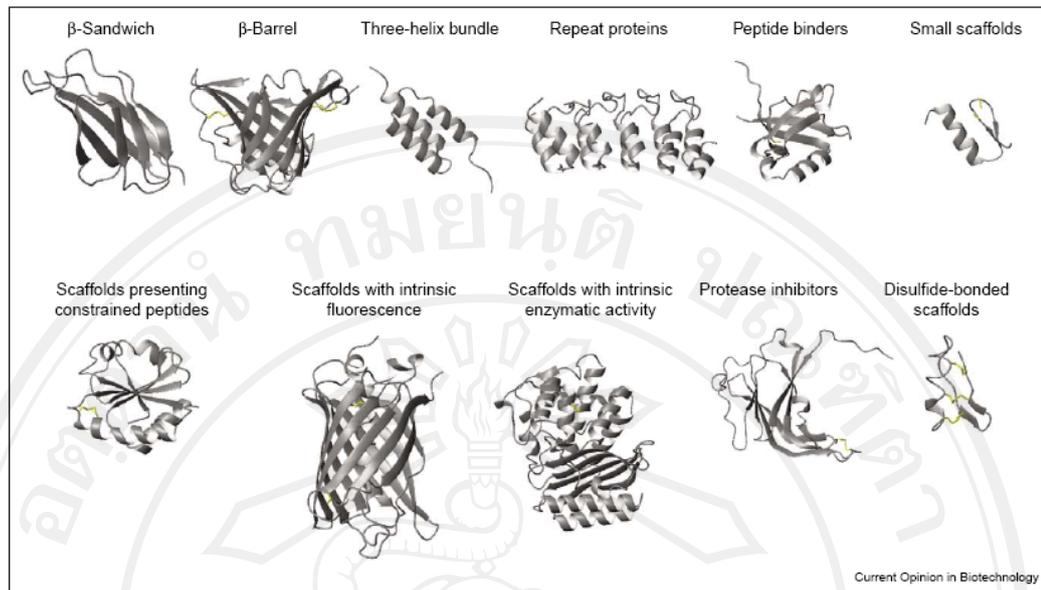
Figure 1.4 The different protein backbones used as scaffolds for the generation of protein-binding agents, classified in groups (Binz and Pluckthun, 2005).

1.2.3.2 Evaluation of repeat proteins

Repeat proteins are the most abundant natural protein classes specialized in binding property. They are found in all phyla and involved in diverse biological processes, such as cell cycle control, transcriptional regulation, cell differentiation, cellular scaffolding or bacterial invasion (Tewari et al., 1998; Sedgwick and Smerdon, 1999; Mosavi et al., 2004; Koprivova et al., 2010). Their unique architecture features repeating structural units which stack together to form elongated repeat domains displaying variable and modular target-binding surfaces (Andrade et al., 2001). The tertiary structures of several proteins with structural repeats are shown in **Figure 1.5**.

Other scaffold motifs such as affibodies (Nord et al., 1997), lipocalins (Skerra, 2001), green fluorescent protein (GFP) (Abedi et al., 1998) and fibronectin type III domains (FNIII) (Koide et al., 1998) are also used as binding molecules. Their architectures consist of a structural framework and variable target-binding surface loops. These scaffolds binding surface is limited by the size. Thus, the repeat proteins have evolved another successful binding strategy. They feature repeating structural units stack together resulting in elongated domains with a continuous target-binding surface which is variable in size as the number of repeats can be varied. For such open structures, there is no theoretical limit on their repeat number since incremental addition of repeats is not sterically hindered. These rod-like or super helical structures present an extensive solvent-accessible surface that is well suited to bind large substrates such as proteins and nucleic acids (Andrade et al., 2001).
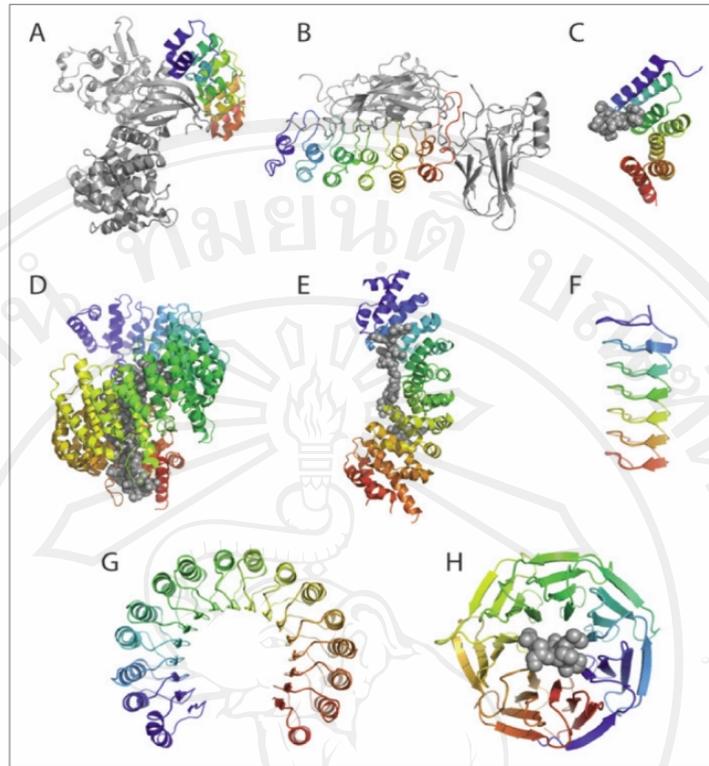
Figure 1.5 The architecture of several repeat proteins. Interaction partners are colored in grey. (A) ANK: the ternary complex between p18^{INK4c}, (B) ANK: The transcription factor complex between NFκB and IκBα (1IKN), (C) TPR: The TPR domain from protein phosphatase 5 in complex with an Hsp90 derived peptide (2BUG), (D) HEAT: Importin β interacting with the IBB peptide of Importin α (1QGK), (E) ARM: Importin α in complex with the NLS peptide (1EE5), (F) β-helix of an antifreeze protein from the *Tenebrio molitor* beetle (1EZG), (G) LRR: The porcine ribonuclease inhibitor (2BNH), (H) WD40: WDR5/Histone H3 complex (2H13). (www.eurekah.com).

1.2.3.3 Ankyrin repeat proteins

In 1987, Breeden and Nasmyth reported a ~33 residue repeating motif in the sequence of two yeast cell-cycle regulators, Swi6p and Cdc10p, and in the Notch and LIN-12 developmental regulators from *Drosophila melanogaster* and *Caenorhabditis elegans* (Breeden and Nasmyth, 1987). Subsequently, the discovery of this sequence containing 24 copies in cytoskeletal protein ankyrin led to the naming of this motif as the ankyrin (ANK) repeat. The ankyrin repeat proteins carry out a wide variety of biological activities and have been detected in organisms ranging from viruses to humans. These molecules are present in the nucleus, cytoplasm and the extracellular milieu. Although, some molecules appear to contain only ANK repeat, others contain insertions between the repeats. Recently, more sophisticated homology search algorithms have been identified ~19,276 sequences in 3608 proteins identified from the non-redundant protein database (http://smart.embl-heidelberg.de/help /smart_glo ssary.shtml) (Li et al., 2006). The ANK repeat is a motif of 33 amino acid residues which exhibits a canonical helix-turn-helix conformation, in which two antiparallel α-helices stacked side by side and connected by a series of intervening β-hairpin motifs. The extended β-sheet projects outward at an approximately 90° angle from the helical pairs resulting in a characteristic L-shaped cross-section (**Figure 1.6**).

From the statistical analyses of ANK sequences by Mosavi et al (Mosavi et al., 2004) and Kohl et al (Kohl et al., 2005a), the consensus sequences of this motif have been identified. There is a consistent pattern of key residues which are well conserved to keep the structural integrity of this motif and some residues which are randomly. The number of repeat found in a single protein varies greatly as shown

in **Figure 1.7**. In 2003, the first combinatorial libraries of ANK repeat named Designed Ankyrin Repeat Protein library (DARPins) was generated from this consensus (Binz et al., 2003). Several specific binders have been isolated from this library using ribosome display and phage display (Amstutz et al., 2005; Kohl et al., 2005b; Huber et al., 2007; Vogel et al., 2007; Zahnd et al., 2007; Interlandi et al., 2008).
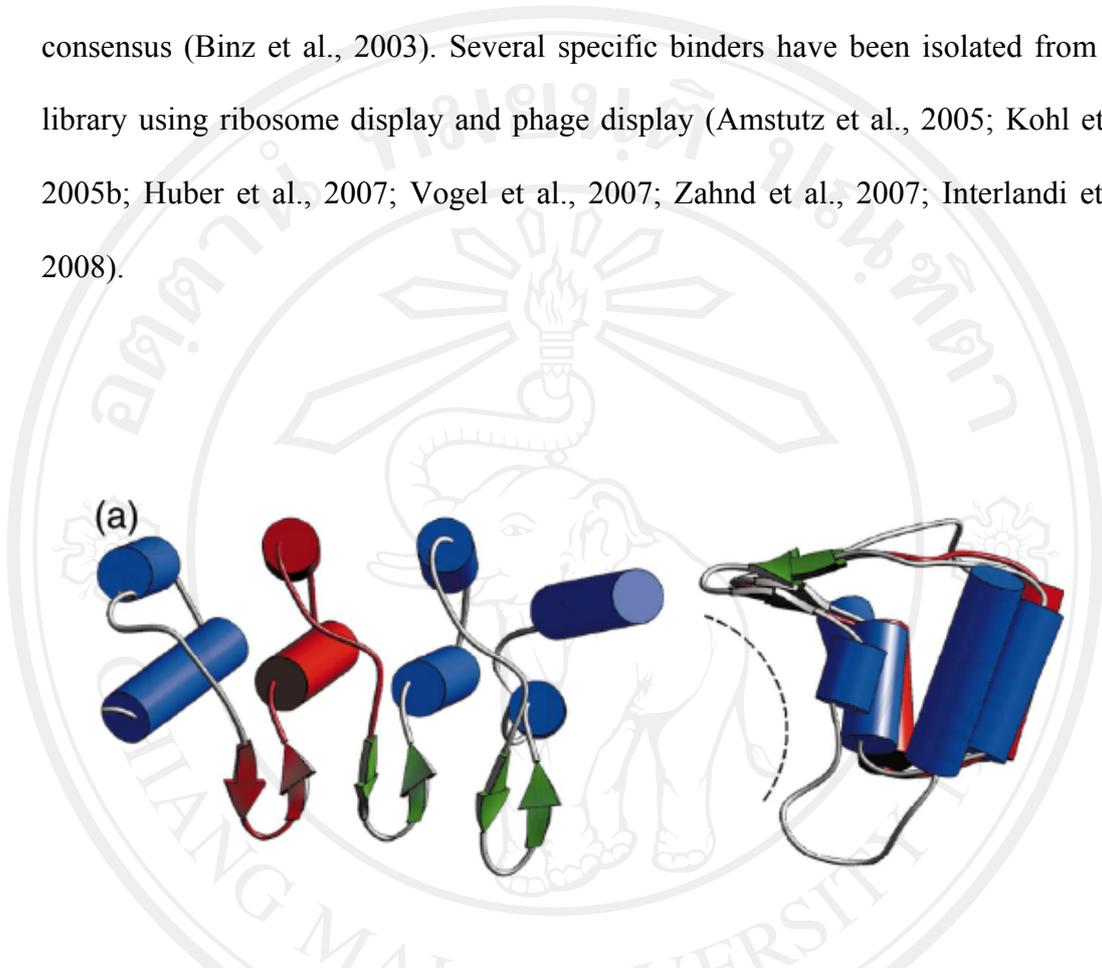
Figure 1.6 Characteristic of ankyrin repeat (ANK) protein architecture.
The arrangement of α-helices and β-hairpins is shown as cylinders and arrows, respectively (Sedgwick and Smerdon, 1999).
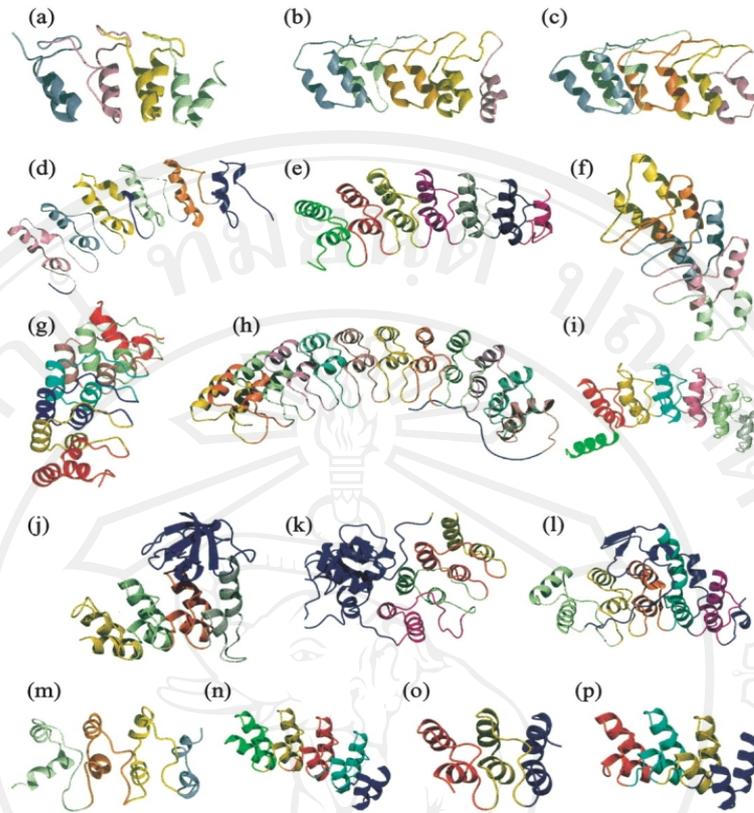
Figure 1.7 Ankyrin repeat proteins with high-resolution structures in the PDB. The ankyrin repeats have been colored differently to illustrate the packing interactions present in this fold. Cell cycle inhibitors (a) p16 (PDB ID: 1BI7), (b) p18 (1IHB), and (c) p19 (1AP7); (d) Iκ-Bα, inhibitor of Nf-κB (1NFI); (e) Bcl-3, a unique member of the Iκ-B family (1K1A); (f) Transcription factor GABP-β (1AWC); (g) Human oncoprotein G ankyrin (1UOH); (h) D34 region of human Ankyrin-R (1N11); (i) Ankyrin repeat domain of *Drosophila* signaling protein Notch (1OT8); (j) Tumor suppressor 53BP2 (1YCS); (k) Signaling protein PAP-β (1DCQ); (l) Transcriptional regulator Swi6 (1SW6); (m) Cardiomyogenic hormone myotrophin (2MYO). Designed ankyrin repeat proteins (n) Sank E3_5 (1MJ0), (o) 3ANK (1N0Q), and (p) 4ANK (1N0R) (Mosavi et al., 2004).
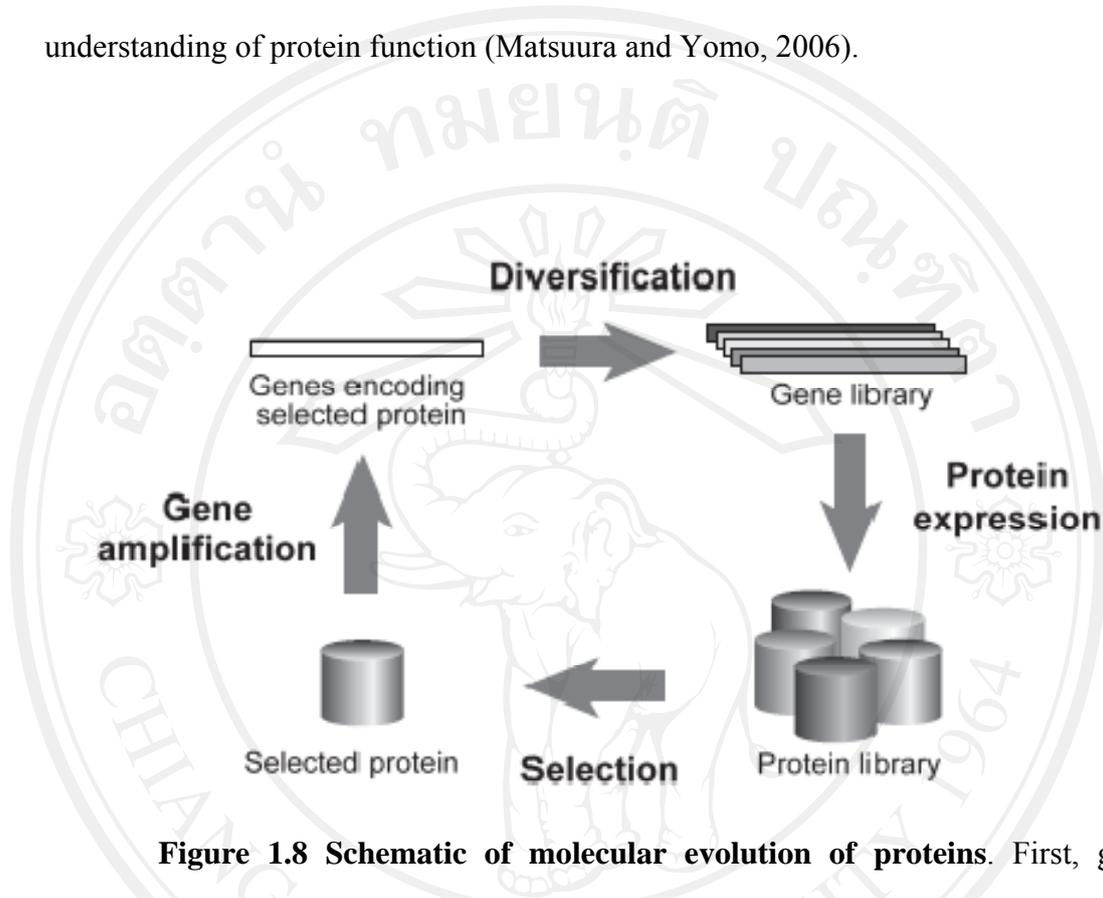
1.2.4 *In vitro* evolution of protein

1.2.4.1 Introduction

Proteins posses a broad range of structural and functional properties that are unmatched by any other class of biological molecules. Nature has inspired many scientists and engineers to design and create their own customized proteins. The concept of laboratory created protein has been documented since the 1970s (Hall, 1973). Recent advances in technology in the field of molecular biology have allowed us to mimic this process at the molecular level and to evolve the desired functions of molecules. Directed evolution is a method used in biology to exploit the power of natural selection and to evolve proteins or RNA with desirable properties not found in nature. More specifically, directed protein evolution is a general term used to describe a reange of methods associating the generation of large collection of protein (variants) and selection of desirable functions (Yuan et al., 2005). Over the last three decades, directed protein evolution has emerged as a powerful technology platform in protein engineering. These engineered proteins can serve as novel molecular tools for scientific, medical, and industrial applications, thus addressing many needs unmet by naturally occurring proteins.

A typical directed evolution experiment involves mainly two steps which are diversification and selection as shown in **Figure 1.8**. First step, the diversification is the process to create large libraries of variant genes by mutation and/or recombination. Second step, selection is the process to isolate molecules or protein possessing the desired properties. However, for efficient evolution, experiments must be designed appropriately, *i.e.*, with appropriate diversification and

selection strategies. Nevertheless, molecular evolution has proven to be a powerful strategy not only for generating proteins for various applications but also for better understanding of protein function (Matsuura and Yomo, 2006).

Figure 1.8 Schematic of molecular evolution of proteins. First, gene encoding the protein of interest is subjected to mutagenesis for the generation of gene library. Subsequently, proteins are expressed from the gene and evaluated for their functions. Then, proteins with the desired properties are selected (Matsuura and Yomo, 2006).

1.2.4.2 Primary sequence analysis

Evolution provides a vastly useful model for protein design. Many families of proteins contain hundreds or thousands of members spread across diverse species. By considering the common features of the sequences, it is possible to presume the key elements that determine protein structure and function even in absence of any explicit structural information. In order to take this approach, several tools are needed. First, given one or a few sequences of a target structure, it is necessary to be able to search through the vast array of known sequences for related protein. Second, this large family of related proteins must be aligned such that conserved positions are in register with one another. Analysis of the degree of conservation at each position can give important insight applicable to protein engineering. Computational methods for this analysis of primary sequence are well established, and many tools are available through Web-based servers **(Table 1.1).**

One of the most straightforward applications of primary sequence data is the use of multiple-sequence alignments to define consensus motifs for a particular structure or function. These sequence signatures focus on the common features of a class, while not corresponding to any natural sequence. The fundamental approach is as follow: first, known sequences of the target protein family are selected. Second, the profile of these sequences is searched in global sequence databases for additional family members. Third, the multiple sequence alignment is performed on this large set of sequences and then the occurrence of each amino acid at each position is measured the statistical enrichment. Finally, this information is used to bias the selection of sequences in an engineering context (Green, 2004). In general, consensus sequences

will be defined by selecting positions that are conserved above a certain threshold. The randomized positions can be also reasonably definite for the creation of protein library.

Table 1.1 Web services and databases for primary sequence analysis

Service	Web Location (URL)	Description
GenBank	http://www.ncbi.nlm.nih.gov/Genbank/	Repository of all publicly available nucleotide sequences.
Swiss-Prot	http://expasy.org/	Annotated database of protein sequences.
BLAST	http://blast.ncbi.nlm.nih.gov/Blast.cgi	Online service to search for related sequences.
ClustalW2	http://www.ebi.ac.uk/tools/clustaw2/index.html	Online service for multiple sequence alignment.
ClustalW	http://www.clustal.org/	Downloadable software for multiple sequence alignment.
Pfam	http://www.sanger.ac.uk/resources/databases/pfam.html	Database for classifying protein sequences

1.2.4.3 Diversification of molecules

In order to modify or optimize an existing protein, a library of variant genes is designed and constructed and is usually a compromise between different requirements. First, library members need to be sufficiently similar to sequence of the starting protein to share a similar structure and function. Second, library members need to be sufficiently different in sequence from the starting protein to be slightly different in structure and thus in the functional property of interest. Other factors that contribute to library design and construction are the difference between accessible library size and the theoretical sequence space of interest, the limited number of approaches, the use and control of randomization, the natural versus synthetic origin of the diverse population and library quality.

Nowadays, several established methods have been proposed. The three major classes of library-construction methods applied to protein engineering are based on random mutagenesis, recombination and site-directed diversification. The random mutagenesis mimics the incidence of such errors over the millions of years of natural evolution, but vastly increases the rate of mutagenesis by artificially increasing the error rate of DNA replication such as UV light, x-ray radiation, chemical mutagen, and error-prone PCR (Fenton et al., 2002). Recombination method such as DNA shuffling mimics a second mechanism of natural evolution in the exchange of pieces between related genes using homologous recombination (Stemmer, 1994). Site-directed diversification is the method that directs to a specific position or set of positions, and the remaining protein sequence is fixed as wild-type (Dordick, 1990).

The most commonly used method is to synthesize a set of oligonuclotides where the wild-type bases in each codon of interest are replaced by mixtures of degenerated nucleotides. Conventional methods for codon randomization employ NNN-, NNB-, NNS- or NNK-containing oligonucleotides (N: A/C/G/T; B: C/G/T; K: G/T; S: G/C), as these combinations, each degenerated codons encode all 20 amino acids with high proportion of stop codon (Patrick and Firth, 2005). A superior strategy for improving the library quality and diversity is to use the reduced codon sets such as MAX, SND, or VDR (M: A/C; D: A/G/T; V: A/C/G; R: A/G) which have no probability of encoding a stop codon (Hughes et al., 2003; Urvoas et al., 2010).

1.2.4.4 Selection of desired functional proteins

The second step, selection is the process to isolate molecules or protein possessing the desired properties. The selection process is typically repeated a number of time, such as in cycle number of selection, to enrich molecules with desired properties. After selection and identification, the selected novel protein is typically produced and characterized in more detail. This step needs the display technologies which physically link the phenotype (polypeptides or proteins) displayed on a certain platform to its corresponding genotype (gene). This enables easy identification and amplification of the selected polypeptides via the nucleic acids, being DNA or RNA. The different technologies can be divided into three categories: cell-dependent display systems, cell-free display systems and non-display systems as shown in **Table 1.2** (Gronwall and Stahl, 2009).

Table 1.2 The different selection systems employed in combinatorial protein engineering (Gronwall and Stahl, 2009)

Selection system	Illustrative references
Cell-dependent systems	
Phage display	Barbas et al. (2001)
E. coli surface display	Daugherty et al. (1998)
Staphylococcal surface display	Löfblom et al. (2005)
APEx *E. coli* display	Harvey et al. (2004)
Yeast display	Boder and Wittrup (1997)
Cell-free systems	
Ribosomal display	Hanes and Plückthun (1997)
mRNA display	Nemoto et al. (1997); Roberts and Szostak (1997)
CIS display	Odegrip et al. (2004)
DNA display	Tabuchi et al. (2001)
Covalent DNA display	Bertschinger et al. (2007)
Microbead display	Sepp et al. (2002); Nord et al. (2003)
Non-display systems	
Yeast-two-hybrid	Parrish et al. (2006)
PCA	Koch et al. (2006)

In the cell-dependent display systems, the polypeptides are displayed on the surface of cells or phage particles, or expressed in a cellular compartment. The most utilized system in this group is phage display (Barbas, 2001a). Focusing on phage display, this technique was first described in 1985 by George Smith *et al*. They demonstrated that the linkage between phenotype and genotype could be established in filamentous bacteriophage and established the new technology of phage display (Smith, 1985a). The foreign DNA fragment is inserted into the genome of the filamentous phage and the encoded foreign peptide is displayed as a fusion to one of the coat proteins on the surface of phage. Since then, phage display has been used as a powerful method to isolate target-specific polypeptides with high affinity (Paschke, 2006). More details on the principle of phage display technique will be described in the section below.

The major advantage with cell surface display systems especially the yeast and bacterial display is the possibility of isolation of antigen binding molecules by fluorescence-activated cell sorting. Typically, fluorescently labeled target is added to the cell displayed protein library in solution and the library is screened with high-speed flow cytometry. A large advantage is that the flow cytometry technique gives the possibility of monitoring and quantifying the relative affinities during the selection process. In addition, no elution of target protein from the affinity proteins is required, since the cells are isolated with bound target protein (Kronqvist et al., 2008; Rockberg et al., 2008; Gronwall and Stahl, 2009). However, there are some limitations using cell-dependent system in the library size which depends on the transformation efficiency of DNA, and the crucial fact that the target protein needs to

be available and preferably in its native form. This limitation has led to the exploration of different cell-free systems.

The *in vitro* display technologies or cell-free systems are powerful tools to select polypeptide binders against various target molecules. The main common feature is the concept of *in vitro* transcription and translation for construction of protein libraries. Consequently, these systems do not require transformation of DNA encoding the combinatorial library into a host. This enables the creation of libraries up to 10^{13} members. Furthermore, these systems allow the possibility to introduce *in vitro* mutagenesis during the amplification which can result in a directed evolution in every selection round (Roberts, 1999). However, the limitations are that the molecular size of affinity protein needs to be small enough to allow efficient *in vitro* transcription and translation and that these are not "user friendly" since they required special reagents and expert researchers. The examples of successful cell-free systems are ribosomal display and mRNA display as shown the principle in **Figure 1.9.**
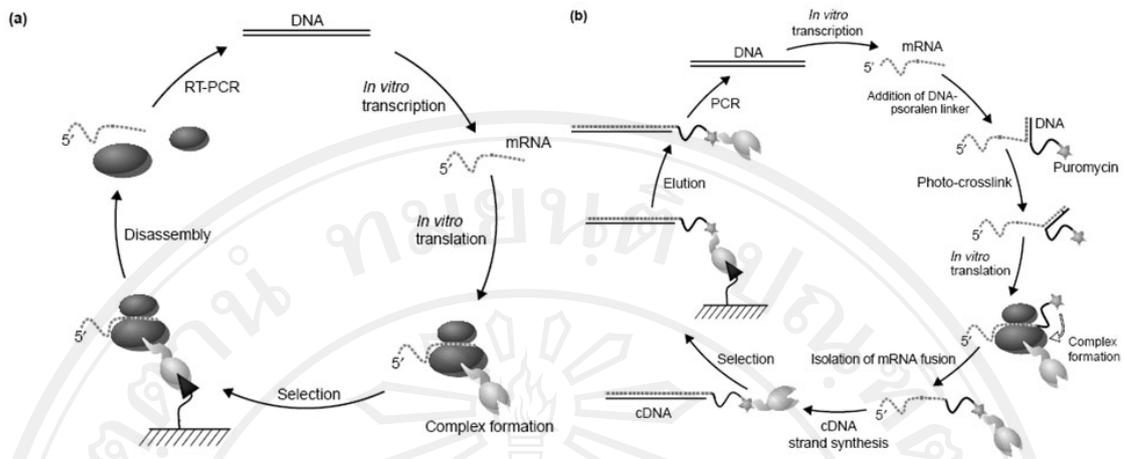
Figure 1.9 *In vitro* display technologies. (A) Schematic representation of a ribosome display selection round. **(B)** Schematic representation of a mRNA display selection round, modified from (Amstutz et al., 2001).

Ribosomal display and mRNA display methods are the DNA sequences encoding the protein library with either a ribosome-binding site containing no stop codon for ribosomal display or covalently link to polypeptide through the antibiotic puromycin for mRNA display, are transcribed and translated *in vitro*. This results in a stalling of ribosome and coupling of the coding mRNA to the translated protein. The mRNA-ribosome-protein complex can then be subjected to a target molecule and various binders can be isolated in a manner similar to panning in phage display. The isolated clones can be recovered by isolation of the RNA and amplified by reverse transcription PCR to generate and enrich a second library that can be used as input to the next selection round.

Moreover, there are additional selection systems that are not based on display of libraries and selection by incubation with a target molecule and isolation of binders. This technique so-called non-display system, the target protein is co-expressed with the individual library members *in vivo*, and the selection is thus not dependent on an available target protein, but instead the target protein can be expressed in a correctly folded form by the host organism (Koch et al., 2006; Parrish et al., 2006). These *in vivo* systems commonly rely on the protein interaction between affinity protein and target generating growth survival or fluorescence activity. Focusing on the protein complementary assay (PCA) (Pelletier, 1998; Remy and Michnick, 2007) as shown the method in **Figure 1.10**, a reporter protein which is usually an enzyme essential for survival of the host is split into two parts and turned into inactive form. Two different vectors are constructed. One vector contains the gene encoding the target protein fused to a gene fragment encoding one part of a reporter protein. The second vector contains the combinatorial library in fusion to another part of the reporter protein. The binding of affinity protein to the target will restore the enzyme activity of reporter protein resulting in survival of the cell. Advantages of this system are that high-throughput selections could in theory be achieved by direct screening for growth survival and that no pre-available target protein is requires for the selection. However, there are still challenges due to unspecific intracellular interactions and poor affinity discriminations that have to be overcome in the selection of new affinity ligand. Moreover, the protein scaffold application must be well suited for intracellular folding to increase the success.
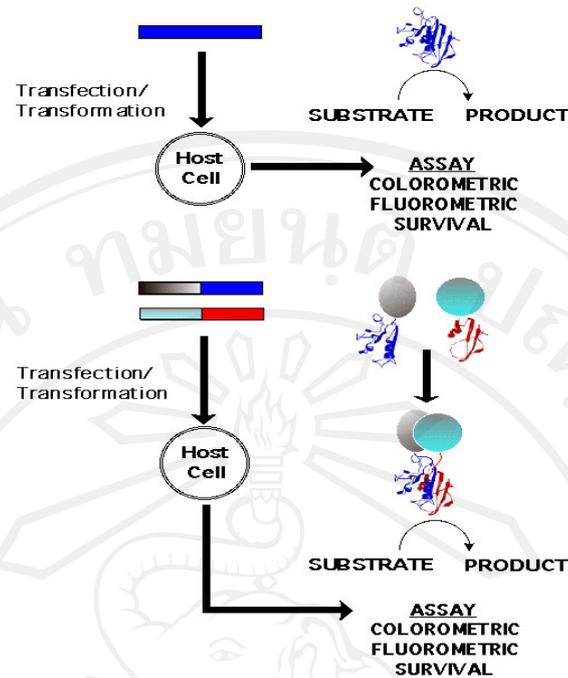
Figure 1.10 General features of a protein complementary assay (PCA).
Top, left, a gene encoding enzyme can be transformed/transfected into a host cell and its activity detected by an *in vivo* assay (right). Bottom, left, oligomerization domains A, B, are fused to N- and C-terminal fragments of the gene for the enzyme. Co-transformation or transfection of oligomerization domain-fragment fusions results in reconstitution of enzyme activity by oligomerization domain-assisted reassembly of the enzyme (right). Reassembly of enzyme will not occur unless oligomerization domains interact (Pelletier, 1998).

1.2.5 Recombinant protein expression

1.2.5.1 Systems for recombinant protein expression

In recent years, the number of recombinant proteins used for several applications including therapeutic and research has increased dramatically. These demands have driven the development of a variety of improvements in protein expression technology. It is now possible to express recombinant proteins using a variety of expression systems such as bacterial, yeast, insect and mammalian systems. Each system has its own advantages and disadvantages as shown in **Table 1.3** (Yin et al., 2007). In order to achieve a desirable expression, the cloned gene must be transcribed and translated efficiently. The yield and biological activity of recombinant proteins differ greatly, and depend on a large number of factors such as solubility, stability and size of protein. Every protein poses unique problems in its expression because of its unique amino acid sequence. In general, the choice of expression system depends on many factors such as the facility and experience of the laboratory, the source of protein of interest and its toxicity in the cells, the modification of expressed protein, the scale of protein production, the purification strategy, the cost of production, safety consideration, etc (Verma et al., 1998). Fundamentally, it should be noted that different proteins pose their own properties and problems and so no 'universal' expression system can be used. The appropriate method for production of protein of interest may be choosen depending on the main characteristics of each host as shown in **Table 1.4.**

Table 1.3 Applications of expression systems in common used (Fernandez, 1999).

Host system	Advantages	Disadvantages	Common applications	Representative commercial products
E. coli	Scale-up, low cost and time, easy operation	Protein solubility, may require protein-specific optimization	Structural analysis, antibody generation, functional assays, protein interactions	pET Expression Systems (Novagen), pGEX Vectors (GST Gene Fusion System), pTrc (Amersham), pQE (Qiagen), pLEX, pBAD, pTrc, pPRO EX HTa, b, c (Invitrogen), PRO Bacterial Expression System Vectors (Clontech), PinPoint^{TM} Xa (Promega), pFLAG-ATS, pT7-FLAG (Sigma)
Yeast	Eukaryotic protein processing, scalable up to fermentation, simple media requirements	Fermentation require for very high yield, growth conditions may require optimization	Structural analysis, antibody generation, functional assays, protein interactions	pPIC, pYC, pYES2 (Invitrogen), pYEX-BX, pGBKT7 (Clontech)
Insect cells	Near mammalian protein processing, greater yield than mammalian systems	More demanding culture condition	Functional assays, structural analysis, antibody generation	pBAC^{TM} transfer plasmids (Novagen), pFastBac^{TM} pMelBac (Invitrogen), BacPak Baculovirus Expression Vectors (Clontech), pPolh-FLAG^{TM} (Sigma)
Mammalian cells	Highest level protein processing	Relatively low yield, more demanding culture condition	Functional assays, protein interactions, antibody generation	pTandem^{TM}-1, pTK-neo (Novagen), pcDNA, pBudCE4.1, pcDNA^{TM}, pBC1, pVAX1 (Invitrogen), pMAM, pCMS-EGFP (Clontech), pRluc-C, pGFP2-C (Perkin-Elmer), HaloTag^{TM} pHT2, pSI, phRG, phRL, pCI (Promega), pFLAG-CMV, pBICEP-CMV (Sigma)

Table 1.4 The characterization of expression systems in common used (Fernandez, 1999; Yin et al., 2007)

Host system	Cell growth	Cost of medium	Expression level	Post-translational modifications						
				Protein folding	N-linked glycosylation	O-linked glycosylation	Phosphorylation	Acetylation	Acylation	Gamma-carboxylation
E. coli	Rapid (30 min)	Low	High	Refolding usually required	None	No	No	No	No	No
Yeast	Rapid (90 min)	Low	Low–high	Refolding may be required	High mannose (mostly)	Yes	Yes	Yes	Yes	No
Insect cells	Slow (18–24 h)	High	Low–high	Proper folding	Simple, no sialic acid	Yes	Yes	Yes	Yes	No
Mammalian cells	Slow (24 h)	High	Low–moderate	Proper folding	Yes	Yes	Yes	Yes	Yes	Yes

1.2.5.2 Recombinant protein expression in *E. coli*

To date, *E. coli* has been used extensively as the cellular host for heterologous protein production. This is due to its ability to grow rapidly and at high density on inexpensive substrates, its well-characterized genetics, ease of genetic manipulation, many supporting expression vectors and mutant host strains have been established and relatively inexpensive. Generally, there are three forms of foreign proteins expressed in *E. coli* which are fusion proteins, secreted proteins and inclusion bodies. Fusion proteins are formed by fusing the gene of interest to other sequences encoding tag domain for example a poly-histidine tag, a FLAG epitope tag, a fluorescent protein, etc. This can further enhance the solubility of proteins, purification, or monitoring, and often affords protection against degradation by intracellular protease as shown in **Table 1.5.** The secreted proteins can accumulate in either the periplasm or extracellular medium. This process can reduce potential degradation by intracellular protease, simplify the purification procedure, and contribute to the correct folding structures of proteins secreted into the medium. The last fraction, inclusion bodies are largely protected from proteolytic degradation by host cell enzyme. However, these systems are unable to properly perform the post-translation modifications such as glycosylation, fatty acid acylation, phosphorylation and especially disulfide-bond formation, which are often required for proper folding of the compact structure and for the functional characteristics of the interested protein. These requirements lead to the accumulation of the expressed proteins in an insoluble and inactive form in the cytoplasm. Recovery of the active protein from these inclusion bodies requires the appropriate strategies which normally result in the low yields of active and soluble protein from the total expressed protein.

Table 1.5 The list of commonly used tag for fusion protein production (modified from (Esposito and Chatterjee, 2006))

Some commonly used solubility-enhancing fusion partners		
Tag	Protein	Source organism
MBP	Maltose-binding protein	*Escherichia coli*
GST	Glutathione-*S*-transferase	*Schistosoma japonicum*
Trx	Thioredoxin	*Escherichia coli*
NusA	N-Utilization substance	*Escherichia coli*
SUMO	Small ubiquitin-modifier	*Homo sapiens*
SET	Solubility-enhancing tag	Synthetic
DsbC	Disulfide bond C	*Escherichia coli*
Skp	Seventeen kilodalton protein	*Escherichia coli*
T7PK	Phage T7 protein kinase	Bacteriophage T7
GB1	Protein G B1 domain	Streptococcus sp.
ZZ	Protein A IgG ZZ repeat domain	*Staphylococcus aureus*

Some commonly used affinity purification tags		
Tag	Protein	Affinity matrix
His6	Hexahistidine tag	Metal chelates
GST	Glutathione *S*-transferase	Glutathione
MBP	Maltose-binding protein	Amylose
FLAG	FLAG tag peptide	Anti-FLAG antibody
BAP	Biotin acceptor peptide	Avidin
Strep-II	Streptavidin-binding peptide	Streptavidin
CBP	Calmodulin-binding peptide	Calmodulin

1.2.5.3 Recombinant protein expression in yeast

The main advantage of yeast over the other expression systems is related to the fact that it is both a microorganism and a eukaryote. Comparing to *E. coli*, the similar advantages are culture simplicity, rapid growth, and low cost. Furthermore, yeast shares many genetics, and biochemical features with higher eukaryotes and is suitable for large-scale industrial fermentation. Therefore, it provides advanced protein folding pathways for heterologous protein that have undergone all the post-translational modifications resulting in soluble form of expressed protein. In addition, yeast can secrete correctly folded proteins into culture medium, when yeast signal sequences are used, and also processed proteins modification such as glycosylation. Moreover, the degradation of heterologous proteins which is often a problem in *E. coli*, is usually reduced in this system (Verma et al., 1998). The first yeast expression platform was based on the traditional baker's yeast *Saccharomyces cerevisiae*. Although successfully applied to the production of pharmaceuticals like insulin or hepatitis B vaccines, some important disadvantages became quickly apparent and limit its general use in biotechmology. Glycoproteins are often over-glycosylated, and terminal mannose residues in N-linked glycans are added by an α-1, 3 bond which is suspected to be allergenic. Instead, non-allergenic terminal are found to be present in *Hansenular polymorpha* and *Pichia pastoris* (Guengerich et al., 1991). A further benefit is that strong promoters are available to drive the expression of a foreign gene of interest, facilitating the production of large amounts of target protein more inexpensively than most other eukaryotic systems.

1.2.5.4 Recombinant protein expression in insect cell *via* baculovirus expression system

The Baculovirus Expression Vector System (BEVS) has been widely used in research and scientific industrial communities for the production of high levels (up to 1000 mg/mL) of properly post-translationally modified (folding, disulfide bond formation, oligomerization, glycosylation, acylation, and proteolytic cleavage), biologically active and functional recombinant proteins. The concept of this system is based on the introduction of a foreign gene into a plasmid-based transfer vector, which is then incorporated into genome of the baculovirus *via* homologous recombination resulting in recombinant baculovirus encoding heterologous protein which can be expressed in cultured insect cells and insect larvae (Huynh and Zieler, 1999). Currently, the viral vectors used in this system have mainly been isolated and constructed form *Autographa californica* nuclear polyhedrosis virus (*Ac*NPV) or *Bombyx mori* nuclear polyhedrosis (*Bm*NPV). *Ac*NPV can infect dozens of moth species and all *Spodoptera frugiperda* ovarian cell lines. In contrast, only cells of silkwarm origin are susceptible to *Bm*NPV. Nearly, a thousand of high-value foreign proteins have been successfully produced in the system and the insect baculovirus may be applied in production of vaccines (Kang, 1997), gene therapy (Ghosh et al., 2002; Cao and Wu, 2006) and recombinant baculovirus insecticides (Bonning et al., 1999; Assenga et al., 2006).

Therefore, this system is inferior to prokaryotic and yeast systems in terms of its capacity for continuous fermentation. Nevertheless, there are some limitations, for examples, the host cells which are infected by nuclear polyhedrosis virus will eventually die, so the heterologous gene cannot be expressed continuously.

Every round of protein production requires the infection of new insect cells. Moreover, insect cells and mammalian cells differ in their glycosylation patterns, such as in the lengths of oliogsaccharides and in mannose content, so bioactivity and immunogenicity of insect expression products are somewhat different from those of the natural product (Kulakosky et al., 1998). However, researchers have successfully addressed this limitation by transforming established insect cell lines with constitutively expressible mammalian genes. This approach has yielded the transgenic insect cell lines with normal growth properties that can support viral infection, have new N-glycan processing enzyme activities, and can produce humanized recombinant glycoproteins (Jarvis, 2003; Harrison et al., 2006).

1.2.5.5 Recombinant protein expression in mammalian cell

Mammalian cell expression systems have several advantages over the systems discussed above in that they promote signal synthesis, process, and can secrete and glycosylate proteins, particularly eukaryotic proteins. This system can be used for either transient or stable expression, depending upon the purpose of the expressed product. Although the relatively high cost, complicated technology, and potential contamination have been bottlenecks for its use in large-scale industrial production. This system is often utilized to express many heterologous proteins including viral structural protein and bioactive peptide for specific functional analysis. The examples of well-known systems are COS-cell-based transient expression systems (COS TES) and Chinese hamster ovary (CHO) cell stable expression systems (CHO SES) (Yin et al., 2007).

Moreover, newly expression systems have been emerged including both individual animal and individual plant expression system. Modern molecular biological techniques allow researcher to transform recombinant DNA into fertilized animal ovum which can be integrated into the host genome at certain frequency, and expressing the foreign protein in specific tissue or organs of the body depending on the promoters used. Until now, a variety of transgenic animal species have been used to produce recombinant proteins or built as disease models.

1.2.6 Phage display technology

1.2.6.1 Introduction

In 1985, Smith first demonstrated that an exogenous protein can be expressed on the surface of the filamentous M13 phage (Smith, 1985a). This was achieved by inserting the gene that encoded a part of the *Eco* RI endonuclease into the ORF of the phage's minor capsid protein III (pIII) and also detected the ability of phages carrying the chimeric *Eco*RI-pIII protein to specifically bind the antibodies. Furthermore, phage with this insertion could be selected from a mixture containing wild-type phages by affine enrichment using antibody against the *Eco* RI endonuclease. These experiments led to two important conclusions. First, using DNA-recombination method, it is possible to create phage populations of different representatives (10^6-10^{11} variants), of which each individual phages displays a random peptide on its surface. Such populations are named "combinatorial phage libraries". Second, a direct physical linkage is created between the displayed protein

and its encoding gene. This linkage endows the protein with the two key characteristics of molecular evolution , replicability and mutability (Smith and Petrenko, 1997).

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 various 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).

1.2.6.2 Biology and structure of M13 filamentous bacteriophage

Filamentous bacteriophages are virus capable of infecting a variety of gram-negative bacteria through the F conjugative pilus of cells. The most characterized class of these viruses is the Ff class including M13, fd and f1 strains. The filamentous bacteriophages particle consists of a circular single-stranded DNA genome surrounded by a proteinaceous coat. The entire genome includes 11 genes, and is encased in a long protein capsid cylinder with diameter of 7 nm and a length of 900-2000 nm. The length of the cylinder is composed of approximately 2,700 copies of major coat protein, called gene VIII protein (pVIII). Fives molecules each of pVII

and pIX are present at one end of the particle, while the other end harbors 5 molecules each of pIII (minor coat protein) and pVI as shown in **Figure 1.11** (Gao et al., 1999) and **Table 1.6** (Barbas, 2001a). Protein VI and pIII are crucial for host recognition and phage infectivity, whereas pVII and pIX are required for phage assembly (Gailus et al., 1994). The pIII is made up of three domains separated by glycine-rich regions. 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 (CT domain) is essential for forming a stable phage particle (Smith and Petrenko, 1997; Barbas CF. III, 2001; Barbas, 2001b; Sidhu, 2001).
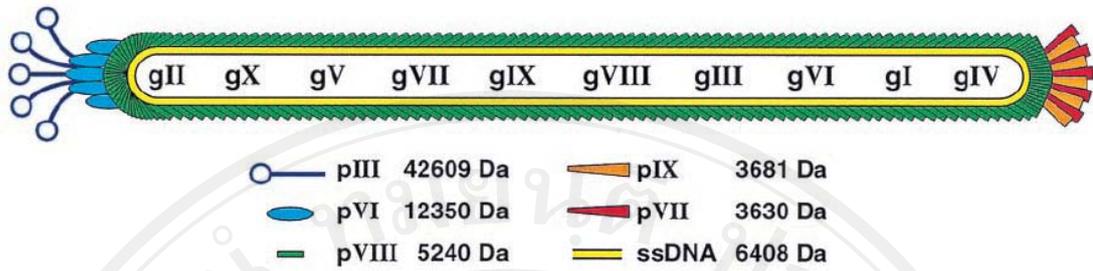
Figure 1.11 Structure of a filamentous bacteriophage. A diagram of the bacteriophage particle represents the single-stranded DNA core surrounded by a proteinaceous coat (Gao et al., 1999).

Table 1.6 The list of 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

Phage infection is initiated by attachment of the phage through the N2 domain of the pIII end to the tip of F pilus of *E. coli* cell and releaseof the N1 domain which becomes free to interact with the C terminal domain of TolA (TolAIII). Docking of the phage is thought to be followed by retraction of the pilus, bringing the pIII end of the phage near the surface of the bacteria (**Figure 1.12**). Through the action of TolA, the major capsid protein, pVIII, and probably the pVII and pIX minor coat proteins depolymerize and become integrated into the host inner membrane. The viral single-stranded circular DNA is translocated into the cytoplasm and then the complementary strand is synthesized by bacterial enzymes resulting in the double-stranded DNA called the parental replicative form (RF) DNA. The (-) strand of this RF is a template for transcription and the resulting mRNAs are translated into all phage proteins. Of the 11 phage-coat proteins, three molecules (pII, pX, pV) are required to generate ssDNA, three (pI, pXI, pIV) for phage assembly and the other five proteins (pIII, pVI, pVII, pVIII, pIX) are components of the phage particle.

Assembly process 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 1.13**). 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.
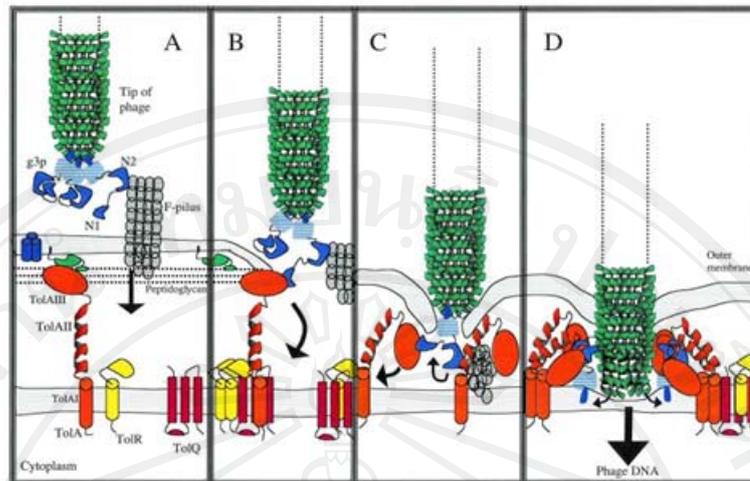
Figure 1.12 Model of filamentous phage infection. A) The phage N2 domain of pIII protein interacts with the F-pilus on the outside of the bacteria. The outer membrane proteins OmpF (blue cylinders) and Pal lipoprotein (greenish) are also included, as there have been reports of TolA interacting with these 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).
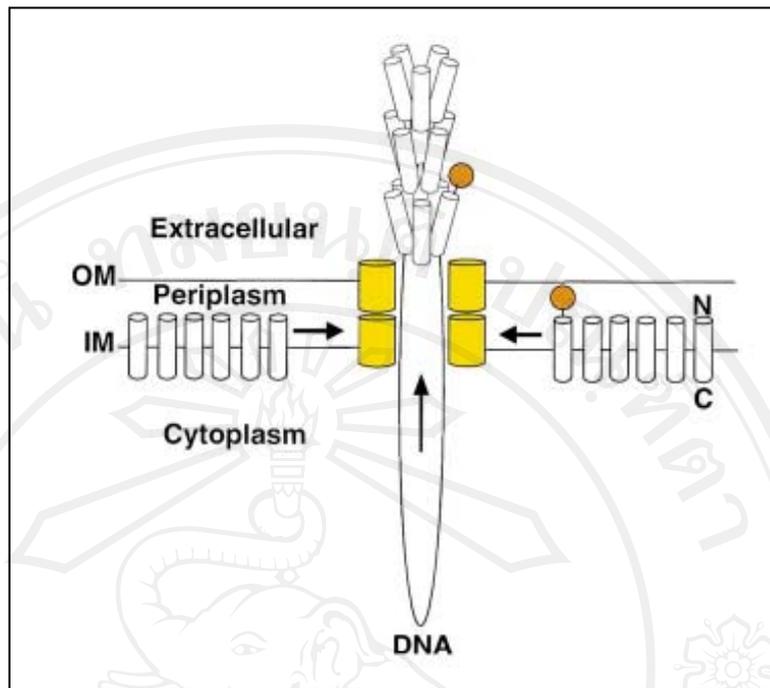
Figure 1.13 M13 bacteriophage assembly. Newly-synthesized coat proteins (white cylinders) are embedded into the inner membrane (IM) with their N termini in the periplasm and their C termini in the cytoplasm. Single-stranded viral DNA is extruded through a pore complex (yellow cylinders) that spans the 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).

1.2.6.3 Filamentous phage-based vector systems

A number of novel vector molecules have been constructed 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 at N-terminus of phage coat protein (pIII or pVIII), and an antibiotic-resistance gene for selection and propagation (Barbas, 2001b). 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 particles. They also carry a kanamycin resistance gene to allow antibiotic selection of helper-infected cells. 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 order to restore infectivity of phages with inserts, specialized vector systems were developed using introduction of an additional gene encoding wild-type pIII or pVIII protein. The vector systems, termed 33 or 88 (**Figure 1.14**), carry both the wild-type phage gene and the recombinant gene. The vector systems referred to as 3+3 and 8+8 **(Figure 1.14)** have the recombinant gene on a phagemid, while the additional wild-type gene is introduced into the *E. coli* cells *via* a "helper" phage. In both cases, the replicated phages carry both normal and hybrid proteins and can replicate, despite the presence of exogenous inserts. Focusing on the 3+3 system which is mostly used for selection

of antibody fragments, the main characteristic of the 3+3 system is its virtual monovalence. Both recombinant and wild-type pIII proteins are exposed on the surface of the phage particle. The copy number of recombinant protein varies from 0 to 5 for each single virion. Notably, only 10 % of the phages carry even one copy of the chimeric protein, and the percentage of phages carrying 2 or more molecules of recombinant pIII is considerably smaller. About 90 % of the phages carry no chimeric protein (Griffiths and Duncan, 1998). This low valence leads to limited avidity, which in turn allows the selection of high-affinity molecules. However, the ratios of polypeptide-pIII fusion protein to wild type pIII may range between 1:9 and 1:1000 depending on the type of phagemid, growth conditions, the nature of the polypeptide fused to pIII, and the proteolytic cleavage of antibody-pIII fusions (Azzazy and Highsmith, 2002).
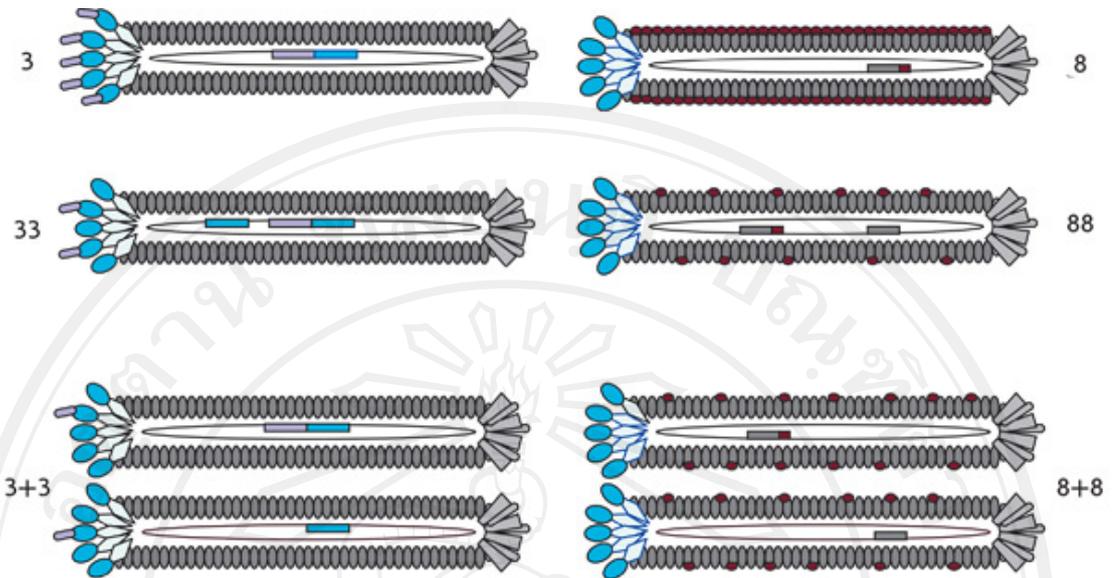
Figure. 1.14. Types of vector systems for peptide and protein display based on filamentous phages. Lilac and red represent the exogenous inserts into the genes encoding the pIII and pVIII proteins, respectively. Foreign peptides displayed on pIII are either appended to the N-terminal domain (type 3 systems) or replace the N-terminal domain (type 3+3 and mostly type 33 systems). In type 8 systems, the foreign peptide is displayed on all copies of the major coat protein pVIII (2700 copies in wild-type virions), whereas in type 88 and 8+8 systems, only a minority of the pVIII copies display the foreign peptide. (http://actanaturae.ru/article.aspx?id=174)

1.2.6.4 Applications of phage display

Since the phage display reported in 1985 (Smith, 1985b), this technology has rapidly evolved into an efficient tool used by structural biologists for the discovery and characterization of diverse ligand-receptor-binding interactions. Furthermore, 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 isolating 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 carrying antibody gene can be easily analyzed the sequence, mutated, and screened to improve binding activity. Moreover, phage display of a functional protein has now become a standard first step prooving of principle prior to the application of combinatorial strategies (Bass et al., 1990; Roberts et al., 1992). A natural functional domain can represent the end product of a highly directed evolutionary process. Moreover, phage-display approaches can create variations of the domain with altered binding affinity or 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).

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 particles contains only one type of recombinant coat proteins encoded by the corresponding gene fusion present inside the same phage particle. Individual phage can be rescued from libraries by an interaction of the displayed protein with the cognate ligand by a panning step, which allows selection 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).

1.3 Objectives

1.   To design and construct the artificial ankyrin repeat proteins library.

2.   To produce the fusion protein of HIV-1 matrix and capsid (H_6MA-CA) target protein using baculovirus expression system.

3.   To select the specific H_6MA-CA binders using phage display strategy.

4.   To investigate the binding activity of H_6MA-CA binders.

5.   To generate the Sup-T1 stable cell line expressing ankyrin binders for further evaluating the intracellular functions.

6.   To evaluate the role of selected binders in the interference of viral assembly and viral maturation *in vitro*.