CHAPTER 1

INTRODUCTION

More than 9 million people worldwide are in HIV treatment which are typically includes the use of combined antiretroviral drugs known as highly active antiretroviral drugs therapy (HAART). The primary goals of HAART are to reduce HIV-related illness, prolong survival, enhance quality of life, bring back immunological response and also prevent HIV transmission. Nevertheless, the emergence of multi-drug resistant mutants and the undesirable side effects of HAART are the major obstacle for the adequate management of HIV infection [1]. The development of a safe and effective HIV-1 vaccine would definitely be the best solution for the ultimate control of the worldwide AIDS pandemic, unfortunately HIV-1 vaccine development efforts have not yet been proven successful [2, 3]. Consequently, the infancy of HIV-1 vaccine gives rise to the alternative therapy, such as protein-based and gene therapy.

HIV entry is the first key step in HIV life cycle. In ordered to prevent the HIV to propagate, there are many extensive studies to invent novel HIV entry inhibitors. Antibody-based therapeutic agents and also their engineered derivatives as seen in neutralizing antibody against HIV infection have been broadly studied. However, the proper structure of antibody, including disulfide bond formation and folding requires special culturing system. To overcome this limitation, many studies have concentrated on other scaffold molecules with high specificity and high affinity properties to target protein comparable to antibody. Furthermore, the high solubility and stability of the alternative protein scaffolds promote them to be the ideal candidates against antibody.

Designed Ankyrin Repeat Proteins (DARPins) are protein architecture built up with varying numbers of structural motifs stack to form the repeat modules. According to

the ability to be selected against the desired targets, DARPins have been raised as alternative to antibodies [4]. The previous study showed that CD4-specific DARPin bound specifically to CD4 molecule and successfully inhibited the binding of HIV gp120 to CD4 *in vitro* [5]. Despite the efficient binding of CD4-specific DARPin to CD4+ cells *in vivo*, the inhibition of simian-human immunodeficiency virus (SHIV) in SHIV-infected rhesus macaques was not successful due to the rapid clearance from the plasma with the rate proportional to the dosage [6]. In order to improve the binding affinity *in vivo*, further study on the binding site and molecular modification needs to be done.

The computational study on the binding of CD4 molecule and CD4-specific DARPin using a combinatorial protein docking criteria and mathematical calculation unveils the crucial amino acids. This study aims to evaluate the key amino acid of these two molecules to further understand the mechanism the DARPin used to inhibit HIV-1 entry. With the application of adenoviral vector to generate mutated CD4 molecules on cell surface, the interaction of CD4 and CD4-specific DARPin will be investigated by flow cytometric analysis. Combining these studies together, the improvement of this CD4-specific DARPin will be performed and encourage this molecule to be one of the potent HIV-1 entry inhibitor *in vivo* in the near future.

1.1 Literature review

1.1.2 HIV entry

The HIV entry is the first key step to HIV replication cycle. The key element on HIV particle in order to enter host cell is the HIV envelope (Env). The Env protein is type I integral membrane protein. It is first synthesized in endoplasmic reticulum (ER) as gp160 precursor. The gp160 travels to Golgi apparatus where it is cleaved into gp120 and gp41 subunits by host protease furin. A trimer of heterodimers are formed, since the soluble surface of gp120 binds to the transmembrane portion of gp41.

The process often, but not always, begins with the attachment of viral envelope protein to the attachment factors on host cell surface, which can be either nonspecific via heparin sulfate proteoglycans or specific via α4β7 integrin or dendritic cell specific intercellular adhesion molecule 3-grabbing non-integrin (DC-SIGN) [7]. In consequence, the HIV envelope will come closer to the viral receptor CD4 and coreceptor, raising the infection efficiency. Nonetheless, this event may not be necessary as the mechanism *in vivo* remains unclear.

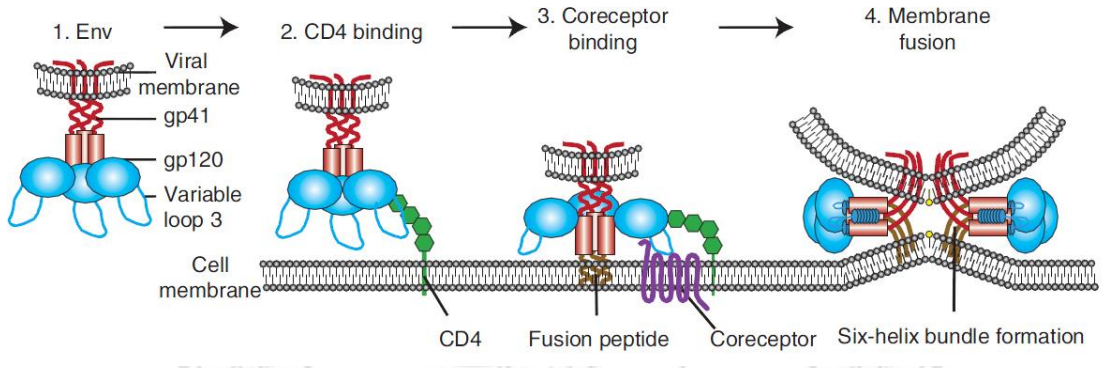

Figure 1.1 Schematic view of HIV entry, introduction of HIV gp120 to CD4 (2) induces the conformational change in gp120. The exposed V3 loop in gp120 binds to host cell coreceptor (3), bringing the viral particle close to host cell to form the membrane fusion (4) [8].

The major phenomenon of HIV entry is the binding of HIV gp120 to its primary receptor, CD4. The conformational change occurs in the viral gp120, resulting in the rearrangement of V1/V2 and the exposure of V3 loop. The binding site for chemokine coreceptor lies within the V3 region, which is the main factor to determine whether CXCR4 or CCR5 will be used. The virus strains which favor CXCR4 are termed X4-tropic, infecting both primary $CD4^+$ T-cells and T-cell lines. CCR5 is utilized in R5-tropic virus which can infect macrophage and $CD4^+$ T-cells but not T-cell lines.

The virus then moves to the site where productive membrane fusion occurs. Finally, the viral envelope compromises the membrane fusion as the hydrophobic gp41 fusion peptide is uncovered and inserted into host cell membrane. This phenomenon exposes amino-terminal helical region (HR-N) and carboxy-terminal helical region (HR-C) from each gp41 to be organized into a six-helix bundle (6HB) [9]. The formation of

membrane fusion pore leads to the dismissal of HIV contents into host cell cytoplasm. However, the number of HIV Env trimers needed to complete membrane fusion is not clarified [8]. The overall mechanism of HIV entry is illustrated in Figure 1.1.

1) CD4 Structure

CD4 is 58 kDa type I integral membrane glycoprotein found on the surface of thymocytes, T helper lymphocytes and cells in macrophage or monocyte lineage. The main three parts of CD4 structure are composed of the extracellular domain, transmembrane domain and short cytoplasmic tail. The extracellular domain is formed up with four immunoglobulin domains, which made the molecule later being classified into the immunoglobulin superfamily. These domains organize themselves as a rod-like unit.

The early x-ray crystallographic study revealed the structure of CD4, mainly on domain 1 and domain 2 (D1 and D2) (Figure 1.2A). D1 has nine β strands forming into two β sheets, one sheet with strand AGFCC′C″ and the other with strand BED. D2, with only seven β stands, has two β sheets, one sheet with strand GFCC′ and the other with strand ABE. The disulfide bond found in D2 is in a non-standard location, connecting the C and F strand within a sheet. These two domains connect each other by the opposite ends of the two IgV-like structures. The strand G of D1 continues without any interruption to form the strand A of D2 [10]. Further study on the crystal structure of rat CD4 on D3 and D4 provided the homologous structure of D3 to D1 and the smaller D4 to D2 [11]. The G strand of D3 continuously fuses to the A strand of D4. Unlike D1, D3's disulfide bond which connects the B-strand of one sheet to the F-strand of another sheet is absent (Figure 1.2B) [12].

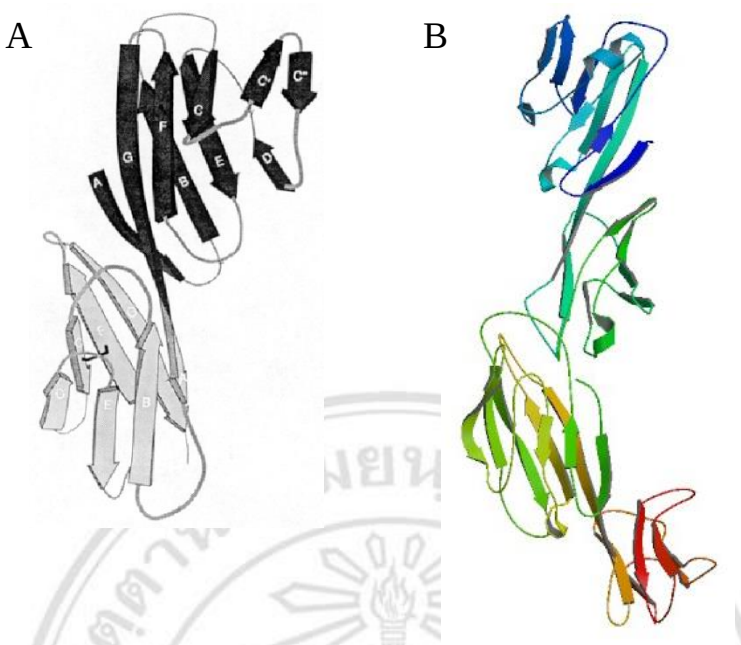

Figure 1.2 Structure organization of human CD4 D1 (black) and D2 (gray) [10] (A) and the whole extracellular domain, D1 – D4 (from top to bottom) (B). (http://www.rcsb.org/pdb/images/1wiq_bio_r_500.jpg?bioNum=1)

2) HIV-1 envelope: gp120 and gp 41

HIV-1 Env glycoproteins are initially synthesized as a polyprotein precursor gp160 on rough endoplasmic reticulum (RER). *N*-linked and some *O*-linked glycosylation occurs at the same time as translation. Trimers of gp160 are formed in ER but dimers and tetramers can be observed. In Golgi complex, the precursor gp120 is cleaved by furin or furin-liked proteases to surface glycoprotein gp120 and transmembrane glycoprotein 41. Trimers of each gp120 and gp41 organize themselves via noncovalent interactions into heterotrimeric HIV-1 glycoprotein spike. The spikes are transported to plasma membrane. There, the Env glycoproteins are immediately recycled through endocytosis. Due to weak interaction between gp120 and gp41, Env spikes are packed with virus particle at approximately 10 spikes/virion. However, low expression of gp120 and gp41 on host cell surface has aided the HIV-1 to avoid host immune response. Three dimensional (3D) structure of gp120 and gp41 are illustrated in Figure 1.3 [13].

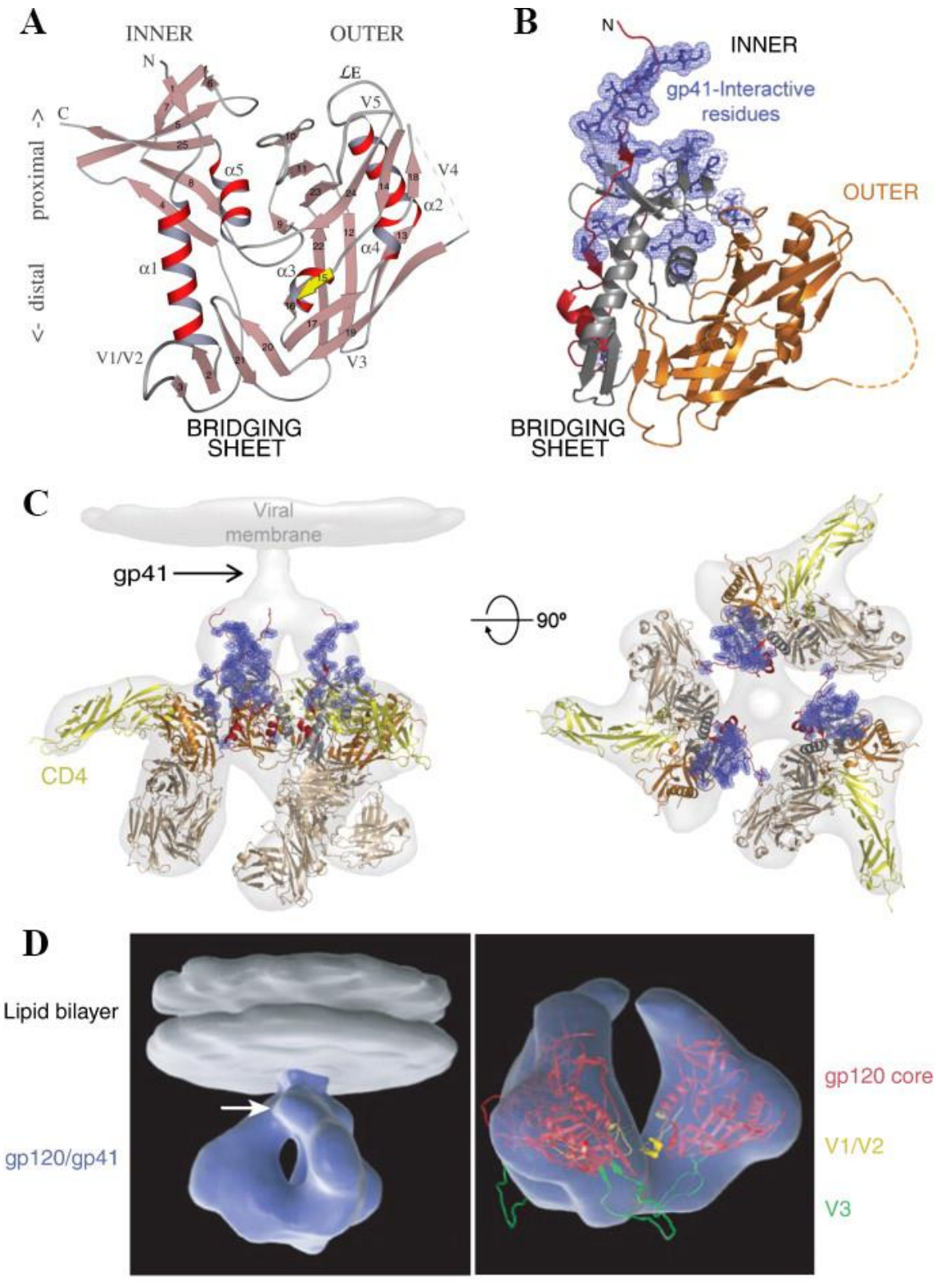

Figure 1.3 Three-dimensional structure of HIV-1 gp120 and gp41. Structure orientation in this figure illustrates HIV particle toward the top and target cell membrane toward the

bottom. Monomeric structure of gp120 shown in ribbon diagram contained 5 α-helices (α1 – α5) and 25 β-strands also with variable loops (V1 – V5). Four-β-strand bridging sheets are formed when gp120 binds to CD4 leading to separation of the structure into inner and outer domains (A) [14]. The same gp120 structure is shown ribbon diagram with the presence of gp41 interaction site (blue). The outer domain is in orange and inner domain in red (N-terminus) and gray (B). Stable gp120 was fixed with trimeric gp120 in interaction with CD4 (yellow) and Fab from neutralizing Ab 17b (brown). This complex structure was superimposed on to electron density obtained from cryoelectron tomography. Beside is the same complex structure from bottom view (C) [15]. CD4-bound conformation of HIV-1 Env in three dimension. The trimeric Env (blue) embedded in lipid bilayer of viral membrane (gray). Proposed location of gp41 was pointed with white arrow. The right figure showed the formation of V1/V2 (yellow) and V3 loop (green) superimposed on the density map (D) [16].

3) CD4 as the key for HIV entry

The binding site for HIV gp120 locates on D1 of CD4 molecule, as revealed by many mutational studies, soluble CD4 and neutralizing antibodies. Specific interaction was further mapped to the C′C″ ridge of CD4 molecule [17], declaring Phe43 and Arg59 being the crucial amino acids [10]. Only the aromatic ring of Phe43 can fit into the pocket in the heart of gp120 (Figure 1.4). The neighboring positively charged of lysine residues at position 29, 35 and 46 also associates with this interaction [10, 14].

As D1 plays an important role in the HIV binding, the others provide the flexibility, conformation and function. The D2 aids the HIV envelope-mediated fusion. The flexibility provided by the short linker between D2 and D3 (hinge region) and the transmembrane portion of D4 permits the bending of CD4 molecule and later the presentation of gp120 to the coreceptors at the cell surface [16, 18].

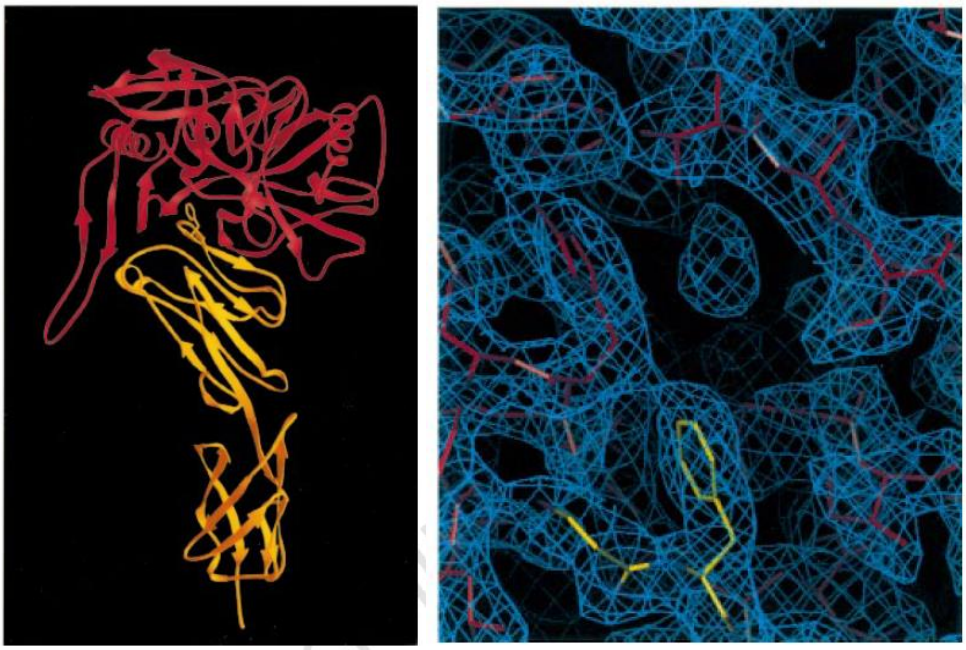

Figure 1.4 X-ray crystallography of HIV gp120 (red) and human CD4 (yellow). Being the crucial amino acid residue of human CD4, Phe43 inserts its aromatic side chain into the heart of HIV gp120 [14].

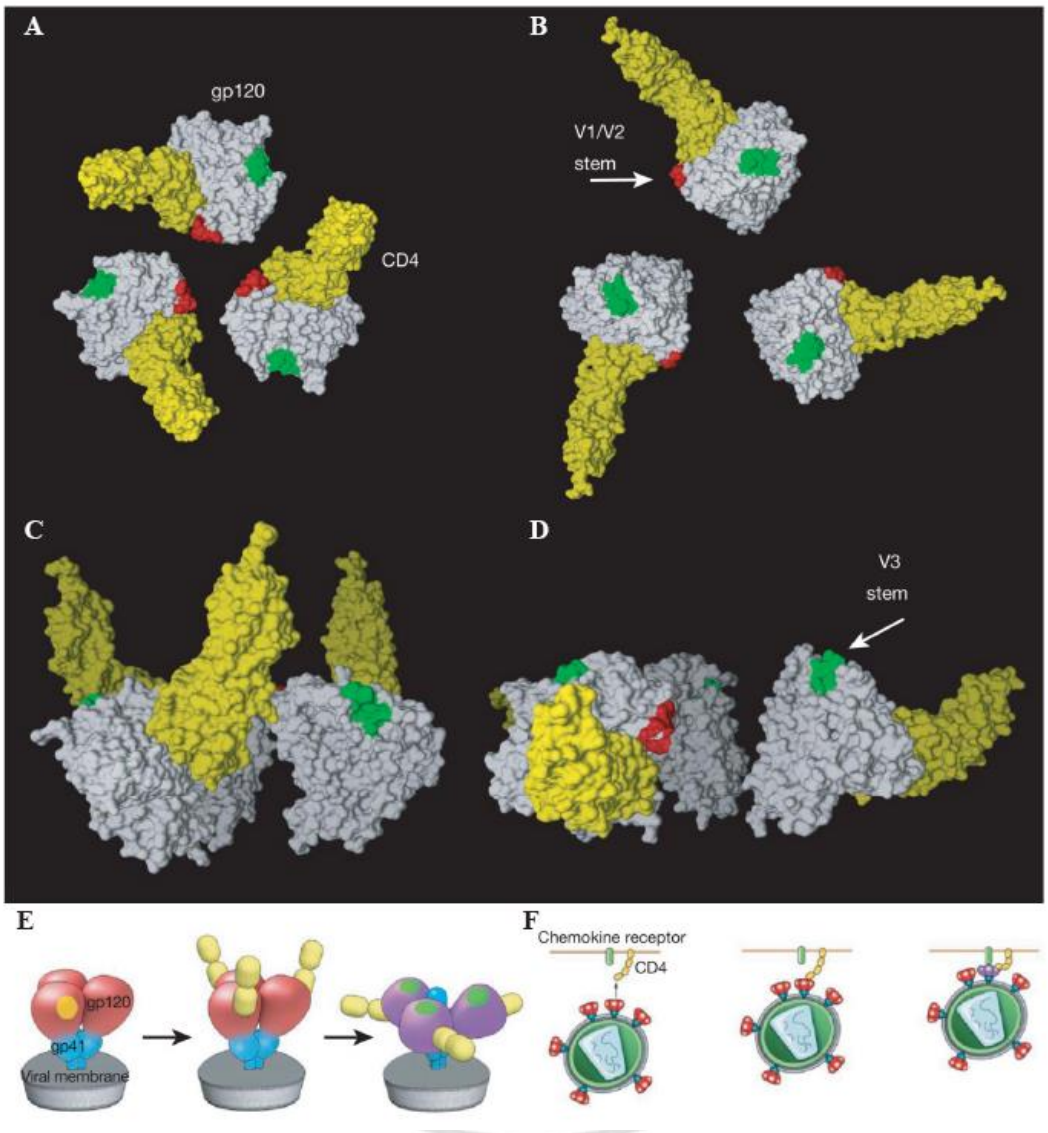

Figure 1.5 CD4-induced conformational change in HIV gp120. The gp120 core (white), CD4 (yellow), V1/V2 (red) and V3 stem (green) are shown above. For unliganded state, top view (A) and front view (Cc), the orientation of V3 stems are hidden. In CD4-bound state, the trimeric gp120 rearrange themselves and expose V3 stem, top view (B) and front view (D). The schemetic picture describes the same phenomenon as a-d. Yellow circle on gp120 is where CD4 binding site lies (E). Attachment of CD4, also bring the virus in closer proximity to host cell as well as induce the conformational change in gp120 leading to exposure of V3 loop to bind to host cell coreceptor (F) [16].

4) Targeting CD4 binding site on gp120 as a key to inhibit HIV-1 entry

Targeting the entry of HIV is a very appealing strategy to control the transmission of the virus. The most effective agent against HIV-1 entry is antibodies to CD4. Unfortunately, their uses are limited because of their immunosuppressive effects. The soluble CD4 (sCD4) has been proved to block the viral attachment *in vitro*, but not successful *in vivo*. Nevertheless, this soluble mimic of CD4 still gets the attention as a new therapeutic agent. The CD4-IgG2 contains the D1 and D2 domains of human CD4 instead of the Fv portion of both heavy and light chains. This tetravalent fusion protein has high avidity for cell-free HIV-1 virions. The CD4-IgG2 minimizes the possibility to mediate functions such as antibody-dependent enhancement of infection or transplacental transmission of HIV-1. Moreover, it also neutralizes various strains of HIV-1 *in vitro* [19] and *in vivo* in Hu-PBL-SCID mice [20].

Table 1.1 HIV entry inhibitors [21]

Class	Compound	Development status	Reference
Attachment inhibitor	PRO 2000	Phase III trial showed no efficacy	Rusconi et al. (1996)
Attachment inhibitor	Carraguard	Phase III trial showed no efficacy	Skoler-Karpoff et al. (2008)
Attachment inhibitor	Cellulose sulfate	Phase III trial showed no efficacy	Halpern et al. (2008) and Van Damme et al. (2008)
CD4-binding site inhibitor	BMS-378806	Preclinical development	Lin et al. (2003)
CD4 mimetic	Soluble CD4	Phase I study showed no efficacy	Daar et al. (1990) and Schacker et al. (1994)
CD4 downmodulators	CADA derivatives	Preclinical development	Vermeire et al. (2002, 2007, 2008)
CCR5 antagonist	Maraviroc	FDA approved	(2009)
CCR5 antibody	PRO 140	Phase II trials	Jacobson et al. (2008)
CXCR4 antagonist	Plerixafor	FDA approved for hematopoietic stem cell mobilization	Flomenberg et al. (2010)
Fusion inhibitor	Enfuvirtide	FDA approved	Wild et al. (1992, 1993)
Fusion inhibitor	D-peptides	Preclinical development	Welch et al. (2007)

The PRO 542 is a third generation of CD4-IgG2 which has the high potency of inhibiting the HIV-1 entry. Phase II clinical trial reveals that it can reduce the viral concentration in HIV-1 infected individuals at 25 mg/kg dose. During the period of 4–6 weeks of follow-up, the viral load reduces, suggesting that PRO 542 is an effective

prophylaxis against HIV-1 infection from sexual transmission and also cell-cell transmission [18, 22, 23]. There are more HIV entry inhibitors which is not limited to the inhibition of CD4 and gp120, as summarized in Table 1.1.

1.1.2 Ankyrin

1) Ankyrin repeat protein

Ankyrin repeat protein is protein architecture build up with varying number of structural motifs stack to form the repeat modules. Each repeat consists of 33 amino acid residues which form a helix-turn-helix conformation. The two α-helices are orchestrated in the antiparallel fashion. The β-sheet projects away from the helical pairs at an approximately 90° angle, following the formation of hairpin-like β-sheet with the prospective loops. The number of the repeats varies from 2 to 29 continuous repeats in a single protein, but usually found at 4 to 6 repeats. The assembly of N-terminal capping, repeat and C-terminal capping generates a mature ankyrin repeat protein [24-26].

Ankyrin is found naturally in many living organisms with the main function in mediating the protein-protein interaction. Ankyrins are found in many proteins ranging from transcriptional regulators, cytoskeletal organizers, developmental regulators, and toxin [26].

2) Designed Ankyrin Repeat Protein (DARPin)

DARPin originates from the natural ankyrin repeat proteins with the similar structural organization. The consensus sequence was analyzed from natural ankyrin repeat protein sequences in various databases in order to clarify a DARPin library module, which represents the fundamental block for DARPin libraries. This consensus sequence resembles to human sequence as many are acquired from the human genome (approximately 67% identity). As in the natural structure, these repeat modules are capped with specialized terminal repeats to protect the hydrophobic core of the internal repeats. The DARPin is defined as NxC, where N stands for the N-terminal capping repeat, C for the C-terminal capping repeat and x for the number of internal repeat modules (Figure 1.6). The generated DARPin from this library commonly range from two to four internal

repeat modules of 14 – 21 kDa. Most amino acids in the repeat are conserved; leaving six to be variable (Figure 1.7). The variable sequence can be any amino acid except cysteine, glycine or proline. This makes the theoretical diversity rise higher than 10^7 per repeat [27].

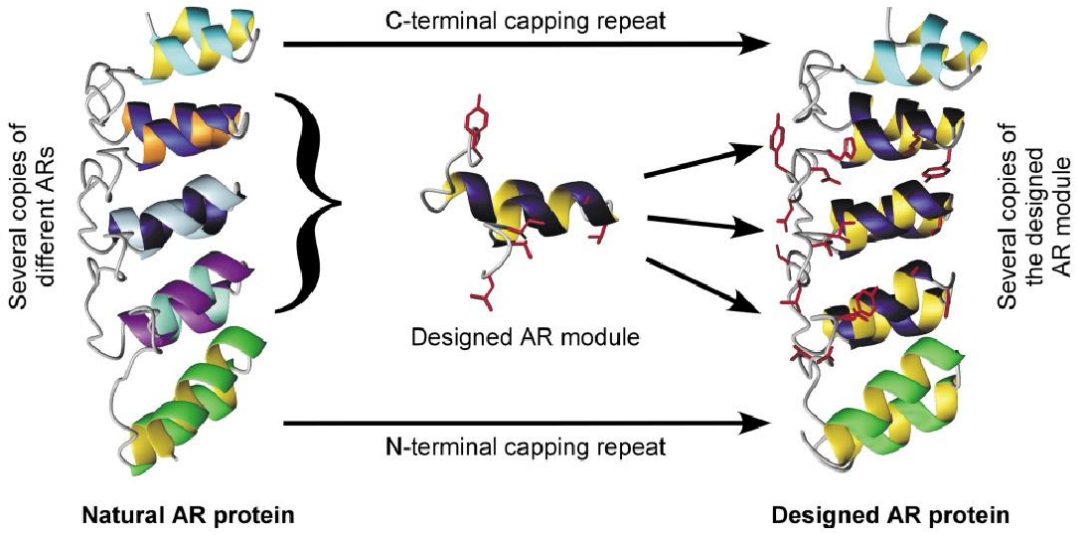

Figure 1.6 Structural modification of natural ankyrin protein to generate DARPin. With the similar structure, DARPin only differ in the repeat protein module, which can be designed with specificity to the desired target [27].

The consensus analysis also exposes the incredible properties of DARPins over natural ankyrin repeat proteins. These proteins can be expressed at high levels in the cytoplasm of *Escherichia coli* as soluble monomers with great stability either in shake flask or fermenter system. The thermodynamic stability of DARPins increases with the higher number of repeat modules as they are resistant to denaturation by boiling or guanidine hydrochloride [27].

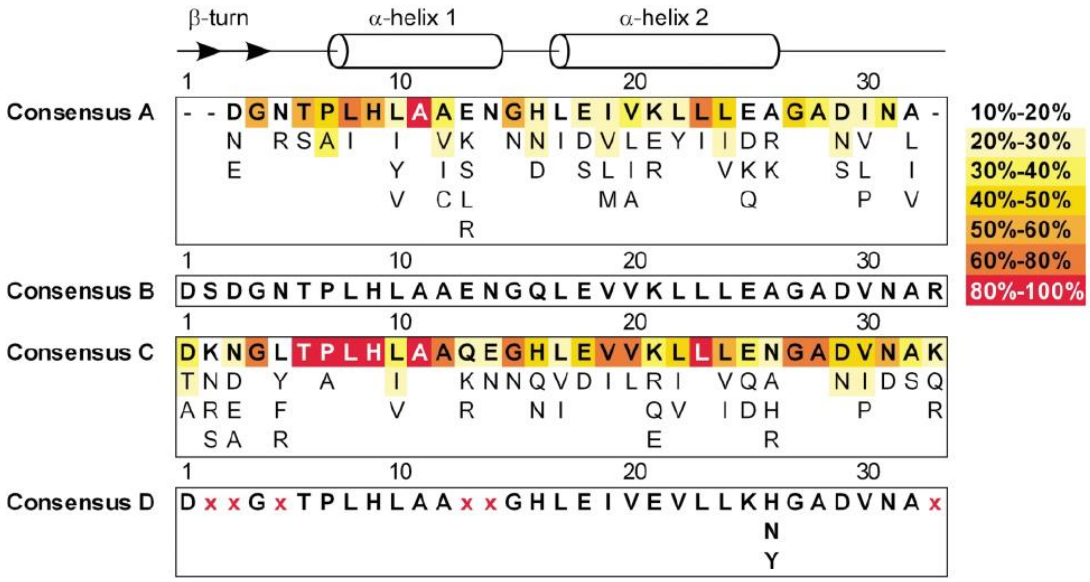

Figure 1.7 Consensus sequence of ankyrin repeat protein. Within the 33 amino acid residues of internal domain, most of them are conserved, leaving only 6 positions to be variable as shown in the letter x [27].

According to the ability to be selected against the desired target with high affinity and specificity at low production cost, DARPins have been raised as alternative to antibodies. The higher popularity in DARPin gives rise to the large diversity in therapeutics. The extraordinary characteristics in its stability and no tendency in aggregation narrow down the limitation in immunogenicity of DARPin-based drug. Current studies have focused the use of DARPin and monovalent and conjugated form. Even though DARPin drugs have not yet been approved in clinical trials, but the favorable preclinical results have been shown [4].

3) CD4-specific DARPin

From the study of Schweizer *et al.*, CD4-specific DARPin was selected from a DARPin library using ribosome display against CD4-IgG_2. This study showed that selected CD4-specific DARPins bound specifically to D1 and D2 of CD4 molecule. Moreover, it successfully inhibited the binding of various strains of HIV gp120 to CD4 *in vitro* without interference of the basic cellular functions and no downstream signaling events were observed. The cross-reactivity with rhesus macaque CD4 and inhibition of

SIV entry were seen in some clones of selected DARPins [5]. Despite the efficient binding of CD4-specific DARPin to CD4+ cells *in vivo*, the inhibition of SHIV in SHIV-infected rhesus macaques was not successful due to the rapid clearance from the plasma with the rate proportional to the dosage [6].

1.1.3 Computational study: molecular docking

In the new era of proteomics, the computational study has come to aid in analyzing the data from 3D structure and protein dynamics, which is largely introduced in the drug discovery industry. With the advancement of technology and the promising of high-efficiency and low cost of production, the computational-aided drug design gets the attention. The term docking in computational schemes refers to the attempt to find the best matching between two molecules, a receptor and a ligand based on a range of assumptions and approximations. The *in silico* homology modeling is introduced to generate protein structure. The *ab initio* prediction and molecular docking of these protein structures with interacting partners are two associated steps in computational proteomics. However, docking of known ligands improves the quality of crude homology models based on sequences alone.

Protein-protein docking has matured into a distinct computational study as it gather the knowledge and techniques from a broad spectrum of sciences including physics, chemistry, biology, mathematics, and computing. The most frequently used docking algorithms for energy-based scoring strategies are fast Fourier transform (FFT), geometric hashing and Monte Carlo (MC) techniques [28-31].

ZDOCK protocol performs rigid body docking of two or more proteins. This method performs a systematic search of a uniform sample of docked protein poses and uses an internal scoring algorithm to predict the optimal interactions. The following steps are performed: calculate docked protein poses, filter docked protein poses, rerank docked protein poses and finally cluster docked protein poses and calculate density. The data from ZDOCK poses are then refined using CHARMm energy minimization.

1) Three dimensional (3D) structure of protein

The study of protein structure provides vast information which can be applied to study the interaction of protein *in vivo*, *in vitro* and *in silico*. The Protein Data Bank (PDB) is mostly consisted of three-dimentional (3D) protein structure obtained from X-ray crystallography and later by nuclear magnetic resonance (NMR) spectroscopy and some other techniques.

1.1) X-ray crystallography

X-ray diffraction is an ideal method for large protein complexes over 100 kDa. This method requires soluble and rigid proteins in high concentrated crystal form. The crystallization of the sample is principally inducing the solution off the sample by mixing with a precipitant, buffer, pH and salt in some combination. This step is varied from sample to sample, and often is the rate limiting step of the whole process. To obtain the structure, the high quality X-ray generated from accelerating electrons at a synchrotron light source will be used to beam the crystallized sample. The diffraction data will later on generate the electron density map using the fourier transformation. The other bioinformatics of the complex will participate in the refining step and allow the researcher to determine the final minimized structure. Via X-ray crystallography, a high resolution of 3D protein structure can be accessed along with detailed atomic information. However, not all proteins can form crystal and diffract. Furthermore, crystallization might take weeks or months for some proteins.

1.2) Nuclear magnetic resonance (NMR) spectroscopy

NMR can provide both low and high-resolution binding site mapping. Instead of using the X-ray, NMR uses the magnetic field to excite the atomic nuclei and obtains the re-emitted electromagnetic radiation. NMR normally requires small protein molecule of 5 – 30 kDa with limited maximum weight of 60 kDa in solution. The method can analyze the atomic structure of flexible protein, which can be applied in kinetic reaction studies. The best NMR structure is a representative of 10 – 20 most energetically favored structure

in superimposes position. This method will allow the authors to determine the overall possible structures of the complexes and report the best fitted structure.

The spectra obtained from the 1 dimension NMR are too many and make it difficult to be differentiated and get useful information. The selected 1D data will move on to the second dimension or a space, defined by two frequency axes. The Nuclear Overhauser Effect (NOE) is the transfer of nuclear spin polarization from one nucleus to another through cross-relaxation. The atoms in close distance can give a NOE. The 2D NMR using NOE is called Nuclear Overhauser Effect spectroscopy (NOESY). This technique provides more precise refined structure of the molecule or the complexes. The most representative structure is selected based on the most energetically favorable structure.

1.3) Homology modeling

3D structure of a protein of known amino acid sequence can be obtain from homology modeling in case of unavailable 3D structure of that specific protein, either too large for NMR or unable to be crystallized for X-ray crystallography. Homology modeling predicts the 3D structure of a protein based on a former known experimentally established structure. By using the homologue protein as a template, the structure of a desired protein is constructed *in silico*. The first step to homology modeling is template identification and initial alignment. The identity percentage between plausible template and target over 40% is more favorable as seen in Figure 1.8.

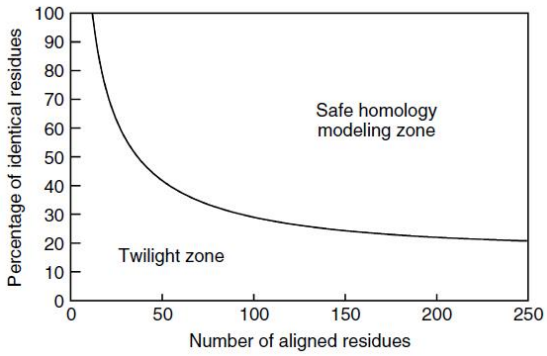

Figure 1.8 Zone of acceptable homology modeling. Two protein sequences are folded into the same structure if the length and percentage of sequence identity access into the safe homology modeling zone [32].

To identify the template, the amino acid sequence of target protein will be blast to database. A series of PDB structures were selected according to the alignments. In some cases, there is a difficulty to align two sequences and results in low identity in some region. Multiple sequence alignment has come to give additional information. Modeling can be achieve by generation of backbone. The omitted region can be fulfilled and refined by loop and side-chain modeling. To obtain the better prediction accuracy, the backbone of the model needs to be optimized. The energy minimization is applied to the overall protein structure. This step will remove the major errors from the structure, but in the same time introducing many minor errors to the structure too. The process continues until it reach the precise energy functions (forces fields). The last step is model validation. Every models contain errors, but they will be less problematic if they can be localized [32]. The overview of homology modeling is illustrated in Figure 1.9.

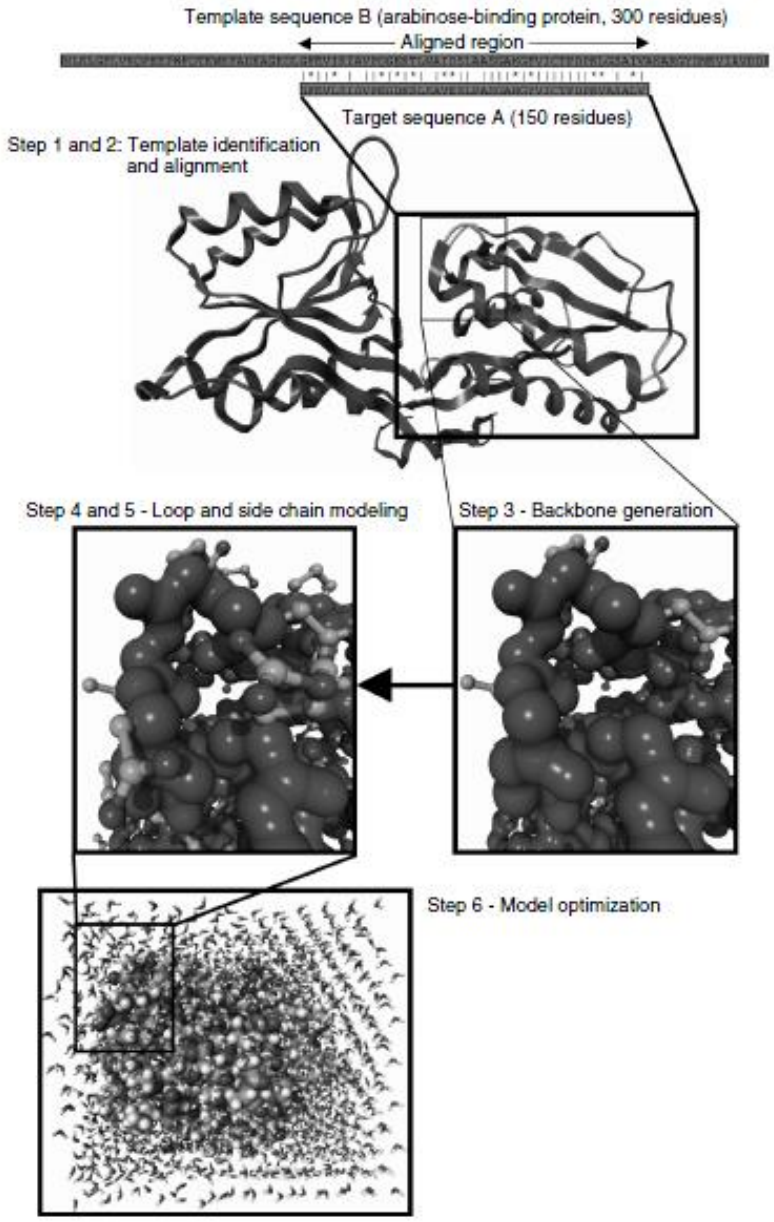

Figure 1.9 Basic concept of homology modeling [32]

2) Computational study in CD4-specific DARPin

The earlier study of Schweizer *et al.* suggested that the binding site of CD4-specific DARpin 23.2 lies within the D1 of CD4 [5]. Further specific binding site of these two molecules was investigated using computational study. Three dimensional (3D) structure of CD4-specific DARPin 23.2 was constructed *via* homology modeling using DARPin binding to human epidermal growth factor receptor 2 (HER2) (PBB code: 2JAB) as a template. The best model resembled 81.5% sequence identity, 89.5% sequence similarity and 56.01 verify score. The Ramachandran plot exhibited 111 residues (91.7%), including amino acids residues located in the binding site within core region and 7 residues (5.8%) in allowed region. The root mean square distance (RMSD) obtained from the homology modeling of CD4-specific DARPin 23.2 was less than 1 Å.

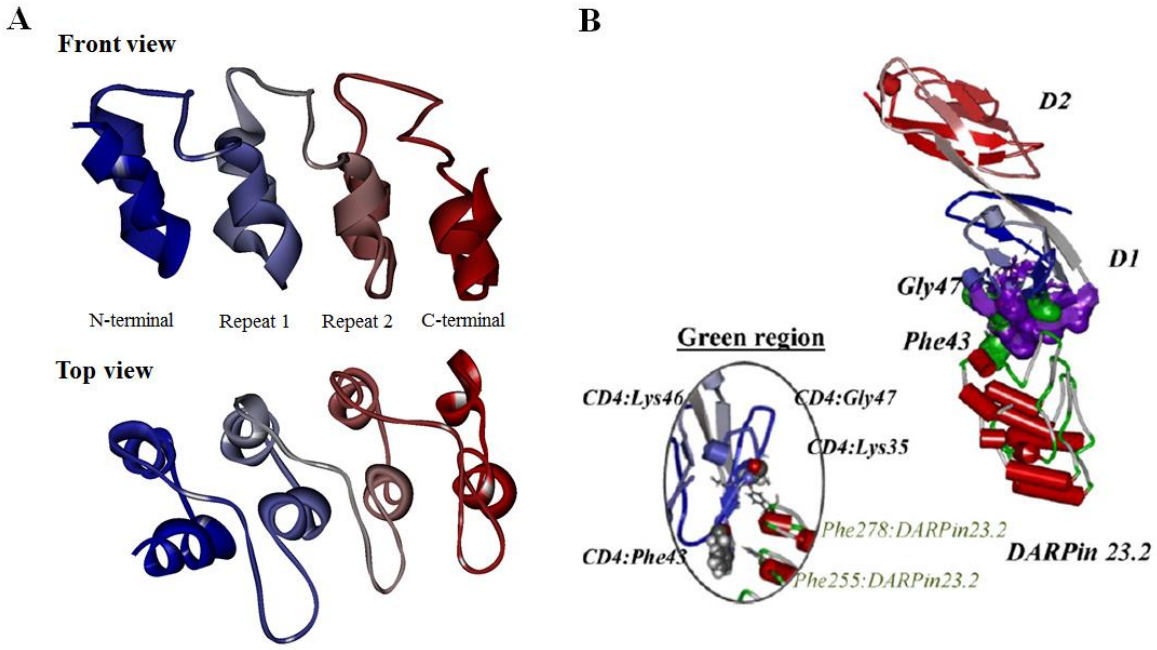

Figure 1.10 Structure and binding interaction of CD4-specific DARPin 23.2. Bimodular CD4-specific DARpin 23.2 structure, generated from homology modeling, was shown in solid ribbon model in front and top view (A). Binding interaction between CD4-specific DARPin 23.2 (stick model) and D1 of CD4 molecule (ribbon model) was evaluated from MD simulations [33].

The molecular docking using the x-ray crystallography of human CD4 (only D1 and D2) (PDB code: 3CD4) and the homology modeling of CD4-specific DARPin 23.2,

seen in Figure 1.10, revealed the binding sites which can be divided into clusters and differ in the binding energy. Free energy analysis using the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) and Molecular Mechanics Generalized Born Surface Area (MM-GBSA) methods were performed to clarify the binding interaction between the two proteins. Furthermore, the decomposition energy of the CD4-specific DARPin 23.2-CD4 complex was computed comparing with the ankyrin control to elucidate the decisive amino acid residues in the binding interface. The interaction was represented by the negative value. Three interaction regions were at position 31 – 36 (-219 kcal/mol), 65 – 69 (-153 kcal/mol), 97–102 (-88 kcal/mol) as calculated in Table 1.2. As a result, the three possible key residues of CD4-specific DARPin 23.2 were Leu34, Val67 and Phe100. The study also suggested the mutational study in order to improve the binding affinity of DARPin to CD4 molecule [33].

Table 1.2 MM-PBSA calculation (GBTOT value) in the three interaction regions [33]

Region 1		**Region 2**		**Region 3**	
Position	Gibbs free energy (kcal/mol)	Position	Gibbs free energy (kcal/mol)	Position	Gibbs free energy (kcal/mol)
31	-38.45	65	-81.53	97	-66.71
32	-45.70	66	-83.23	98	-41.96
33	-29.75	**67**	**9.32**	99	-1.56
34	**9.05**	68	2.79	**100**	**18.97**
35	2.19	69	-0.37	101	1.80
36	-116.60			102	0.52
Sum	**-219.26**	**Sum**	**-153.02**	**Sum**	**-88.94**

Table 1.3 Mutation of Phe100 into other amino acids to ameliorate the binding affinity of CD4-specific DARPin to CD4 molecule

Mutation	**Interaction Energy (kcal/mol)** Phe26- Lys29:CD4+ Lys35-Gln40:CD4+ Tyr45-Gly47:CD4/ Gln275-Lys280:ankyrin	**Van der Waals Energy(kcal/mol)**	**Electrostatic Energy (kcal/mol)**
Aromatic R groups			
Phe100 (WT)	**-17.39772**	**-17.16212**	**-0.23560**
Trp	61.63661	61.87538	-0.23877
Tyr	-12.20562	-11.99058	-0.21504
Polar, uncharged R groups			
Ser	-9.04374	-8.81731	-0.22642
Thr	-9.45569	-9.24469	-0.21100
Cys	-8.83569	-8.60907	-0.22662
Pro	-9.79523	-9.56852	-0.22671
Asn	-11.04878	-10.84985	-0.19893
Gln	-13.14088	-12.89312	-0.24776
Positively charged R groups			
Lys	-14.66072	-14.38972	-0.27099
Arg	**-20.59643**	**-20.47971**	**-0.11671**
His	-14.00152	-13.77864	-0.22288
Negatively charged R groups			
Asp	-11.45809	-11.11120	-0.34689
Glu	-13.52742	-13.25577	-0.27166
Nonpolar, aliphatic R groups			
Ala	-8.17878	-7.94039	-0.23840
Gly	-6.99591	-6.75667	-0.23924
Val	-10.55371	-10.31815	-0.23556
Leu	-13.70751	-13.46586	-0.24165
Met	-13.33690	-13.10247	-0.23443
Ile	-12.17384	-11.95135	-0.22249

Phe100 was proposed to be mutated in order to increase the binding affinity between DARPin and CD4 molecule as it displayed the highest Gibbs free energy (-88.94 kcal/mol) among the three regions. Table 1.3 showed the mutation of Phe100 into other amino acids (unpublished data). Considering the interaction energy and comparing with wild type Phe100, mutation to arginine would offer the stronger interaction. Nevertheless, mutation to glycine would abolish the binding.

In contrast, Wisitponchai *et al* studied the interaction between CD4-specific DARPin 23.2 and CD4 molecule *via* the key amino acid residues of CD4 [34]. By combining the computational algorithms and the mathematic calculation, the strategies to consider the key amino acid in the interaction were generated as demonstrate in Figure 1.11. The advantage of using this strategy is the more precise prediction. The DARPin-CD4 docking simulations were performed by a ZDOCK and RDOCK protocol in the Discovery studio (DS) 2.5.

The protein-protein interaction between homology model of CD4-specific DARPin 23.2 (using DARPin with PDB code 2JAB as a template) molecule (PDB code: 3CD4) was conducted by using ZDOCK protocol. Two thousand poses were screened and ranked. The top 20 poses which displayed the highest ZDOCK score were selected to be further analyzed by RDOCK protocol. Only 11 poses, further divided into 3 groups, showed the specific binding to D1 of CD4 molecule as reported by Schweizer *et al,* shown in Figure 1.12. The rational complexes were deeply evaluated and classified according to the criteria;

- Number of DARPin's amino acid per each CD4's amino acid
- Number of interactions per each CD4's amino acid
- Number of CD4's atom type per each CD4's amino acid
- Percent of CD4's atom type per each CD4's amino acid
- Interaction between atoms making hydrogen bond

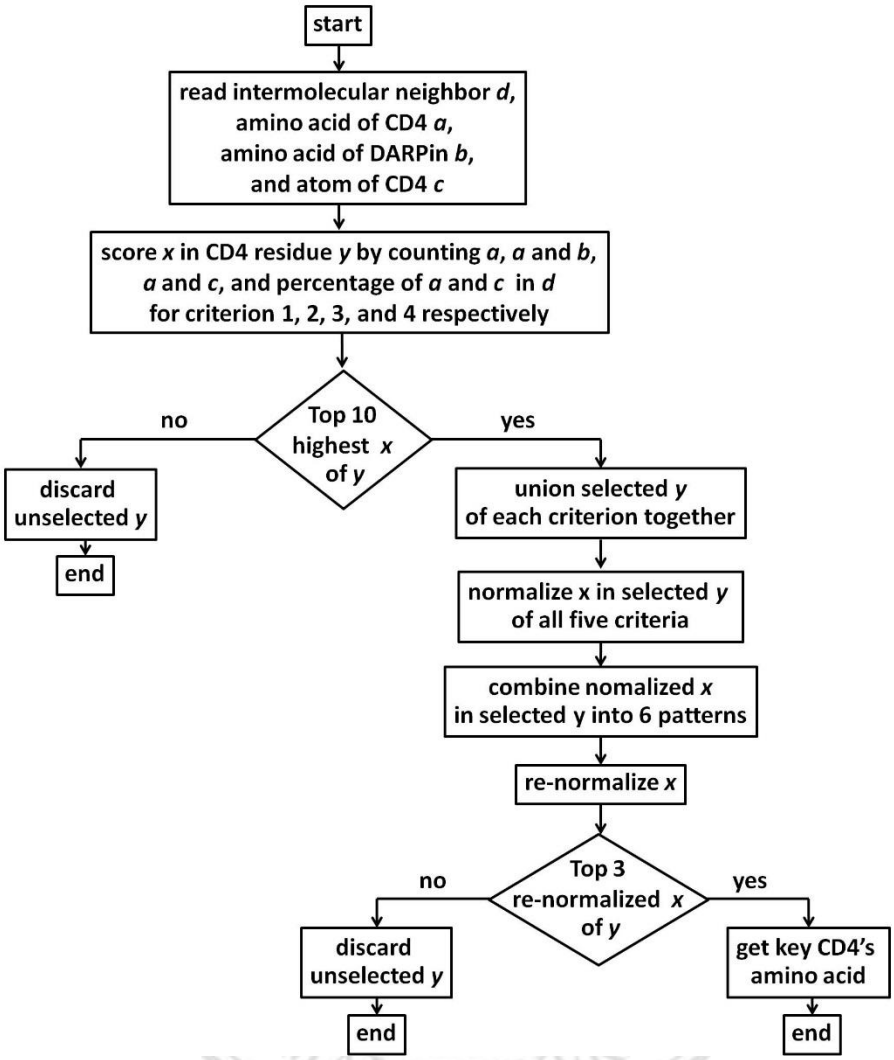

Figure 1.11 Generated algorithm used in the identification of crucial amino acids on CD4 molecule in the interaction with CD-specific DARPin 23.2

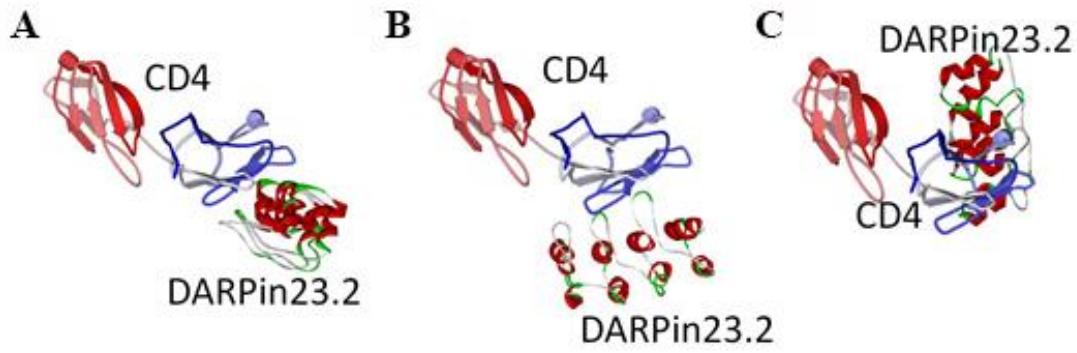

Figure 1.12 Representatives of binding interaction between CD4-specific DARPin 23.2 and D1 of CD4 molecule. From the candidates of 11 poses, the binding positions were classified into 3 groups, A, B, and C [34].

The data obtained from each criterion was evaluated and grouped into 6 patterns. The results were further merged and normalized into histogram. The results obtained from group 1 with the highest group members indicated that Lys35 on CD4 was named the 1st key amino acid residue and Gln40 and Gln25 being the 2nd and 3rd key amino acids respectively. The physical interaction from the study is demonstrated in Figure 1.13. The hot spot procured from this study was comparable to the hot spot obtained from HotPOINT and HSPred [34].

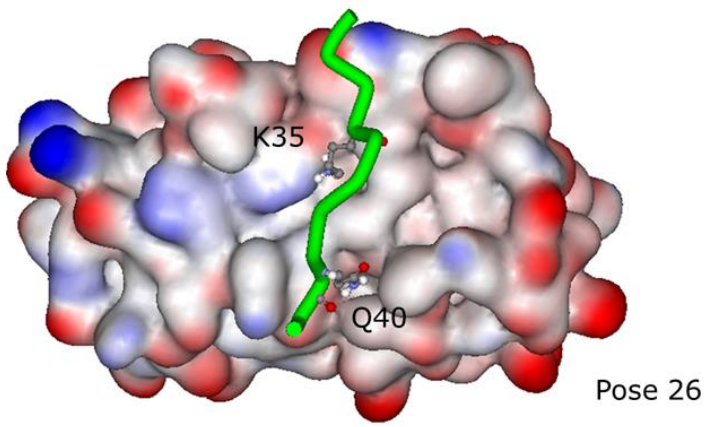

Figure 1.13 Key amino acids of CD4 molecule (green stick), Lys35 (K35) as the first rank and Gln40 (Q40) as the second rank, binding to CD4-specific DARPin 23.2. This data was obtained from pose 26 of group 1.

1.1.4 Adenovirus

1) Adenovirus structure and life cycle

Adenovirus belongs to the family *Adenoviridae* which are deeply classified into more than 50 different serotypes. They are common pathogens in human which exhibit a wide ran+ge of tropisms according to their various serotypes. Adenoviruses are non-envelope icosahedral viruses range from 80 – 110 nm in size. The capsid is composed of three major surface proteins or capsomers which are trimeric hexon protein, pentameric penton base protein and spike trimeric fiber protein. Adenoviral structure is shown in Figure 1.14.

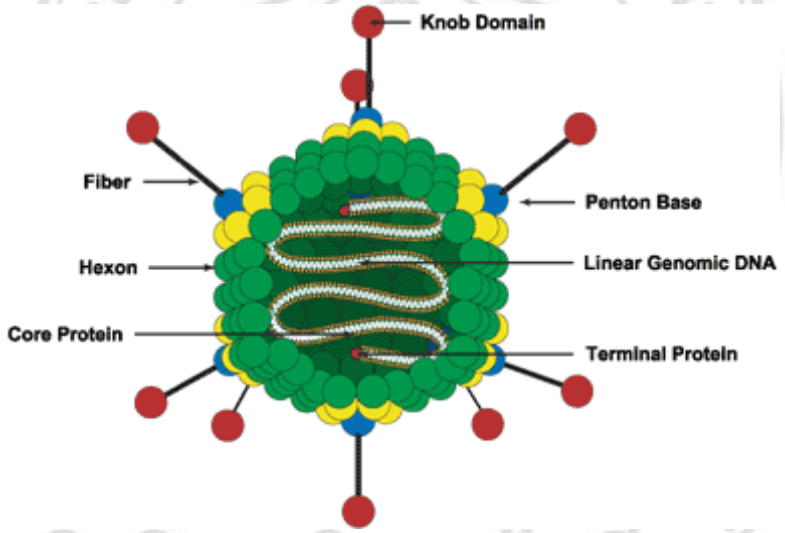

Figure 1.14 Components of adenoviral particle
(http://www.daviddarling.info/encyclopedia/A/adenovirus_infection.html)

Adenovirus enters the cells by binding of viral surface proteins to various cellular receptors such as coxsackie and adenovirus receptor (CAR) and $\alpha_v\beta_3$ integrin [35]. The internalization occurs after the formation of receptor-mediated endocytosis. The virus flee from the endosome to cytoplasm and then to nucleus *via* nuclear pore complex. The viral genome does not integrate into host chromosome but rather acts as an epichromosome. Inside the host nucleus, the virus utilizes host cell's transcriptional and translational

machinery to synthesis the viral protein. The virus production leads to cell death and liberation of progeny viruses [35-37].

2) Adenoviral vectors

Recombinant adenovirus is currently an effective tool for gene delivery and expression, also in vaccination and gene therapy. With the ability to infect most of dividing and non-dividing cells, the target cells for adenovirus are broad. Additionally, the high viral titers can be produced. The protein expression can be observed for both human and non-human proteins with proper folding and posttranslational modification. The most commonly used adenoviral vectors originate from human adenovirus species C (serotype 2 and 5). The viral genome is a 36-kb linear double-stranded DNA with inverted terminal repeats (ITR) at both ends which act as the origin of replication. The genome over 38 kb in size is incapable to be packed in viral particle. As many researches reveal the informative adenovirus life cycle and their genome, some early genes are removed to expand the capacity size of transgene.

The first generation adenoviral vector removed E1 gene from the viral genome and replaced with transgene. E1 gene is responsible for viral replication and producing infectious viral particle in target cells. Nevertheless, adenovirus propagation can be done in HEK-293 (293A) [38], 911 [39], PER.C6 [40] or N52.E6 [41] cells when using the E1-deleted adenoviral vector as it contains a stably integrated copy of the E1 gene that supplies the E1 proteins (E1a and E1b). ΔE1 adenovirus still preserved approximately 80% of viral genetic material. The natural adenovirus requires E3 gene *in vivo* to conquer host immunity which is not necessary for viral production. ΔE3 adenoviral vector increases space for transgene up to 8 kb. However, destruction of transduced cell can occur *in vivo* due to cytotoxic T lymphocyte (CTL) response from enduring adenoviral gene expressions [36, 42-44].

The second generation adenoviral vectors are ΔE1 ± ΔE3 and ΔE2/ΔE4. This vector's capacity of transgene is increased up to 10 kb. The adenovirus produced from this generation of vector cause low cytotoxic effects to host cells and is incapable of stimulation of immune response. To propagate the adenovirus, 911 or 239A cells expressing E4 are required.

The gutless or gutted adenovirus is currently the third generation adenovirus. It offers space for transgene up to 35 kb as most of the viral genome is depleted. Only 5' and 3' ITR and packaging signal (ψ) are preserved from wild type adenovirus. A helper adenovirus, usually ΔE1-adenoviral vector, is indispensable. More than 80 essential proteins for adenoviral genome encapsidation are generated from helper virus. To separate the adenovirus carrying transgene or helper-dependent adenovirus from helper virus, ultracentrifugation is needed as the two adenoviruses have distinct buoyant densities [44, 45]. The diagram comparing adenoviral genome modification is demonstrated in Figure 1.15.

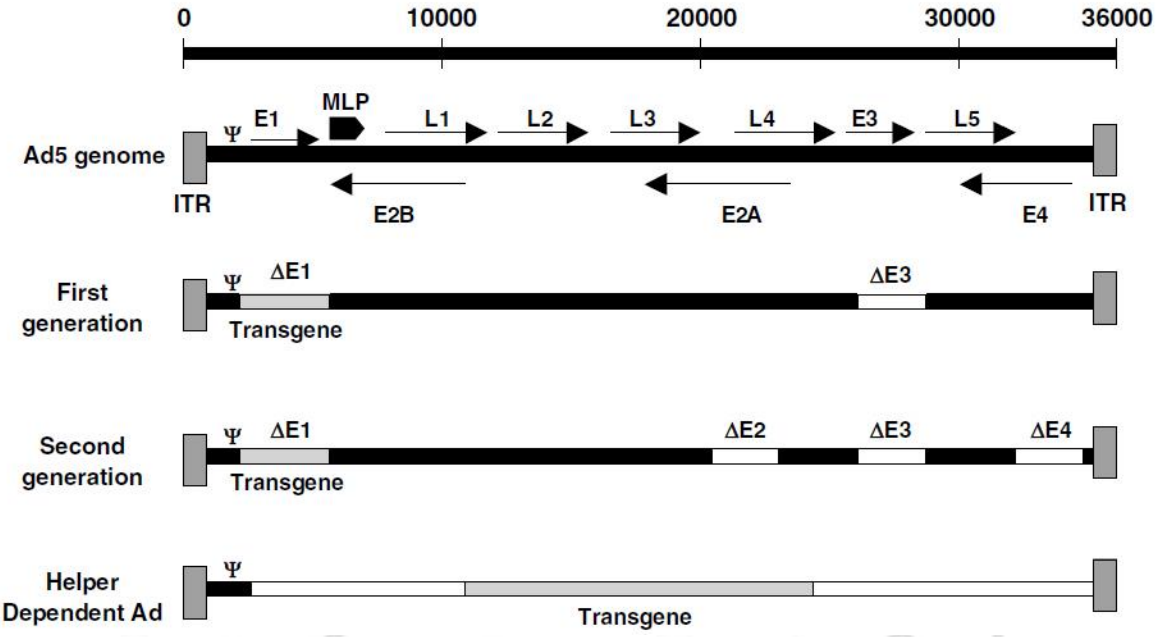

Figure 1.15 Adenoviral genome of 36 kb is consisted of 5' and 3' ITR, packaging signal (ψ), major late promoter (MLP), early transcript genes (E1 – E4) and late transcript genes (L1 – L5). E1 and E3 are deleted in the first generation adenoviral vector. Further deletion of E2 and E4 are found in the second generation. Only 5' and 3' ITR and packaging signal are left for helper-dependent adenovirus in order to increase transgene size [44].

1.1) pAdEasy system

The pAdEasy system was developed to generate a simple and efficient system for recombinant adenoviruses. The system consists of two vectors, backbone and shuttle vectors. The pAdEasy-1 backbone vector provides most of the adenoviral genome with

the deletion of E1 and E3 gene (ΔE1 and ΔE3). The pAdEasy-2 is based on second generation adenoviral vector with additional deletion of E4 gene. The pAdEasy vectors are supplied as supercoiled plasmid rather than as linear viral DNA. This is beneficial in simple backbone amplification and also provides the possibility of generating stable recombinant.

The shuttle vectors suitable for the pAdEasy system come in four versions. The viral sequences or the arms are found on shuttle vector which conduct the homologous recombination with pAdEasy-1. The pShuttle is the basic vector containing a set of multiple cloning sites. Its advantages are the great capacity for insertion of transgene and the flexibility to customize a promoter for transgene expression. The cytomegalovirus (CMV) promoter is added into pShuttle-CMV. The pAdTrack and pAdTrack-CMV vectors both carry the green fluorescent protein (GFP) in order to trace the expression of transgene.

The transgene is subcloned into an appropriate shuttle vector. The *Pme*I-linearized shuttle vector recombines with supercoiled backbone vector in *Escherichia coli* strain BJ5183 (*end*A1 *sbc*BC *rec*BC *gal*K *met thi*-1 *bio*T *hsd*R (*Str*^{r})) [46]. BJ5183 cells, not a *rec*A mutants but rather deficient in other enzymes, promotes the homologous recombination via the left and right arm homology. The isolated colonies grown on the medium containing kanamycin reflect the occurrence of homologous recombination. The recombinant vector is then digested with *Pac*I to generate linear viral genome before being transfected into 293A packaging cell line for virus construction. The viral production is commonly observed for comet-like foci 7 – 10 days after transfection. The schematic production of recombinant adenovirus is shown in Figure 1.16 [42, 47].

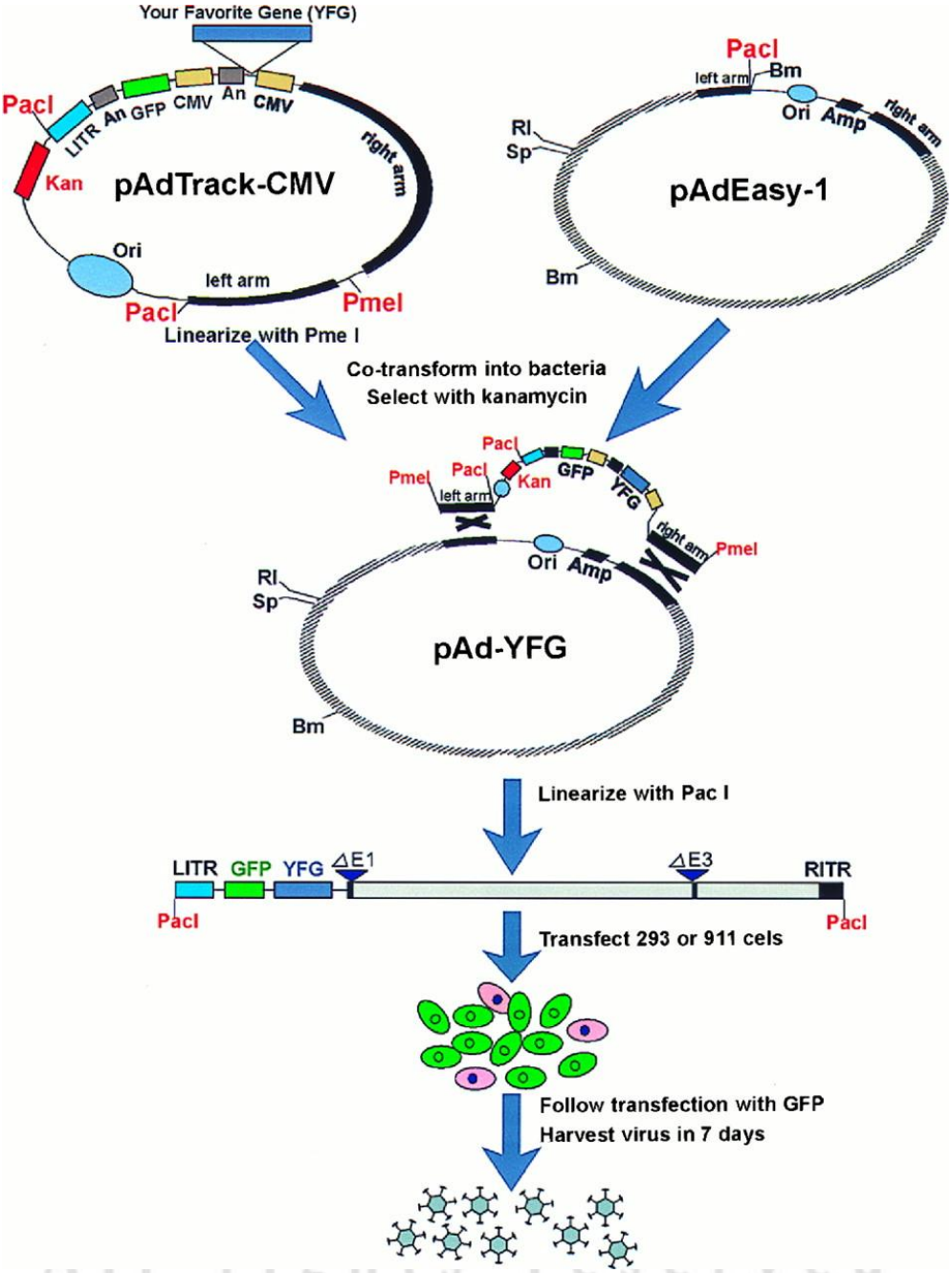

Figure 1.16 Schematic presentation of pAdEasy system to produce recombinant adenovirus [42].

1.2 Objectives

1.2.1 To elucidate the key amino acids in the binding of CD4 molecule and CD4-specific DARPin by combining the computational and mutational analysis

1.2.2 To determine the mechanism of CD4-specific DARPin used to inhibit HIV-1 entry

1.2.3 To improve the binding affinity of CD4-specific DARPin to CD4 molecule

1.2.4 To apply the adenoviral system to generate the expression of CD4 in mammalian cells