CHAPTER I

INTRODUCTION

1.1 Statement and significance of the problem

Human Immunodeficiency virus-1 (HIV-1) is the etiological agent of the acquired immunodeficiency syndrome (AIDS) (Mann *et al.*, 1988). Entry of the virus into a host cell is followed by the conversion of its RNA genome into double-stranded DNA by reverse transcriptase (RT) (Baltimore, 1970; Temin and Mizutani, 1970). Both terminal sequences of viral DNA are then joined non-covalently and enter into the nucleus (McGraw-Hill, 2006; Levy, 2007). The integration step of the double-stranded DNA molecule into the host genome is essential for viral replication (Katz and Skalka, 1994). The viral protein, integrase (IN), removes the terminal GT dinucleotide from each 3'-end long terminal repeat (LTR) of the viral genome to produce new 3'-hydroxyl ends (CA-3'OH) (Sherman and Fyfe, 1990; Bushman and Craigie, 1991), and the processed viral DNA is then inserted into the host chromosome (strand transfer) (Vink and Plasterk, 1993). The conserved CA and TG dinucleotides are crucial to maintain an integrative element at the ends of the viral DNA precursor (Whitcomb *et al.*, 1990). Mutation, insertion or deletion at the circle junction sequence in the virus results in the loss of integration into the host genome (Jurriaans *et al.*, 1992). Integration also stabilizes the viral DNA against degradation, and the unintegrated viral DNA in an acutely infected cell is degraded within hours to

days (Donehower and Varmus, 1984; Kim et al., 1989; Pauza, 1990; Barbosa et al., 1994; Pauza et al., 1994).

Highly active antiretroviral therapy (HAART) can dramatically reduce viral loads and increase CD4+ lymphocyte counts, which lead to improved health and longer lives in HIV-infected individuals (Cavert *et al.*, 1997; Palella *et al.*, 1998). Resistance to HAART has been observed, and some patients do not tolerate this therapy. Therefore, researchers continue to seek anti-HIV agents. IN inhibitors can be divided into two groups as follows: 3'-processing inhibitors and selective strand transfer inhibitors. The latter IN inhibitor (Raltegravir) has been approved for clinical use (Evering and Markowitz, 2008; Grant and Zolopa, 2008). However, resistant strains and toxicities have been observed (Charpentier *et al.*, 2008; Delelis *et al.*, 2009; Hicks *et al.*, 2009). In addition, withdrawal of anti-IN therapy results in the re-emergence of the virus because latent viral particles remain in reservoir cells in lymphoid organs and in the peripheral blood (Schrager and D'Souza, 1998). Therefore, alternatives, such as stem cell gene therapy, must be developed.

Intrabodies, which are mostly expressed as single-chain fragment variable (scFv), are antibodies that function intracellularly. Intrabodies that target IN (Levy-Mintz *et al.*, 1996), reverse transcriptase (RT) (Shaheen *et al.*, 1996), Vif (Goncalves *et al.*, 2002), Rev (Vercruysse *et al.*, 2010), gp120 (Marasco *et al.*, 1993), and Gag p17 (Tewari *et al.*, 2003) have been reported to inhibit the function of HIV-1 proteins. However, the reducing environment in the cytosol is an obstacle for the formation of disulfide bonds required for scFv function. Therefore, alternative scaffolds for conferring long-lasting protection against viral infection may circumvent the limitations of individual scFvs.

The Cys_2His_2 zinc finger protein (ZFP) is a ubiquitous DNA binding motifs in human, which can be designed to target any DNA sequence specifically. Regarding the excellent folding in cytoplasm and homing to nucleus, this molecule would be a candidate framework for long-last protection of HIV-1 infection. Certain reports have been manifested the interferemce of HIV-1 replication at the transcriptional level (Reynolds et al., 2003); (Segal et al., 2004). Hence, our strategy relies on tracking the non-covalent 2-LTR circle junctions using the designed ZFP during the manipulation of HIV-DNA by IN. Therefore, the benefit of targeting the integrase recognition sequence at 3'-end terminal part of HIV-1 LTR should introduce a prospective for the future gene therapy.

1.2 Literature review

1.2.1 HIV-1 and AIDS

HIV-1 remains one of the most serious health problems in the world. It is a pathogen which causes AIDS. Since the first case of AIDS has been reported, many people had already lost their lives by that time to the disease and the number of seroconverters was running with the rate of 7,400 now a day (UNAIDS/WHO, 2009). According to UNAIDS and WHO global estimates of the number of people living with HIV/AIDS as of the end of 2009 were 33.3 million adults and children, whereas people newly infected with HIV were 2.6 million (UNAIDS/WHO, 2010). In the present, the epidemic still extremely threat in many countries such as the nations of southern Africa, while in some developed countries can control excellently a number of new infections.

1.2.2 HIV-1 structure

HIV was classified in family *Retroviridae*, and subfamily *lentivirinae*. The diameters of virus particle are approximately 110 nm with an outer envelope protein. The actual arrangement is usually in the form of icosahedrons. The outer envelop consists of lipid bilayers membrane derived from the infected cell and glycoprotein and some cellular proteins. The spikes of cell surface consist of a transmembrane glycoprotein, gp41, as the stem and a surface glycoprotein, gp120 (Chan et al., 1997; Weissenhorn et al., 1997). Deep inside of the viral envelop is a layer of matrix protein (MA), or p17 within which is a conical-shape protein core containing the two identical single stranded plus sense RNA. A matrix shell containing approximately 2000 copies of the matrix protein (MA, p17) lines the inner surface of the viral membrane and a conical capsid core particle comprising about 2000 copies of the capsid protein (CA, p24) is located in the center of the virus (Frankel and Young, 1998; Turner and Summers, 1999). The genomic RNA tensely associated with the nucleocapsid protein (NC), or p7 and the three enzymes: protease (PRO), or p11, reverse transcriptase (RT), or p66/p51, and integrase (IN), or p31. Other 6 viral proteins, such as Vif, Vpr, Nef, Rev, Tat, and Vpu are also found in the virion, which play an essential role in efficient and normal HIV-1 pathogenesis to infect a cell. The structure of HIV-1 is shown in **Figure 1.1**.

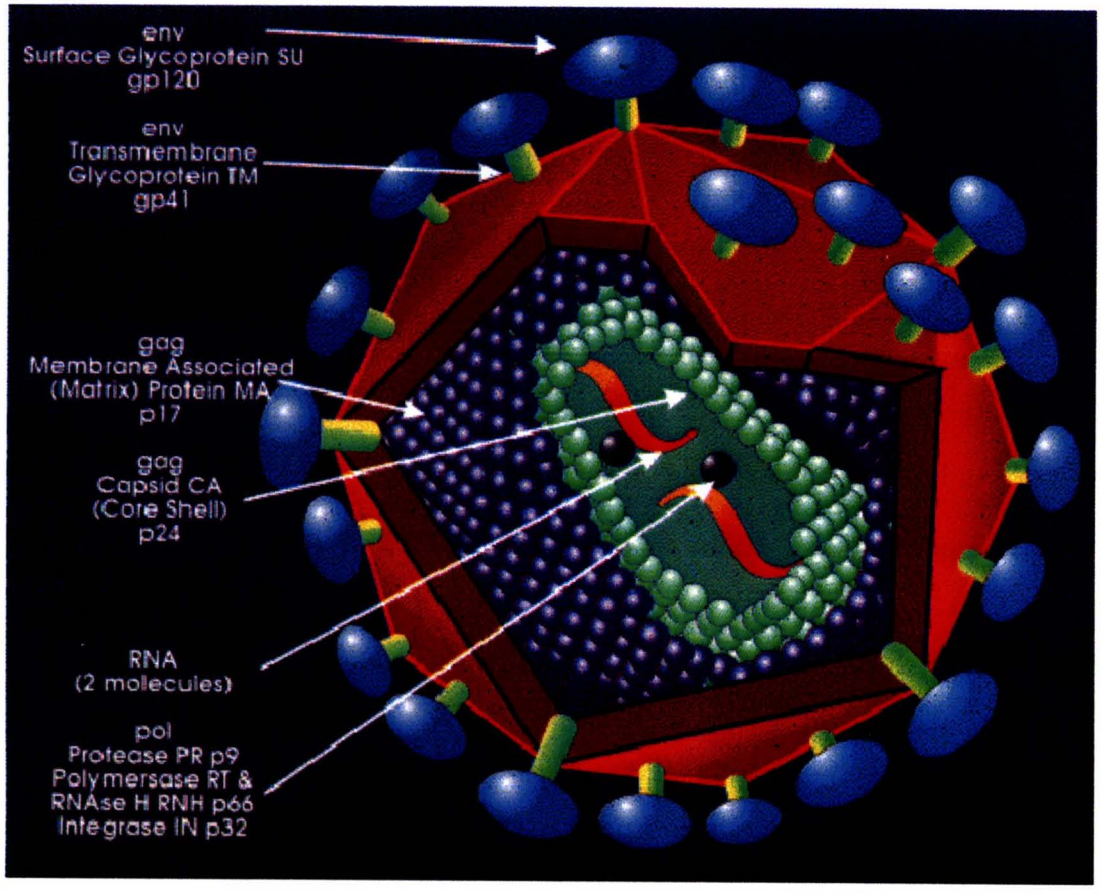

Figure 1.1 The structure of human immunodeficiency virus (HIV).

(Source: www.wellesley.edu/.../ Chem101/hiv/HIV-1.html)

1.2.3 HIV-1 genomic structure

HIV genome is 9.7 kilobases (kb) in length (Hahn et al., 1984). There are three structural genes, *gag*, *pol*, and *env* located from 5' to 3' (Muesing et al., 1985; Ratner et al., 1985). In addition to the three major genes, HIV viruses also contain several small genes including LTR, *vif, vpr, tat, rev, vpu,* and *nef* that control expression of the viral genome. The genomic structure of HIV-1 is shown in **Figure 1.2**.

LTR Long terminal repeat, this part genome is identical repeated non-coding sequences designated U3, R, and U5. The functions of these units are important for transcription initiation and polyadenylation.

GAG Group specific antigens, the genomic region encoding the capsid proteins. The precursor is p55 myristylated protein, which is processed to p17 (MAtrix), p24 (CApsid), p7 (NucleoCapsid), and p6 proteins, by the viral protease.

POL The genomic region encoding the viral enzymes protease, reverse transcriptase, and integrase, which are produced as a Gag-pol precursor polyprotein.

ENV The responsibility of this gene is to produce outer envelope. Precursor is *gp160* which is processed to the external glycoprotein *gp120* and the transmembrane glycoprotein *gp41*.

VIF Viral infectivity factor, it is likely that the Vif protein increases the efficiency of cell to cell and person to person spread in HIV infection.

VPR Viral protein R is a protein, which promotes nuclear localization of preintegration complex, inhibits cell division, arrests infected cells at G2/M.

TAT Transactivator of transcription. Tat protein is necessary for viral expression and replication.

REV The second necessary regulatory factor for HIV expression.

VPU Viral protein U is a protein, which promotes extracellular release of viral particles, and degrades CD4 in the endoplasmic reticulum (ER).

NEF Nef downregulates CD4, the primary viral receptor, and major histocompatibility complex (MHC) class I molecules.

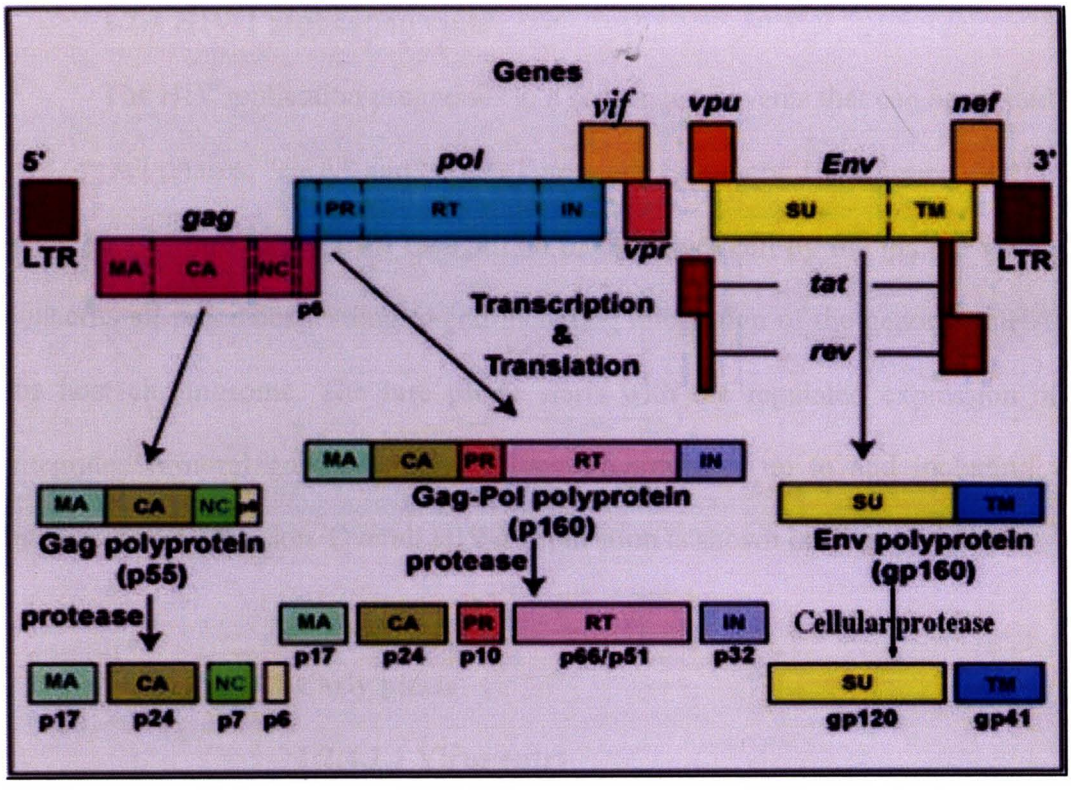

Figure 1.2 HIV-1 genomic structures.

(Source: adapted from www.cat.cc.md.us)

1.2.4 HIV-1 replication cycle

The HIV replication progresses in a sequence of events that can be sprited into two overall phases: "early" and "late" (Turner and Summers, 1999; Freed, 2001). The **early phase** initiates with the recognition of the target cell by the mature virion and concerns all processes leading to and including integration of the genomic DNA into the host chromosome. The **late phase** starts with the regulated expression of the integrated proviral genome, and involves all processes up to and including virus budding and maturation. Overall HIV-1 replication is shown in **(Figure 1.3).**

1.2.4.1 Early phase

1.2.4.1.1 Virus entry

To establish infection, the first step of viral entry occurs by interaction between the surface envelope glycoprotein gp120 and the cellular receptor, CD4, which is present on the surface of a subset of T-lymphocytes and monocytes (Clapham and McKnight, 2002). In addition on CD4 receptor, there must also be a chemokine coreceptor. The coreceptors differ for different cell types. CXCR4 is on T-lymphocyte, whereas CCR5 is on monocytes (Clapham and McKnight, 2002). Binding of the surface subunit gp120 to CD4 and a coreceptor on cell surface triggers conformational changes in the envelope complex, leading to the insertion of the hydrophobic N-terminal fusion peptide (FP) of gp41 into the target cell membrane (Eckert and Kim, 2001).

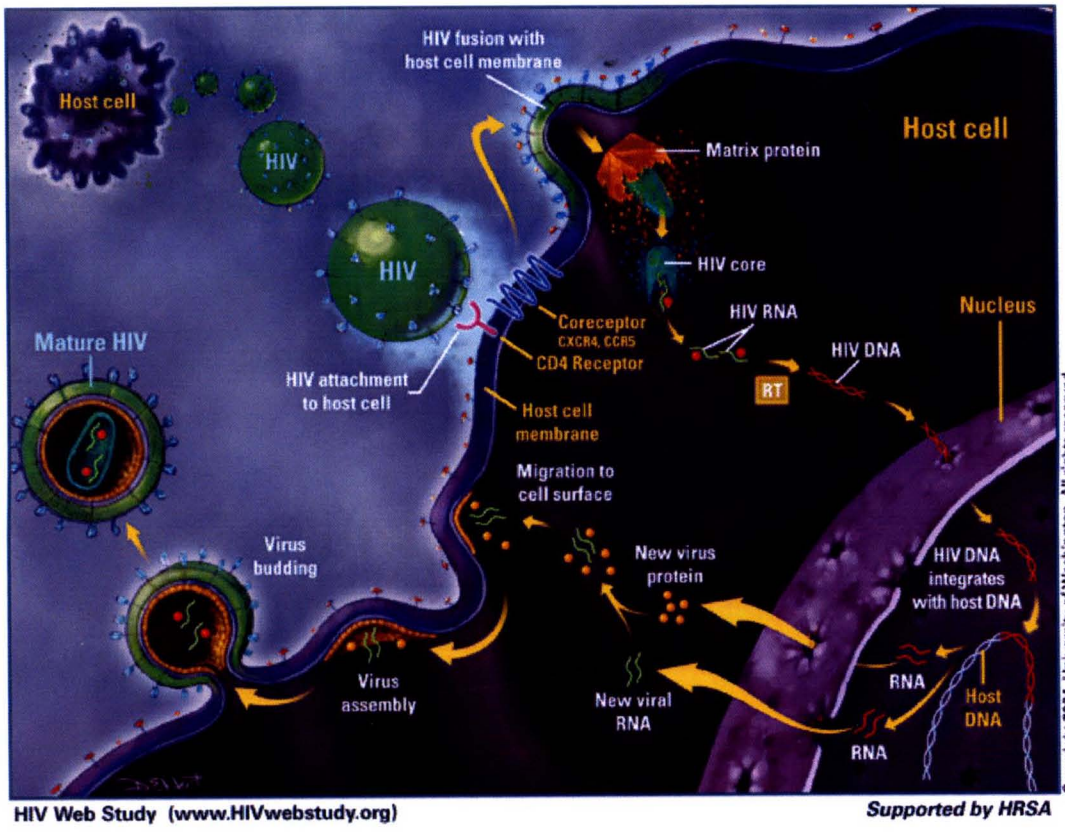

Figure 1.3 HIV-1 replication cycle. (Source: www.HIVwebstudy.org) The HIV-1 replication cycle can be divided into two main phases: early and late. The overall steps of the early phase include viral fusion, reverse transcription, nuclear trafficking, and integration. The late phase begins with host-cell mediated proviral gene transcription and translation consequenced by virion assembly, virion budding, and finally, viral maturation into infectious progeny virion capable of another round of infection with a new target immune cell.

1.2.4.1.2 Reverse transcription

After virus entry to the host cell, the viral capsid is disrupted and the viral RNA is released in to the cytoplasm. During reverse transcription, the two RNA molecules in the virion are converted to a linear double-stranded DNA (Gotte et al., 1999). The whole process of reverse transcription is shown in **Figure 1.4** as a series of steps below:

1. A specific cellular tRNA acting as a primer and binds to a complementary viral RNA at the primer binding site (PBS) and then synthesis a short stretch of DNA (the minus-strand stong-stop DNA)
2. One activity of the reverse transcriptase enzyme so-called RNAse H, which degrades the 5' end of the RNA by removing the U5 and R region.
3. The minus-strand stong-stop DNA is transfered from 5' end to the 3' end of viral RNA of the same, or a second viral RNA molecule.
4. Then, first strand of complementary DNA (cDNA) is extended.
5. The majority of viral RNA is degraded by RNAse H except a polypurine tract (PPT).
6. PPT serves as a primer for synthesis a short stretch of plus-strand viral DNA (plus –strand-strong stop DNA).
7. There is then another 'jump' where the PBS from the second strand hybridizes with the complementary PBS on the first strand.
8. Both strands are extended further and can be processed and ligated into the host genome by the integrase.

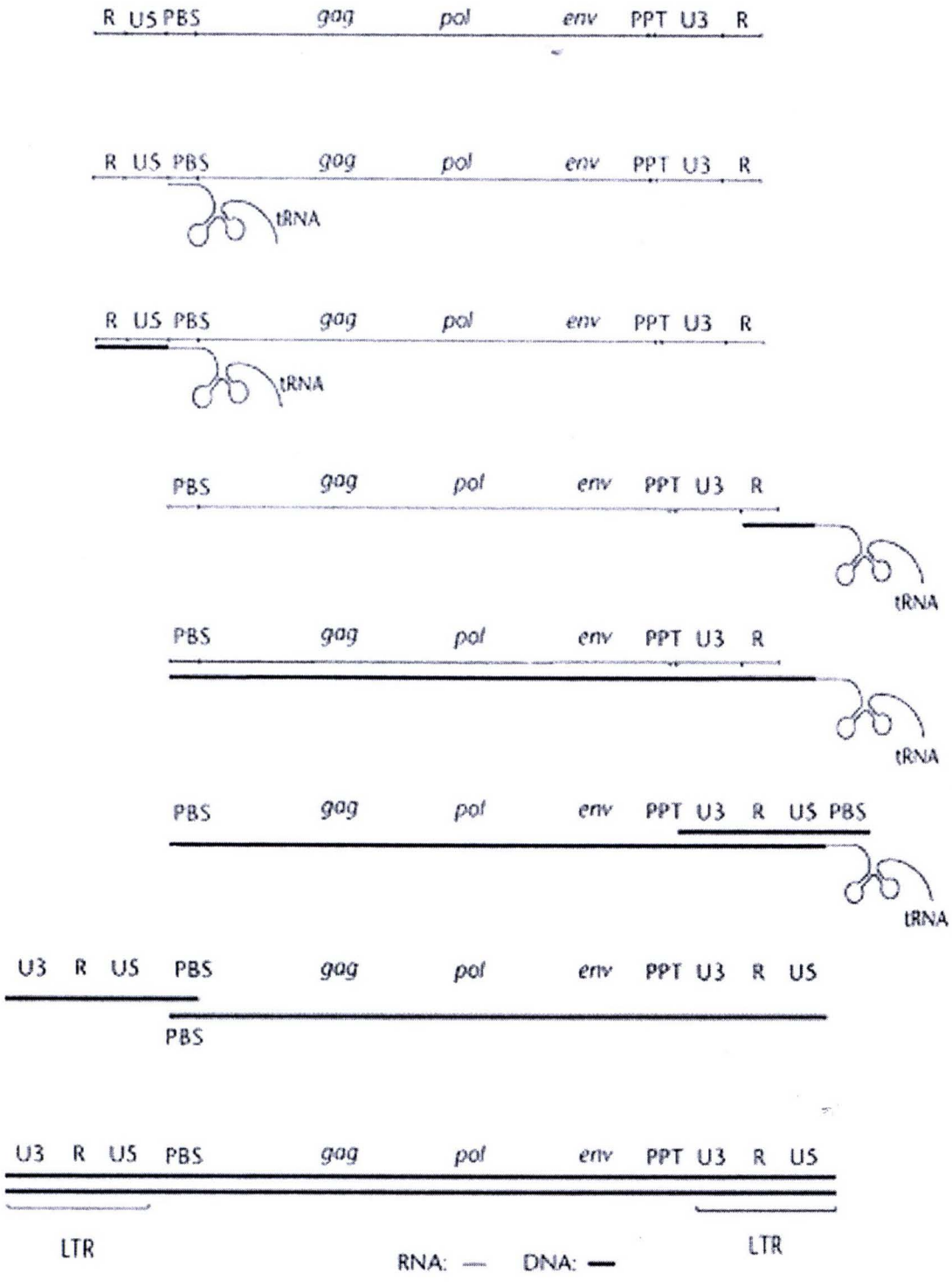

Figure 1.4 Reverse transcription of HIV-1.

(Source: http://www.els.net/WileyCDA/ElsArticle/refId-a0000430.html)

1.2.4.1.3 Nuclear transport and integration

After the step of reverse transcription, the newly synthesized HIV-1 cDNA is transported to the nucleus as part of a preintegration complex (PIC) (Freed, 2001). This complex includes viral DNA and many viral proteins including MA, RT, IN, Vpr and NC (Van Maele and Debyser, 2005). Functional cellular proteins have also been identified functional in PICs (Van Maele and Debyser, 2005). Vpr protein is helped to enhance the HIV-1 preintegration complex transport to the nucleus (Le Rouzic and Benichou, 2005). Moreover, a nuclear localization signal (NLS) from Vpr and MA provide to actively enter the nucleus. The proviral DNA is then randomly integrated in the cellular host genome by the viral enzyme integrase (IN) (Esposito and Craigie, 1999). The integration is a crucial step for HIV-1 replication and it is involved in the focus for integration interference in this research. This HIV-1 enzyme will be described in more detail in a later section.

1.2.4.2 Late phase

After integration of viral DNA into the host chromosome, the integrated provirus functions as the source for the viral RNAs synthesis that encode the viral structural as well as regulatory/accessory proteins used for the production of progeny virus particles.

1.2.4.2.1 HIV transcription

The integrated form of provirus flanked by the tandem LTRs is organized as a eukaryotic transcriptional unit (Turner and Summers, 1999). The 5' LTR contains a strong enhancer/promoter and the 3' LTR contains an efficient polyadenylation site.

The actual transcript begins at R (+1). The primary transcript is a 9.2 kb RNA and is used in three ways. Firstly after capping at 5'end and polyadenylation at 3'end, it is utilized as genomic RNA for the production of a new virion. Secondly, it is also used as the mRNA for viral Gag and Pol proteins. Thirdly, it is spliced to generate subgenomic mRNAs for all other viral proteins.

1.2.4.2.2 HIV translation, assembly and budding

After transportation of mRNA from nucleus to the cytoplasm, 2 kb HIV-1 mRNAs encoding regulatory Tat, Rev and Nef and the 4 kb mRNAs encoding viral accessory proteins Vpr, Vpu, and Vif are translated into individual mature proteins. On the other hand, 4 kb and 9 kb mRNAs encoding for envelop and Gag-Pol proteins, respectively are translated first into precursor polyprotein which were later cleaved during or after viral assembly. Full length of HIV-1 RNA can be used as a mRNA for protein translation or be used either as a genomic RNA for packaging into the virion.

Translation initiation starts with the Gag open reading frame. Gag precursor polyprotein (pr55) is a major translation product. It is modified by adding myristate at its amino terminus, which plays a key role for direction the pr55 Gag precursor to the plasma membrane of infected cell. Pr55 Gag precursor can be processed by viral protease enzyme into the matrix protein (MA) p17, the capsid (CA) p24, the nucler capsid (NC) p7 and the p6 protein which bind to the Vpr protein and also incorporated in the virion. The Gag-Pol precursor pr160 is the minor product from 9 kb viral mRNA. The *Pol* gene products derived from the pr160 which include the viral protease (PR) p11, reverse transcriptase (RT) p66, integrase (IN).

The Env precursor polyprotein (gp160) is made in the endoplasmic reticulum (ER) using the spliced *env* mRNA gene as the template (Cimarelli and Darlix, 2002; Ganser-Pornillos et al., 2008). Env is posttranslationally modified in the ER and Golgi apparatus and is cleaved to produce the surface glycoproteins (gp120) and the transmembrane (gp41), and these are transported to the plasma membrane of the host cell (Freed, 2001) follow with the HIV genomic RNA to pack into a new virion and then bud from the host cell (Demirov and Freed, 2004).

1.2.4.2.3 Assembly, budding and maturation

Assembly is processed by the Gag polyprotein. This process is started from the Gag targets to the site of assembly and interacts with lipid bilayer membrane. At N-terminus of HIV-MA has myristoyl modification. The myristoyl group can bind phosphoinositol (4,5) bisphosphage (PI(4,5)P2, and a phosphoinosotide into the lipid bilayer. 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 following by the encapsidation of genomic RNA and the incorporation of Env into virus particles. The budding process and the virus detaches from the plasma membrane (Simons and Gerl, 2010). The maturation of progeny virion involves final proteolytic cleavage of Gag and Gag-Pol precursor polyproteins to yield mature virion protein. The mature virion is then ready to infect the next cell and initiate a new infection cycle.

1.2.5 Antiretroviral drugs

Current antiretroviral drugs therapy for AIDS involves the use of a multidrug cocktail referred to as highly active antiretroviral therapy (HAART) (De Clercq, 2001). As of December 2008, approximately 4 million people in low- and middle-income countries were receiving antiretroviral therapy—a 10-fold increase over five years (UNAIDS/WHO, 2009). HAART is effective to dramatic reduction of virus loads and higher CD4+ lymphocyte count, which improves health and longer lives in HIV infected individuals (Cavert et al., 1997; Palella et al., 1998). Now a day, there are more than 26 approved antiretroviral drugs but not all are licensed or available in every country. There are five groups which were classified to act at different stages of the HIV life-cycle. Nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs) are available in most countries. Fusion/entry inhibitors and integrase inhibitors are usually using in resource-rich countries. Protease inhibitors are not a common for starting treatment in resource-limited areas due to a high cost per pills, and the particular side effects caused by protease inhibitors. Current licensed antiretroviral drugs are shown in **Table 1.1** (Flexner, 2007).

Table 1.1 Current licensed antiretroviral drugs (adapted from Flexner, 2007).

Name	Trade name	Company	Launched
Nucleoside or nucleotide reverse-transcriptase inhibitors			
Zidovudine	Retrovir	GlaxoSmithKline	1987
Didanosine	Videx	Bristol–Myers Squibb	1991
Zalcitabine	HIVID	Roche	1992
Stavudine	Zerit	Bristol–Myers Squibb	1995
Lamivudine	Epivir	GlaxoSmithKline, Shire Pharmaceuticals	1998
Abacavir	Ziagen	GlaxoSmithKline	1999
Tenofovir disoproxil fumarate	Viread	Gilead	2001
Emtricitabine	Emtriva	Gilead	2003
Non-nucleoside reverse-transcriptase inhibitors			
Nevirapine	Viramune	Boehringer Ingelheim	1996
Efavirenz	Sustiva, Stocrin	Bristol–Myers Squibb, Merck	1998
Delavirdine	Rescriptor	Pharmacia & Upjohn, Agouron, Pfizer	1999
Protease inhibitors			
Saquinavir	Invirase	Hoffmann-La Roche	1995
Indinavir	Crixivan	Merck	1996
Ritonavir	Norvir	Abbott, GlaxoSmithKline	1996
Nelfinavir	Viracept	Agouron, Pfizer	1997
Amprenavir	Agenerase, Prozei	Vertex	1999
Lopinavir + ritonavir	Kaletra, Aluvia	Abbott	2000
Atazanavir	Reyataz, Zrivada	Bristol–Myers Squibb, Novartis	2003
Fosamprenavir	Lexiva, Telzir	Vertex, GlaxoSmithKline	2003
Tipranavir	Aptivus	Boehringer Ingelheim	2005
Darunavir	Prezista	Tibotec	2006
Entry inhibitors			
Enfuvirtide	Fuzeon	Trimeris, Roche	2003
Maraviroc	Celsentri, Selzentry	Pfizer	2007
Integrase inhibitors			
Raltegravir	Raltegravir	Merck	2007

1.2.6 Gene therapy for HIV-1

The efficacy of HAART has brought to an impressive reduction of viral count to undetectable levels in many HIV-1 infected individuals, which resulted to decreased motality rate in these populations. However, obvious eradication of the infection has failed due to the insistence of latent HIV-1 in resting memory CD4+ T-cells. Moreover, the emergence of multi-drug-resistant viral strains in patients exposed to HAART regimens and the rapid spread of infections in developing countries attititudinize a big challenge for AIDS treatment (Chen et al., 2007; Gupta and Pillay, 2007).

The ultimate goal of gene therapy is to develop the strategy for treatment a disease by delivery of an appropriate therapeutic gene: (a) to hematopoietic progenitor cells (HSC) for long-term protection of their differentiated progeny from HIV-1; (b) to the circulating CD4+ cells which are the HIV-1 target, to make them resistant to HIV-1 infection or inhibit HIV-1 replication; (c) to immunize against HIV-1 antigens; and (d) to inhibit HIV-1 in the orther organ target sites (e.g., central nervous system) (Strayer et al., 2005).

A number of gene therapy strategies have been approached to inhibit HIV-1 replication by interfering with the functions of HIV-1 RNAs or proteins. These approaches may be classified into two groups: protein-based and RNA-based. The interfering RNA strategies involve antisense RNAs (Cohli et al., 1994; Lu et al., 2004), ribozymes (Ramezani et al., 2002), RNA aptamers and decoys (Bohjanen et al., 1996), small interfering RNAs (siRNAs) (Lee et al., 2002), and short hairpin RNAs (shRNAs) (Lee et al., 2002). The interfering protein strategies involve *trans-*

dominant negative mutants (Liem et al., 1993), single-chain antibodies (intrabodies) to inhibit the function of intracellular target proteins in specific cellular compartments such as IN (Levy-Mintz et al., 1996), reverse transcriptase (RT) (Shaheen et al., 1996), Rev (Vercruysse et al., 2010), gp120 (Marasco et al., 1993), Gag p17 (Tewari et al., 2003).

Although gene therapy holds promise, a number of disadvantages must be overcome before it could be respected for an alternative therapy in HIV-infected individuals. The therapeutic advantage would rely on the option and mechanism of the transgene to inhibit HIV-1 replication, and on the interference step within the life cycle of virus. Inhibition of viral replication at the earliest step is believed to be perfect for becoming infection and would permit to sustain normal immune functions. Therefore, advance studies are needed before these strategies may be assessed in clinical trials.

1.2.7 HIV-1 integration as an important step for intracellular immunization

Each stage of the HIV-1 replication cycle can be used as a strategy for inhibition. Since integration is an essential step to insert HIV-1 DNA into the host genome, it is the most interesting point for intervention. Many investigators are interested to target IN enzyme activity by using antiviral drug design (De Clercq, 2001; Deprez et al., 2004; Zahm et al., 2008), but resistant strains and toxicities have been observed (Charpentier *et al.*, 2008; Delelis *et al.*, 2009; Hicks *et al.*, 2009). In gene therapy strategy, some studies have been reported by using scFv (Levy-Mintz et

al., 1996; BouHamdan et al., 1999; Strayer et al., 2002), but the main obstacle for using scFv is a disulfide bond formation for its folding. However, the integration process is involved not only with IN enzyme, but also with IN substrate. IN substrate or IN recognition sequence is crucial to maintain an integrative element at the ends of the viral DNA precursor (Whitcomb *et al.*, 1990). Therefore, IN recognition sequence makes an attractive target to inhibit HIV replication. In this thesis, intracellular immunization against IN recognition sequence at 2-LTR circle junctions by using zinc finger protein (ZFP) might be an alternative approach to search a new molecule for HIV integration blockcade. Structure and function of IN will be portrayed in the following section.

.

1.2.7.1. The structure of HIV-1 IN

HIV-1 IN is a member of three HIV enzymes and classified in a group of superfamily of polynucleotidyl transferases. The molecular weight is 31-Kda. It is produced from the C-terminal part after the processing of $Pr^{160\ gag\text{-}pol}$. IN plays an important role for the integration of viral DNA into the host genome. This enzyme cleaves and /or link nucleic acids via phosphodiester bond. IN comprises three domains **(Figure 1.5)**. The N-terminal domain (residues 1-50) which contains a His_2Cys_2 motif, is highly conserved among integrases of all retroviruses, and binds one equivalent of Zn^{2+} (Cai et al., 1997). Although the N-terminal domain is involved for 3' processing and strand transfer, its mechanisms are not clear. The core domain (residues 50-212) contains the catalytic triad motif (D, D, 35E) coordinated with Mg^{2+} metal ion, which is characterized of many polynucleotidyl transferases, which required for 3' processing and strand transfer (Drelich et al., 1992; Bushman et al.,

1993; Engelman et al., 1993; Vink and Plasterk, 1993; Mazumder et al., 1994; Jonsson et al., 1996). The C-terminal domain (213-288) binds DNA nonspecifically, deletion of this domain abolishes 3' processing and strand transfer activities (Drelich et al., 1992; Bushman et al., 1993; Engelman et al., 1993; Vink and Plasterk, 1993; Jonsson et al., 1996). The intact HIV-1 IN appears to function as tetramer (Jenkins et al., 1996), each of dimers work on one LTR end (Faure et al., 2005).

1.2.7.2. The function of HIV-1 IN

A double-stranded DNA was made from viral RNA by RT, then both end of HIV-1 LTR are joined non-covalently and enters into the nucleus (McGraw-Hill, 2006; Levy, 2007). A schematic representation of non-covalent 2-LTR circle junctions is shown in **Figure 1.6**. The steps for HIV-1 integration are followed here and shown in **Figure 1.7** (Chiu and Davies, 2004). Firstly, IN assemble the end of the viral DNA and bind the integrase recognition sequence to removes the terminal GT dinucleotide from each of the viral DNA ends, leaving a recessed CA with a free 3' hydroxyl group and an overhanging 5'-AC on the complementary strand, the so-called 3' processing. Then, reaction is strand transfer by ligating both 3' ends of HIV-1 DNA into the 5' ends of the cleaved target DNA. Finally, the intermediate product of viral DNA-host DNA is gapped and repaired by host cell DNA repair enzymes (Brown, 1990; Yoder and Bushman, 2000).

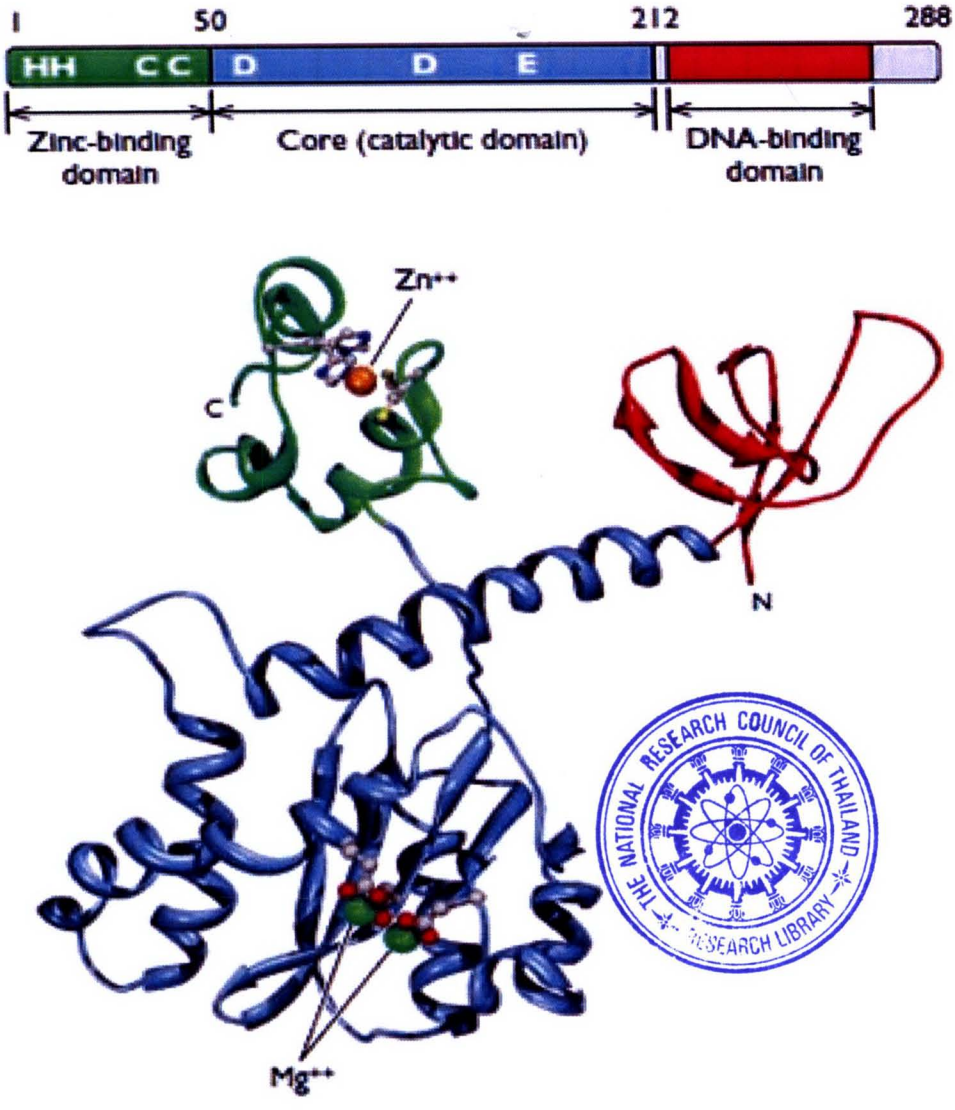

Figure 1.5 The structure of HIV-1 integrase (Skalka, 2011)**.** The HIV-1 IN structure in individual domains is numbered sequentially from the N-terminus. Green color gives structure of the N-terminus for His_2Cys_2 domain that binds Zn^{2+} and is involved in protein multimerization. Catalytic core domain is shown in blue containing the catalytic triad motif (D, D, 35E). C-terminus is in red and this domain has an SH3-like fold and consists of two antiparrallel beta sheets and binds DNA nonspecifically.

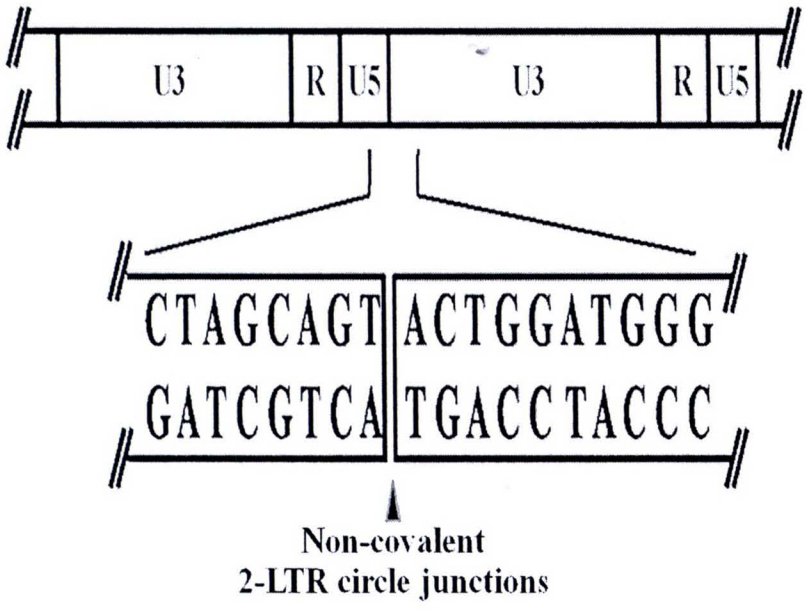

Figure 1.6 Schematic representations of non-covalent 2-LTR circle junctions of HIV-1. Predicted organizations of U3, R, and U5 are shown above.

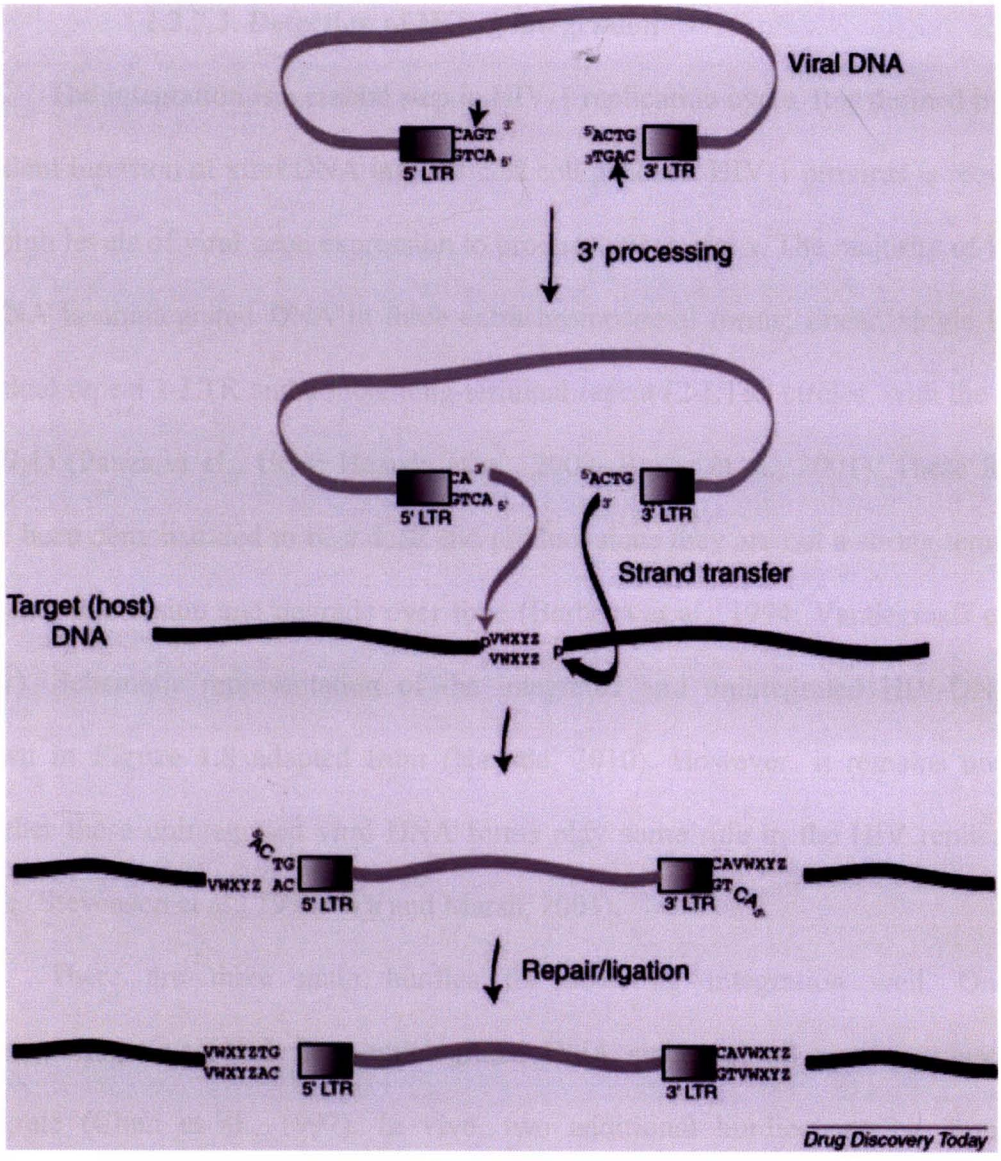

Figure 1.7 3' processing and strand transfer activities of IN. In 3' processing, IN enzyme removes a GT dinucleotide at both 3' end of LTR producing a new 3' hydroxyl ends (CA-OH). Then, the processed CA-OH viral DNA ends are ligated to the 5' O-phosphate ends of host target DNA, so call strand transfer. After integration, host enzymes repair and ligate the viral/cellular DNA junction (Adamson and Freed, 2008).

1.2.7.3. Detection of HIV-1 integration

The integration is a crucial step in HIV-1 replication cycle. It is defined by the covalent insertion of viral DNA into the host cell genome. HIV-1 provirus is required for high levels of viral gene expression to produce new virions. The majority of HIV-1 DNA is unintegrated DNA in three extrachromosomal forms; linear, single long-terminal repeat 1-LTR and double long-terminal repeat (2-LTR) circles with the ratio (20:9:1) (Pauza et al., 1994; Hazuda et al., 2000; Butler et al., 2001). These forms have been demonstrated to be a dead end product since they are not a strong template for gene expression and degrade over time (Barbosa et al., 1994; Vandegraaff et al., 2001). Schematic representation of the integrated and unintegrated HIV-DNA is shown in **Figure 1.8** adapted from (Hazuda, 2010). However, it remains unclear whether these unintegrated viral DNA forms play some role in the HIV replication cycle (Stevenson et al., 1990; Wu and Marsh, 2001).

There are three main hurdles for detecting integration well. One is distinguishing integrated from unintegrated DNA since not all reverse transcripts integrate (Chun et al., 1997). In vivo, two additional hurdles can be found in enhancing sensitivity since the level of integration is low (Chun et al., 1997) and detecting all variants of the integrated population since it is well known that HIV has a high mutation rate (Coffin, 1995).

A number of methods to detect HIV-1 integration level have been reported such as gel separation, inverse PCR, and linker ligation PCR. In the first method, integration in HIV-infected individuals is processed by first using gel separation methods to isolate genomic DNA from episomal DNA (Lehrman et al., 2005; Koelsch et al., 2008) and then detecting the amount of HIV DNA within the genomic DNA by

routine quantitative real-time PCR. However, the gel separation method is too laborious for large numbers of samples and it is unclear how effectively this method separates episomal DNA from genomic DNA. For inverse PCR (Chun et al., 1995), it involves: digesting genomic DNA, ligating the DNA into a circle, and performing outer and nested PCR **(Figure 1.9).** The last method, linker ligation PCR (Vandegraaff et al., 2001), involves digestion of chromosomal DNA with the restriction enzyme *Nla*III generating fragments with cohesive termini for ligation with oligonucleotide linker. Then, the linker and HIV DNA are served as a template for PCR amplification **(Figure 1.9).** However, these 2 assays are labor-intensive and not appropriate for screening in a large scale samples.

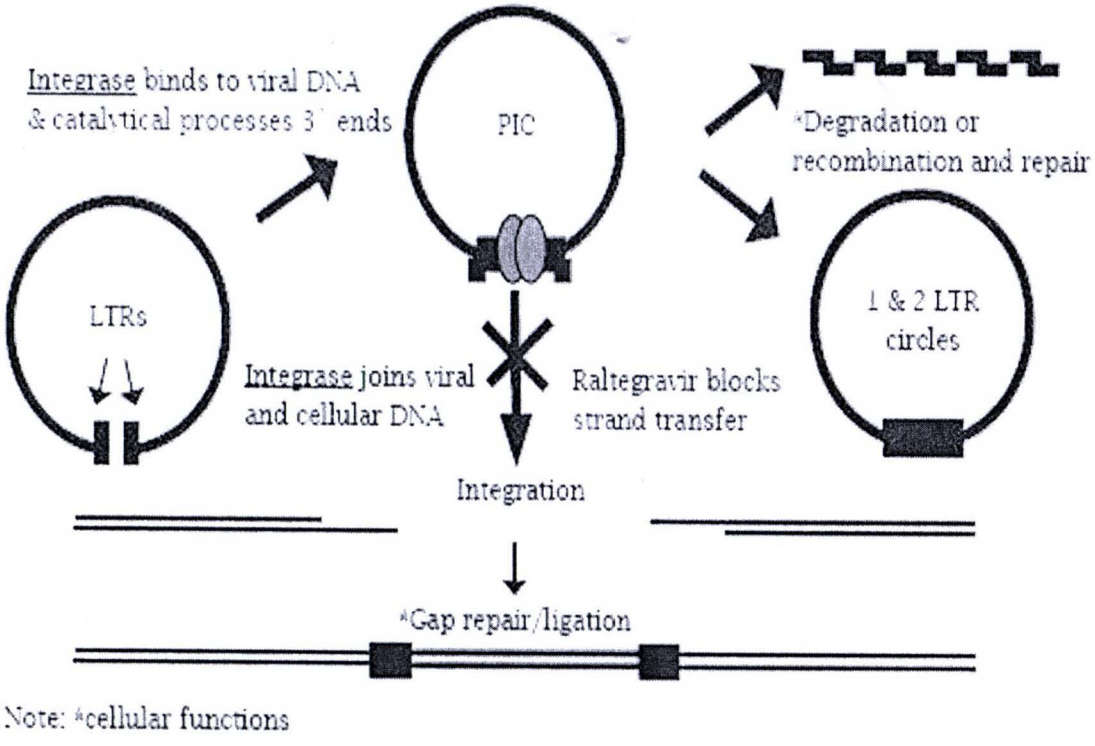

Figure 1.8 Schematic representation of the integrated and unintegrated HIV-DNA (Hazuda, 2010). Firstly, IN assembles at specific sequences at 3' end of LTR regions at each end of the fully reverse transcribed HIV-1 DNA in termed of pre-integration complex or PIC. IN then removes the terminal 3' dinucleotide from each end 3' end and joins viral DNA into the cellular DNA. Inhibition of integration results to accumulate the unintegrated forms of HIV-1 DNA, which have been reported to be a dead end product and degrade overtime.

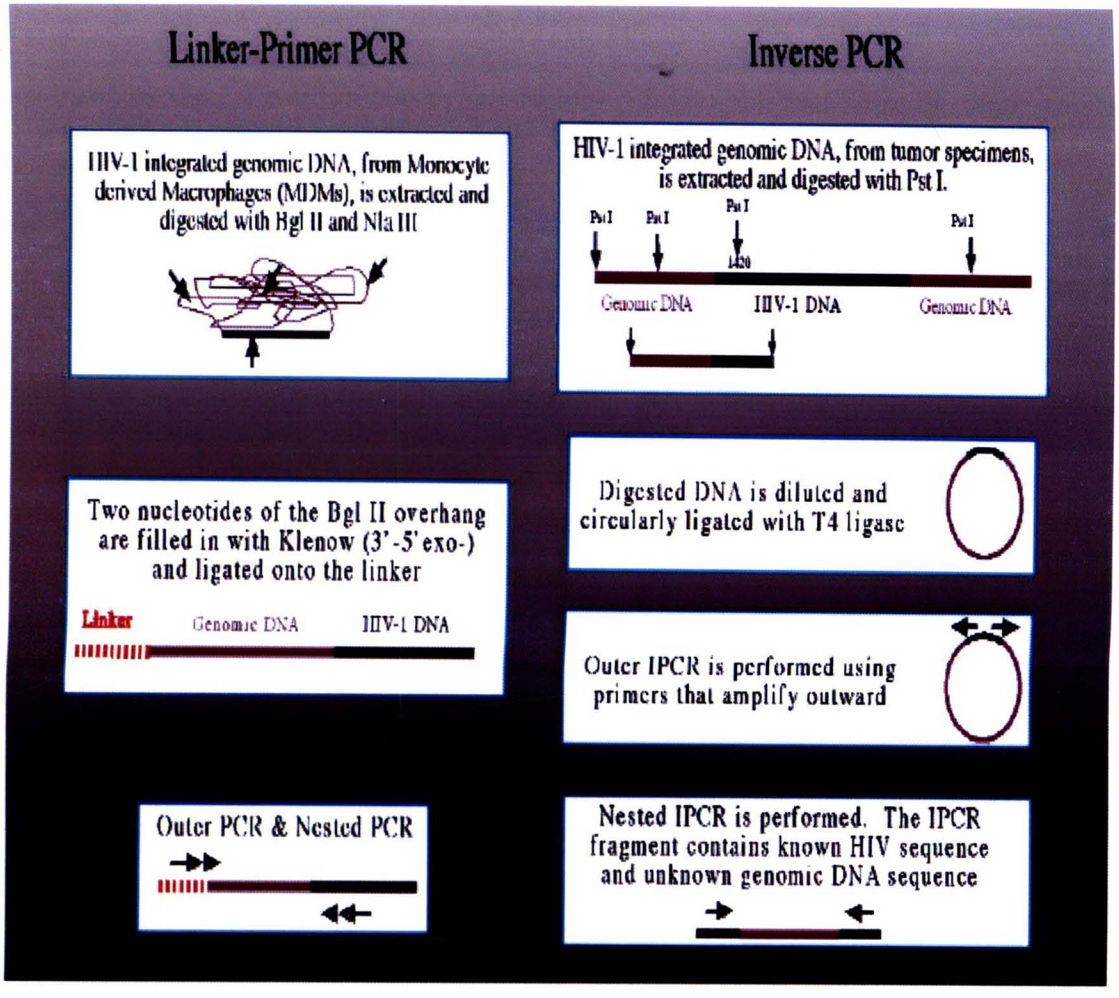

Figure 1.9 Illustration of the linker-primer PCR assay and the inverse PCR assay (Killebrew et al., 2004)**.** The linker-primer protocol involves: digestion of genomic DNA, ligation of the linker onto the cut DNA, and outer and nested PCR. The inverse PCR protocol involves: digesting genomic DNA, ligating the DNA into a circle, and performing outer and nested PCR.

In human genome, *Alu* elements are the most common repetitive and comprising over 1 million copies per diploid cell (Jelinek and Schmid, 1982; Mighell et al., 1997), around 5% of total human genomic DNA. *Alu* elements are randomly distributed, estimately 5,000 bp apart, and are randomly oriented. Early HIV-1 integration assays using *Alu* repeat elements as "anchors" within genomic DNA were sensitive but were not exactly quantitative, since they lacked real-time reaction monitoring. *Alu-gag* qPCR has recently developed (Butler et al., 2001; O'Doherty et al., 2002; Agosto et al., 2007; Liszewski et al., 2009). The principle of this technique is base on conventional PCR in the first round and real-time PCR in a second round. In the first round, one primer specific to *Alu* sequence and the other one primer bind specifically to HIV-1 *gag* sequence in both integrated and unintegrated DNA. For second round qPCR, the reaction is carried out by using specific primers and probe recognized on *LTR* region. In this way, only integrated HIV is exponentially amplified where a linear amplification occurs for unintegrated HIV DNA **(Figure 1.10).** Therefore, this technique does not need a step of digestion with restriction endonuclease and it is the least labor intensive and suitable for a big sample number or clinical use. However, this technique might be failed due to the integration sites lying too far away from an *Alu* element to be detected.

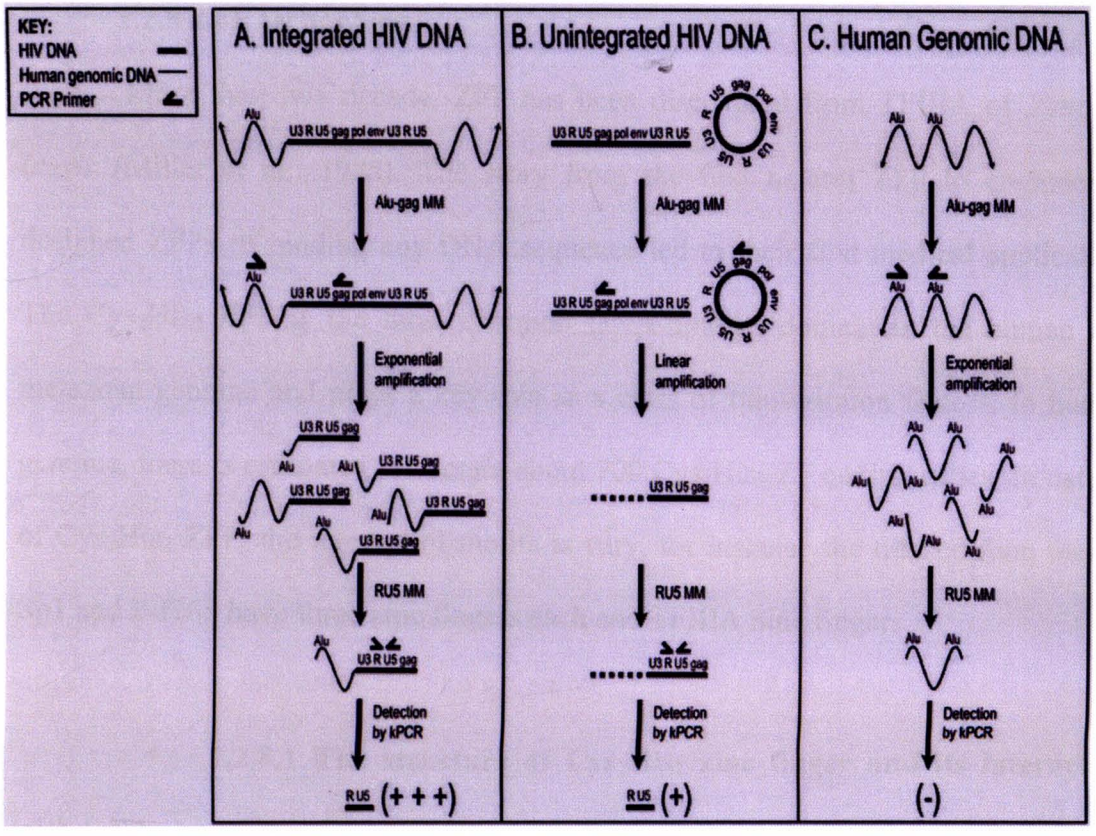

Figure 1.10 An overview of *Alu-gag* qPCR assay (Liszewski et al., 2009). Isolated DNA is a template for amplification. First, the *Alu-gag* master mixture (MM) is prepared by containing primers which bind to *Alu* and *gag* sequences. The first round PCR is then performed. **(A)** Exponential amplifying with integrated HIV DNA, as *Alu* and *gag* primers anneal to opposite strands of the same template DNA. **(B)** Linear amplifying with unintegrated HIV DNA (both circular and linear forms), as the *gag* primer anneals to only one strand of the HIV template. **(C)** *Alu–Alu* fragments of human genomic DNA are also exponentially amplified when the *Alu* elements are near together and in opposite orientation. However, the *Alu–Alu* amplicons can not be detected, only amplicons containing HIV DNA are detected because a second nested qPCR is carried out using a forward primer to R and a reverse primer to U5 and a specific probe complementary to the sequence of the HIV LTR.

1.2.8 ZFP technology

More than two decade, ZFP has been discovered from TFIIIA of *Xenopus laevis* (Miller et al., 1985). The story from the first natural ZFP to engineering designed ZFPs of binding any DNA sequence led to their first medical application. The Cys_2His_2 ZFP is the most common DNA-binding domain in the human and metazoan genome and plays a key role as a class of transcription factors. In human genome, there is estimated to contain about 700 Cys_2His_2 ZF coding genes. In natural of Cys_2His_2 ZFP, the number of motifs is vary, for instance the transcription factors Sp1 and Zif268 have three zinc fingers each and TFIIIA nine fingers.

1.2.8.1 The structure of Cys_2His_2 zinc finger and its interaction with DNA

The family of zinc-coordinating DNA-binding proteins can be subdivided into at least three classes on the basis of their structure: (i) One class is classified by presence of the Cys_6-zinc cluster motif (sometimes called Cys_6 binuclear cluster) that can be found in metabolic regulators of Fungi (e.g. yeast GAL4) (Pan and Coleman, 1990); (ii) A second class is Cys_2Cys_2 (or Cys4) zinc fingers that contains conserved Zn-binding consensus of Cys–X2–Cys–X13–Cys– X2–Cys which are predominately found in nuclear steroid or hormone receptors (e.g. glucocorticoid receptor); and (iii) The third class contains the "classical" zinc finger domain Cys_2His_2 which are present in a large number of regulatory proteins in many species of organism.

The structure of Cys_2His_2 ZFP consists of a simple ββα fold of 30 amino acids in length containing a conserved sequence of (Tyr, Phe)-X-**Cys**-X2-5-**Cys**-X3-(Tyr, Phe)-X5-leu-X2-**His**-X3-5-**His** (where X represents any amino acid; highly conserved

cysteins and histidines appear in boldface). It is stabilized by hydrophobic interactions and chelation of a single zinc ion between two histidine residues and two cysteine residues (Miller et al., 1985; Lee et al., 1989). Three dimension picture of the Cys_2His_2 ZFP is shown in **Figure 1.11**. Within the residue at the N-terminus of the α-helix, the positions -1, 3, and 6 typically bind to three contiguous base pairs in a major groove of the target DNA. However, it should be noted that in some cases, ZF domain at position 2 of the α-helix has a specific contact by overlapping with the other strand of DNA that complementary to target recognized by the amino acid in position 6 of the proceeding finger, so call target site overlap (TSO) **(Figure 1.12)**. This issue mostly happens with GNG tripets and TSO interactions can potentially reduce the independent modularity of the tripets (Mandell and Barbas, 2006).

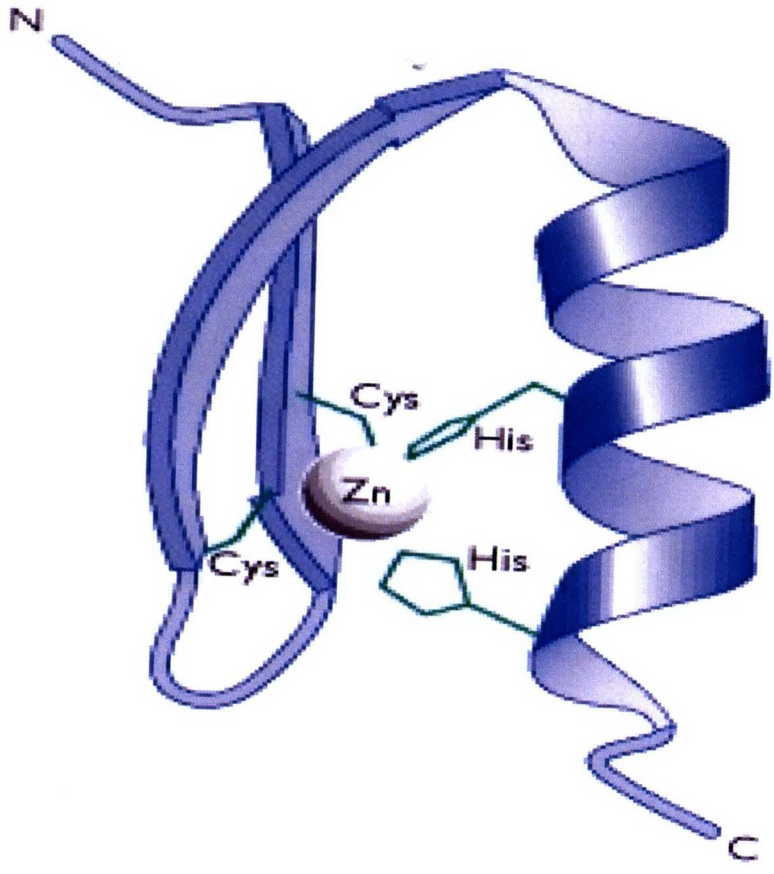

Figure 1.11 Three dimension picture of the Cys_2His_2 ZFP.

(Source: http://www.ncbi.nlm.nih.gov/books/NBK21115/)

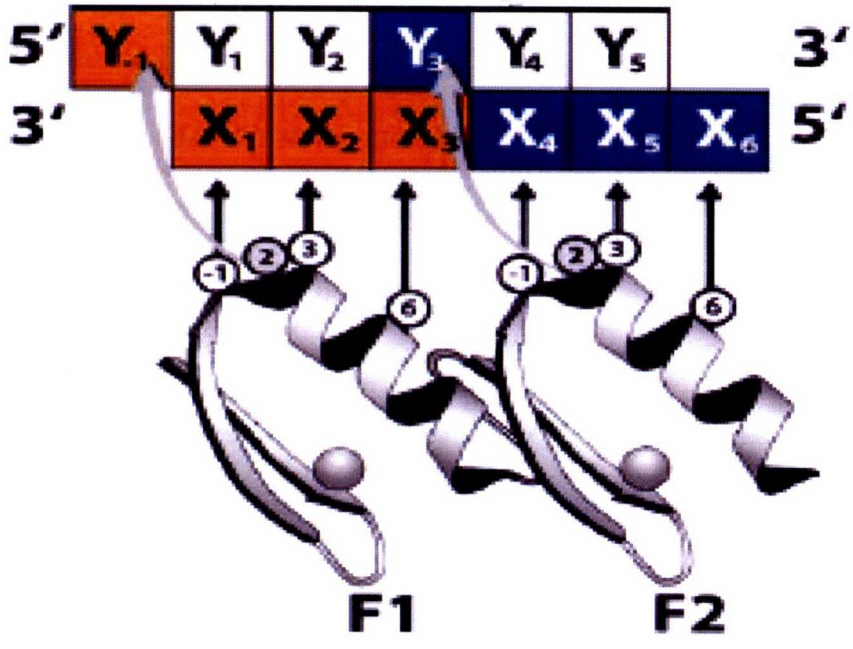

Figure 1.12 The model of DNA recognition by the classical zinc finger proteins (Papworth et al., 2006). Each classical zinc finger comprises of two β-strands (represented as large arrows) and an α-helix internally stabilized by chelatation of a single zinc ion (grey sphere). Contacts to the primary DNA strand (positions X1 to X6) are made by amino acids in positions −1, 3 and 6 of recognition α-helix (interactions marked as straight black arrows), whereas the complementary strand of DNA (positions Y−1 to Y5) is interacted by amino acid in position 2 (gray circle). Note that zinc fingers bind anti-parallel to 5′–3′ orientation of the primary DNA strand.

1.2.8.2 Design and selection of novel zinc finger proteins

Many strategies have been developed to construct Cys_2His_2 zinc fingers to bind desired DNA targets. Most of engineered zinc finger arrays are based on the zinc finger domain of the murine transcription factor Zif268, however some groups have generated them based on the human transcription factor SP1. Zif268 has three individual zinc finger motifs which specifically bind to 9 bp sequence (Christy and Nathans, 1989). The crystal structure of this protein bound to DNA was characterized in 1991(Pavletich and Pabo, 1991) and resulted in research into engineered zinc finger arrays. In 1994 and 1995, some studies used phage display to select the binding specificity of a single zinc finger of Zif268 (Choo and Klug, 1994; Jamieson et al., 1994; Rebar and Pabo, 1994; Wu et al., 1995). Normally, designed ZFP has between 3 and 6 individual zinc finger motifs and bind DNA target ranging from 9 to 18 bps in length. Arrays with 6 zinc finger motifs are the most attractive because they recognize a target sequence that is long enough to have a good chance of being unique in the mammalian genome (Liu et al., 1997). There are two main methods currently used to generate engineered zinc finger arrays, modular assembly and selection system.

The most popular method to generate new zinc finger arrays is to combine smaller zinc finger "modules" of known specificity which recognize a 3 basepair DNA sequence to generate 3 to 6 finger arrays. This concept was first described in 1991 publication about the structure of the zinc finger protein Zif268 bound to DNA (Pavletich and Pabo, 1991). Another method uses selection methods. Initially, the Barbas Laboratory of the Scripps Research Institute used phage display to develop and characterize zinc finger domains that recognize most DNA triplet sequences

(Segal et al., 1999; Dreier et al., 2001; Dreier et al., 2005). This technique is difficult to use on more than a single zinc finger at a time, so many use a processes that generates a completely optimized 3-finger (Greisman and Pabo, 1997).

Recently, Barbas's lab also developed a simple system for creating zinc finger (ZF) libraries based on three unique vectors containing all of the GNN, ANN or CNN domains which easy to use as a "cut and paste" methodology. This system was named SuperZif (Gonzalez et al., 2010). It should help researchers who are not well skilled in ZF technology to easily construct libraries and designed polydactyl ZF proteins for their studies. The advantage of this method is that it increases capability of scientists in the field of gene therapy and gene regulation to specific DNA target. The protocol can be completed in 9–11 days to create a standard six-finger protein.

1.2.8.3 Applications of designer Cys_2His_2 zinc fingers

The applications of engineered ZFP are described here and the mechanism of action on the DNA target site and the effector domain (if any) attached to ZFPs.

One application of ZFPs is to bind a specific DNA sequence. These ZFPs are not fused to any functional domains. There are 2 types of actions. One action can be used as a natural barrier for transcriptional complex progression **(Figure 1.13 A-I)**, whereas another action will be to compete with other DNA binding proteins e.g. transcription factors **(Figure 1.13 A-II)**. Basically, the binding affinity of ZFP without a functional domain is quite weak, but it can be improved by combining many modules for a stronger effect.

The second application is that ZFPs are fused to effector domains, which then bind directly with DNA. A variety of catalytic domains that act directly on DNA such as DNA methyltransferase, integrase, and nuclease can be included. These catalytic domains have been used to modify DNA in a site-specific manner **(Figure 1.13 B)**.

The other application for ZFPs is to be fused to effector domains that interact with other proteins. Adding activation or repressor domains generate new transcription factors to up or down regulates the expression of a gene of interest **(Figure 1.13 C)**. Many kinds of repressor domains have been used such as Kruppel associated box (KRAB), ERF repressor domain (ERD), or even parts of TATA-box binding protein (TBP) (Thiesen et al., 1991; Beerli et al., 1998; Stege et al., 2002; Bartsevich et al., 2003) **(Figure 1.13 C-I)**. On the other hand, widely used transcription activation domains contain the herpes simplex virus VP16 domain as well as a tetrameric repeat of its minimal activation domain (called VP64) (Beerli et al., 1998) or the p65 domain of the human endogenous transcription factor NFκB (Yaghmai and Cutting, 2002) **(Figure 1.13 C-II)**. New regulatory molecules can be constructed by using chromatin-modifying domains (e.g. histone deacetylases, acetyltransferases or methyltransferases) fused with zinc finger domain **(Figure 1.13 C-III).**

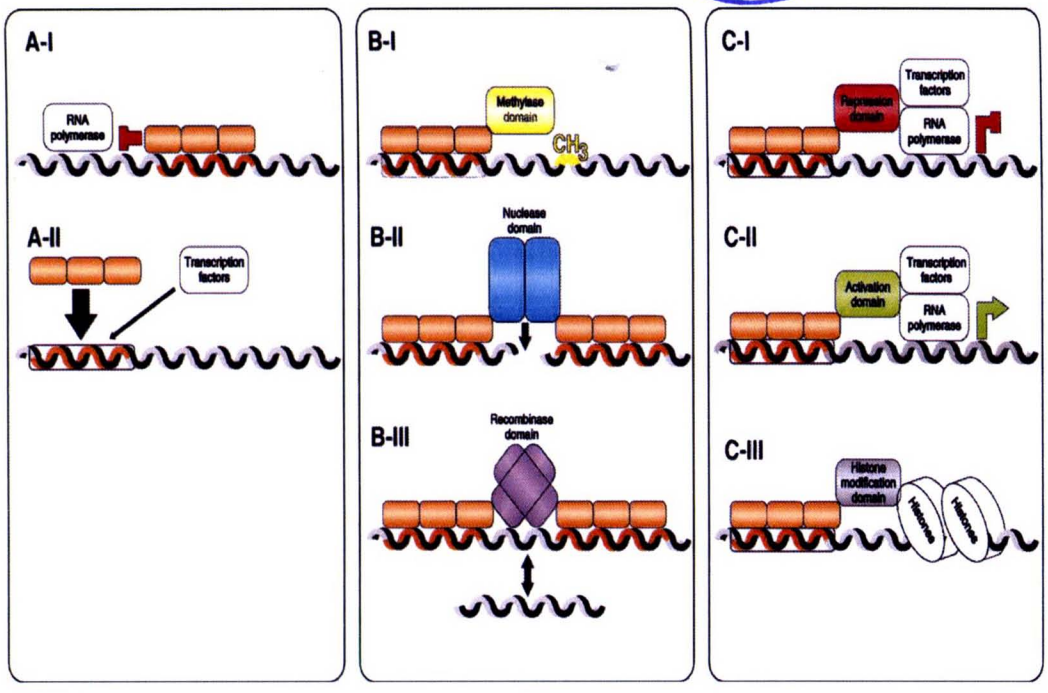

Figure 1.13 Applications of designer Cys_2His_2 zinc finger proteins (Papworth et al., 2006). The mechanism of action of engineered ZFP depends on the DNA target sequence (grey and black DNA helix) and the effector domain (if any) linked to zinc fingers (3 orange rectangles) to provide the desired function. In each case DNA target bound by ZFP is marked in orange and framed, if it sits within promoter region. No effector domains **(panel A)** ZFP can be used as a natural blockade for the transcriptional complex progression (A-I) or to compete with other proteins (natural activators or repressors), that reside the promoter region (A-II). ZFP can be also applied to produce chimeric enzymes **(panel B)** which can modify on DNA sequences e.g. methylating (B-I), cleaving (B-II), mediating recombination/integration (B-III). Designed ZFP can be used as recombinant transcription factors binding to the promoter region of the gene of interest **(panel C)** which linked to the repressor domain (C-I), or activation domain (C-II), engineered chromatin remodeling factors (C-III), respectively.

1.2.8.4 Intracellular targeting of designed zinc finger proteins

ZFPs are inherently nuclear without defined linear nucleus localisation signal (NLS), which localise in the nucleus. However, most of them were designed to target nuclear DNA and have been provided with a linear NLS to ensure their efficient transition from the cytosol to the nucleus. The examples of this strategy have been shown by targeting various DNA targets in the cell nucleus including promoter (Pollock et al., 2002), non-promoter DNA target such as telomeres (Patel et al., 2004), integrated viral DNA such as HIV-1(Isalan et al., 2001; Reynolds et al., 2003; Segal et al., 2004). However, the possible future applications of ZFP to bind or regulate its target site in various cellular compartments and sub-compartments have been discussed (Papworth et al., 2006) and show in **Figure 1.14.**

Designed Cys_2His_2 zinc finger proteins are made and correctly folded in the cytoplasm, however there are no natural targets for them in this intracellular compartment. Since natural and artificial zinc fingers can bind a sequence specific manner with RNA as well as with DNA (Brown, 2005), they can be applied to target cytoplasmic nucleic acids. This possibility includes genomes of RNA viruses, which replicate in the cytoplasm such as critical human pathogens e.g. Hepatitis A, and C, rubella, influenza, and polio. Large DNAviruses such as poxviruses and a family of African swine fever viruses replicate in the cytoplasm and their DNA could be also targeted with designed zinc fingers for not only research, but also therapeutic interventions.

Moreover, a variety of genetic disorders have been showed to associate with point mutations, deletions or rearrangements in human mitochondrial DNA (Schon, 2000). There is no efficient treatment for these diseases resulted from mtDNA

mutations, whereas gene therapy has not been successful proven due to difficulties in transfering active DNA to mitochondria (Dimauro et al., 2004). Therefore, zinc finger technology has the potential to be used as a novel therapeutic tool to treat patients with mitochondrial defects by targeting mutations with specific ZFP to inhibit the replication of mutant mtDNA allowing propagation of only the wild-type DNA molecules (Choo et al., 1994).

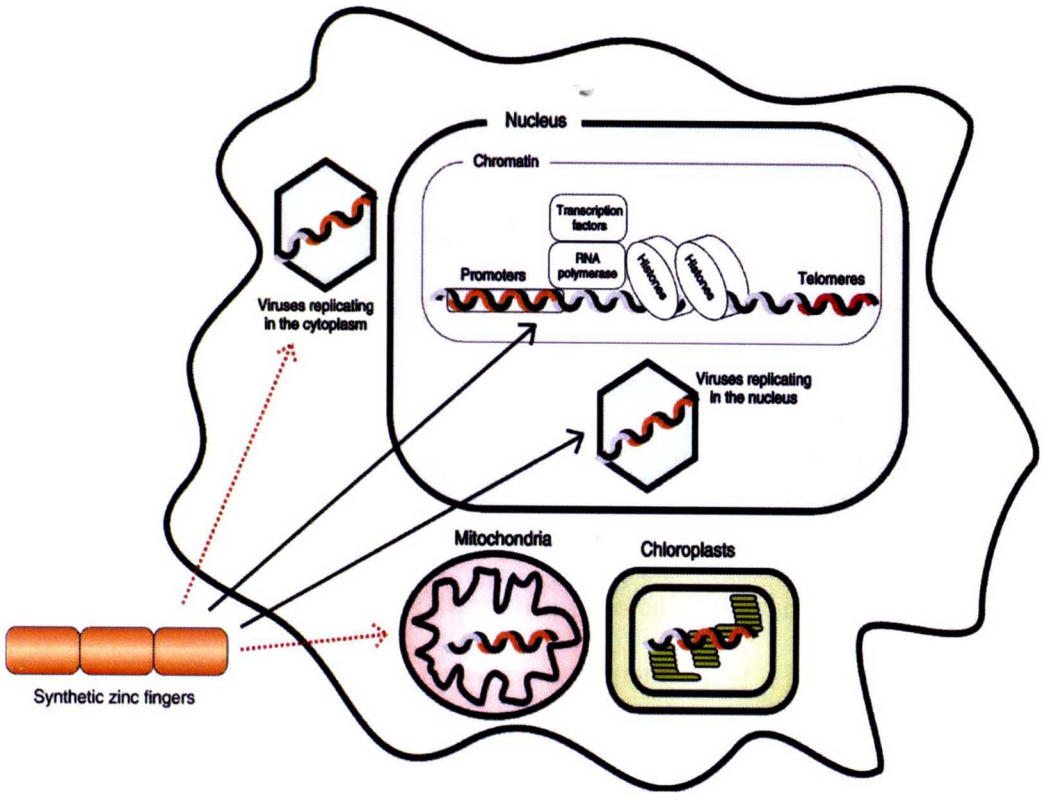

Figure 1.14 Intracellular targeting of designed zinc finger proteins (Papworth et al., 2006). Artificial ZFPs based on the natural nuclear proteins have been successfully synthesized against various DNA targets in the cell nucleus (black solid arrows). This includes promoter (orange, frame) and non-promoter targets in chromatin (including integrated viruses), structural elements like telomeres (red) and not integrated viruses. Since all the ZFPs originate in the cytoplasm, it should be capable to direct them to intracellular compartments other than nucleus, which also contain DNA (red dotted arrows) such as mitochondria or chloroplasts. It is also possible to use ZFPs targetting DNA or RNA viruses, which replicate in the cytoplasm as well as other cytoplasmic targets e.g. specific RNA.

1.2.9 Lentivirus gene transfer system

Viral vectors are more familiar in both clinical and non-clinical research. The evolution and the capability of viruses to deliver recombinant DNA molecules (transgenes) into host cells have led them to be useful for many applications. The most applications for clinical gene therapy in current worldwide are using viral vectors derived from retroviruses (including lentiviruses), adenoviruses, adeno-associated viruses, herpes viruses and poxviruses (Walther and Stein, 2000).

Lentiviral vectors are a type of retrovirus that can infect not only in dividing cells, but also in non-dividing cells as well (Bukrinsky et al., 1993; Gallay et al., 1997). It has been developed replication defective vectors from the lentivirus HIV and demonstrated that they transduce target cells independent of mitosis (Naldini et al., 1996b). It showed highly efficient for gene delivery in vivo and obtained stable long-term expression of transgene in several target tissues, such as the hematopoietic system (Amado and Chen, 1999), the retina (Greenberg et al., 2006), the neurons (Naldini et al., 1996a), muscle and liver (Kafri et al., 1997).

To construct a lentiviral gene therapy vector, a reporter gene or therapeutic gene is cloned into a vector sequence that is flanked by LTRs and the Psi-sequence (Ψ) of HIV. The overall construction of lentiviral gene therapy vector is showed in **Figure 1.15.** The LTRs are needed to integrate the therapeutic gene into the genome of the target cell, just as the LTRs in HIV integrate the dsDNA copy of the virus into its host chromosome. The Psi-sequence (Ψ) acts as a signal sequence and is compulsory for packaging RNA with the reporter or therapeutic gene in virions. Viral proteins which make virus shells are made in the packaging cell line, but are not in

construct of the LTRs and Psi-sequences (Ψ) and so are not packaged into virions. Thus, virus particles are generated that lack of replication duty, so are designed to be unable to persist to infect their host after they deliver their therapeutic transgene (Neala, 1999).

In a first generation of HIV-derived vectors, viral particle comprises all of the HIV-1 proteins except the envelope from the vesicular stomatitis virus (VSV-G) (Naldini et al., 1996a). In second version of lentiviral vector, four additional genes, encoding proteins likely to act as a crucial virulence factors: *Vpr*, *Vif*, *Vpu*, and *Nef* were deleted to improve the biosafety safeguards (Zufferey et al., 1997). The third generation packaging unit of HIV-1-based vectors which conserves only three of the nine genes of the parental virus: *gag*, *pol*, and *rev* (Dull et al., 1998). This construction was eliminated the possibility that a wild-type virus will be reconstituted through recombination.

Normally, lentiviral vectors are designed in a transient transfection system in which a cell line is transfected with three individual plasmid expression systems. These contain the transfer vector plasmid (parts of the HIV provirus), the packaging plasmid or construct, and a plasmid with the heterologous envelop gene (ENV) of a different virus (Amado and Chen, 1999). The three plasmid vectors are tranfected into a packaging cell which are then formed the virion. The virus portions of the vector contain transgenes so that the virus cannot replicate inside the cell system.

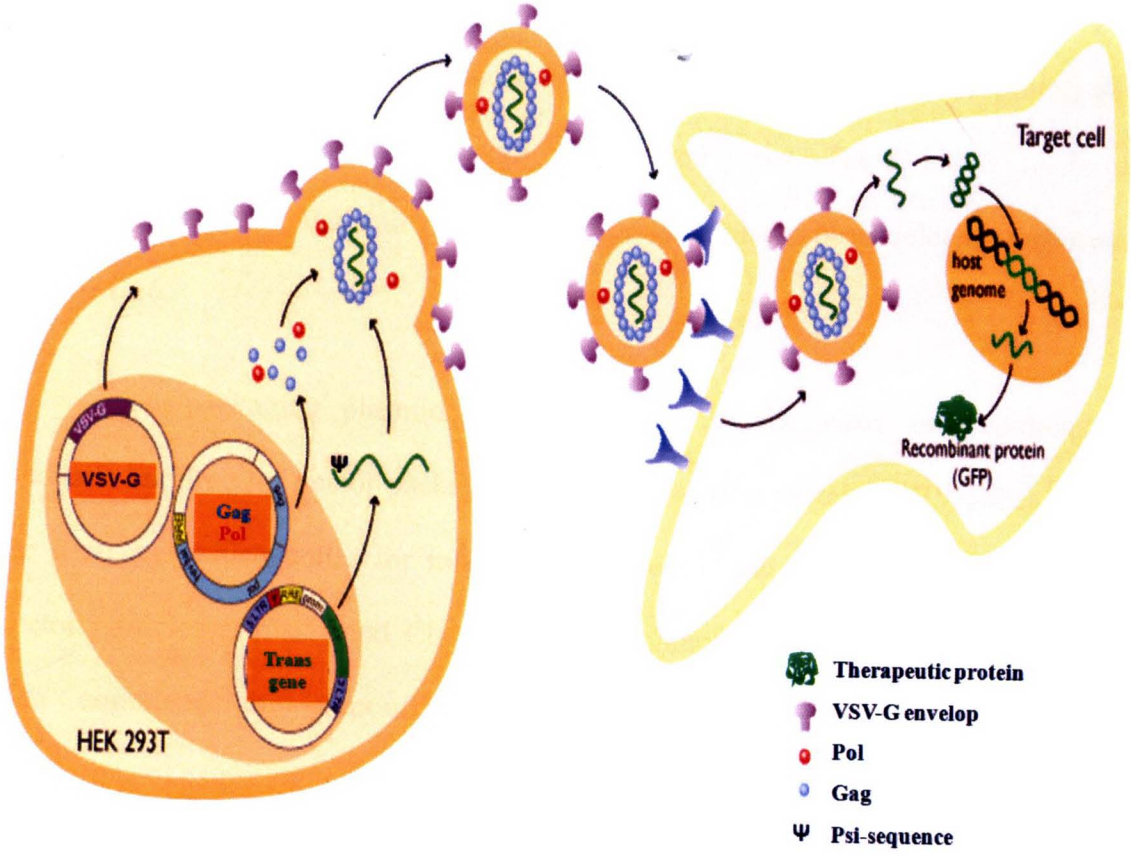

Figure 1.15 The overall picture of lentiviral gene therapy. (Adapted from http://www.invivogen.com/sscat.php?ID=116&ID_cat=16) Transfer vector is flanked by LTRs and the Psi-sequence (Ψ) of HIV. This vector is transfected or cotransfected with packaging plasmids into a packaging cell line which can produce viral proteins to form virus shell. The RNA of transgene is packed into virion. This virion contains only RNA of transgene without sequences of viral components. Therefore, it lack of responsibility of replication after they deliver their therapeutic transgene. RNA transgene is reversed into double-stranded DNA by RT and is integrated into a host genome by IN. Then, translation occurs to produce therapeutic protein.

The transfer vector plasmid consists cis-acting genetic sequences necessary for the vector to infect the target cell and for delivery of the therapeutic (or reporter) gene and contains restriction sites for insertion of desired genes. The 3' and 5' LTRs, the original envelop proteins, and *gag* sequence promoter have been deleted as shown in **Figure 1.16.**

The packaging plasmid is the backbone of the virus and contains the elements required for vector packaging such as structural proteins, HIV genes (except the gene *env* which codes for infection of T cells), and the enzymes that produce vector particles (Amado and Chen, 1999). The human cytomegalovirus (hCMV) is also contained for responsibility in the viral proteins expression during translation. The packaging signals and their adjacent signals are removed as shown in **Figure 1.17.** Thus, the packaging sequences will not be incorporated into the viral genome and the virus will not reproduce since it has infected the host cell.

The third plasmid's envelope gene of a different virus specifies what type of cell to target and infect instead of the T cells (Amado and Chen, 1999). Normally HIV can infect helper T-cells by using gp120 protein to bind with the CD4 receptor. However, it is useful for lentiviral vector to infect the different cell type on gene therapy by exchange the CD4 receptor-binding protein instead of another protein. This facilitates the HIV lentiviral vector a broad range of possible target cells. There are two kinds of heterologous envelope proteins as the amphoteric envelop of MLV and the G glycoproteins of the vesicular stomatitis virus, known as VSV-G as shown in **Figure 1.18.**

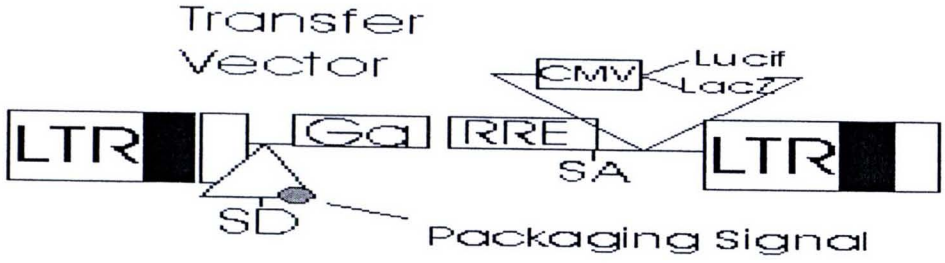

Figure 1.16 The construction of transfer vector (Neala, 1999).

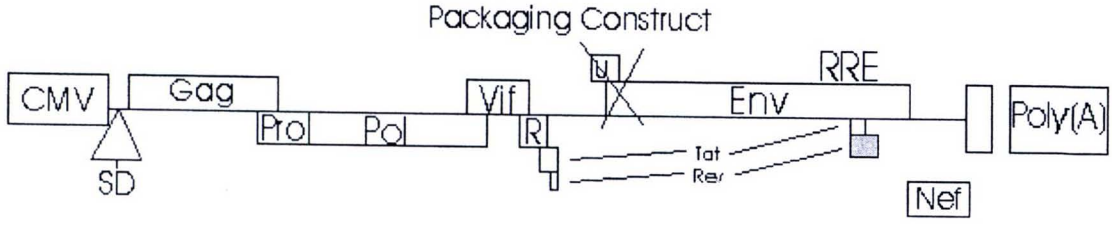

Figure 1.17 The construction of packaging vector (Neala, 1999).

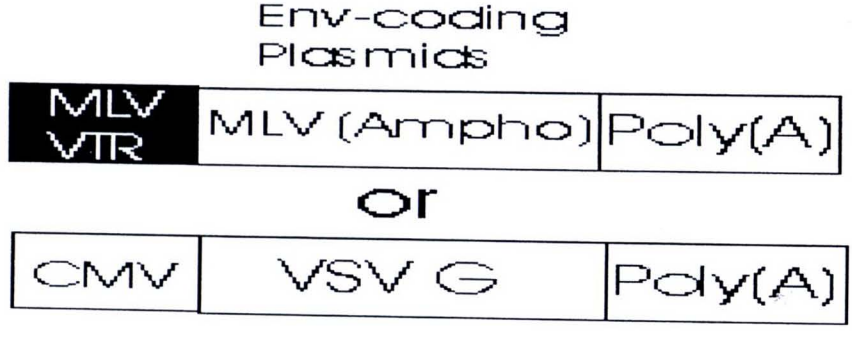

Figure 1.18 The construction of envelop genes in the third plasmid

(Neala, 1999).

Because the lentiviruses, such as HIV have been successfully shown the efficient of gene delivery vehicles, the development of lentiviral vector has now turned its attention to producing vectors with built-in safety features to prevent the opportunity of replication competent lentiviruses (RCL). However, even the earliest studies with HIV lentiviral vectors did not generate RCL in vitro or in vivo (Amado and Chen, 1999), but precautions are still concerned.

The Salk Institute's researchers have created HIV lentiviral vectors that are self-inactivating. The scientists are working on packaging a defective HIV genome that contains only the necessary elements for gene transduction into a virion that has a broad host range. A deletion of the transcriptional enhancers and promoter in the U3 region of the 3' LTR of the DNA so-called self-inactivating (SIN) vector **(Figure 1.19)** used to produce the vector RNA. During reverse transcription, this deletion is transferred to the 5'LTR of the proviral DNA **(Figure 1.20)**. This sequence is eliminated to destroy the transcriptional activity of the LTR, which the generation of full-length vector RNA in transduced cell is abolished. The potential for replication competent lentiviruses (RCL) production in the target cell is further minimized (Amado and Chen, 1999). A 4 plasmid vectors is used in the SIN vector system **(Figure 1.21)**. These vector designs provide significant biosafety features. The contribution of HIV is reduced to a fraction of *cis*-acting sequences in the vector, leaving out in particular most of the LTR, and to only three genes, *gag*, *pol*, and *rev*, in the packaging constructs, compared with the nine genes necessary for the in vivo replication and pathogenesis of wild-type HIV-1 (Aldrovandi and Zack, 1996; Haynes et al., 1996).

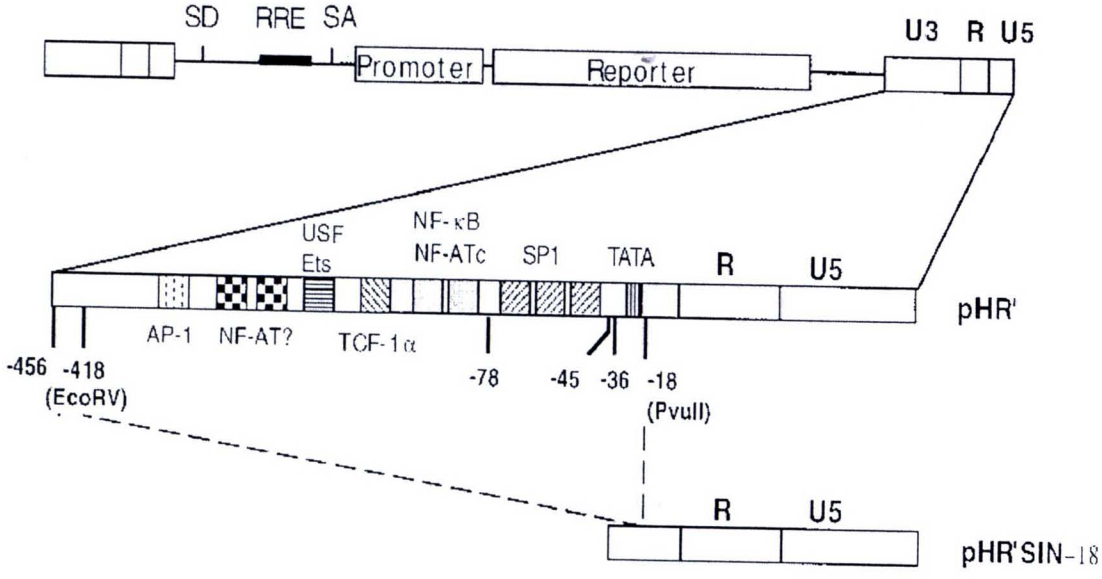

Figure 1.19 Structure of SIN HIV-derived vectors (Zufferey et al., 1998). A schematic representation of an HIV-1 vector with enlarged 3' LTR to show the binding sites for different transcription factors on U3 is shown (not to scale). Although the 3' LTR is depicted, the nucleotide numbering refers to the cap site at the beginning of R as +1 as for a 5' LTR. Position -418 is the 5' limit of all deletions and positions -78, -45, -36, and -18 indicate the 3' limits of the different type of the SIN-vector. SD, splice donor; RRE, Rev-response element; SA, splice acceptor. The GenBank accession number for the wild-type 3' LTR is M1991.

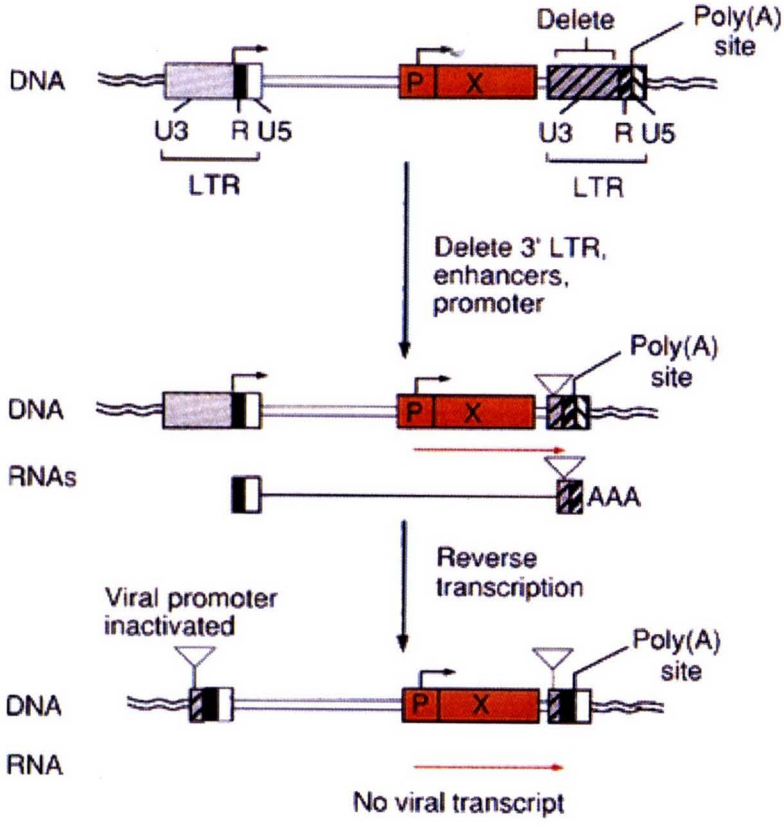

Figure 1.20 Construction and replication cycle of self-inactivating vector (Coffin et.al., 1997). The enhancers and/or promoter in the 3'LTR of a standard lentiviral vector (*top*) is deleted to make the SIN vector (*middle*), which, after reverse transcription, has partially deleted LTRs at both ends (*bottom*). Arrows indicate promoters.

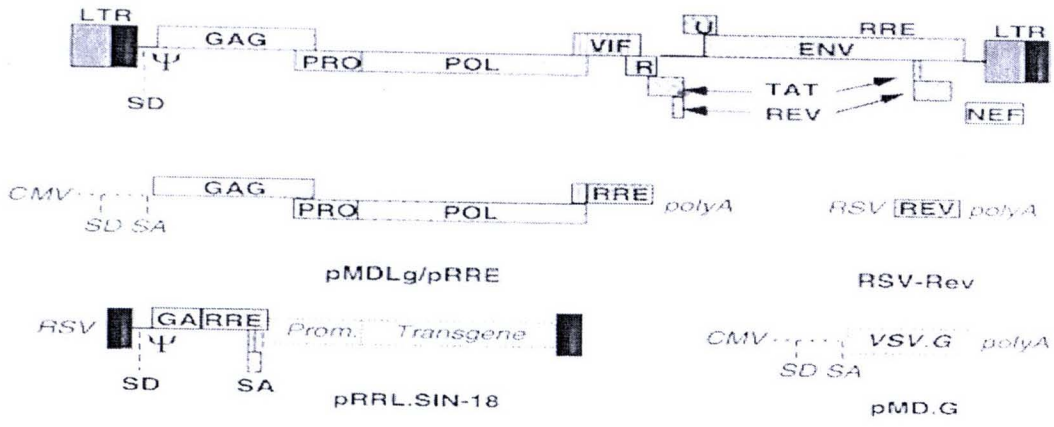

Figure 1.21 Schematic drawing of the HIV provirus and the four constructs used to make a lentivirus vector of the SIN system (Dull et al., 1998). The viral LTRs, the reading frames of the viral genes, the major 5' splice donor site (SD), the packaging sequence (Ψ), and the RRE are boxed and indicated in bold type. The conditional packaging construct, pMDLg/pRRE, expresses the *gag* and *pol* genes from the CMV promoter and intervening sequences and polyadenylation site of the human β-globin gene. As the transcripts of the *gag* and *pol* genes contain *cis*-repressive sequences, they are expressed only if Rev promotes their nuclear export by binding to the RRE. All *tat* and *rev* exons have been deleted, and the viral sequences upstream of the *gag* gene have been replaced. A nonoverlapping construct, RSV-Rev, expresses the *rev* cDNA. The transfer construct, pRRL.SIN-18, includes HIV-1 *cis*-acting sequences and an expression cassette for the transgene. The 5' LTR is chimeric, with the enhancer/promoter of Rouse sarcoma virus (RSV) replacing the U3 region (RRL) to rescue the transcriptional dependence on Tat. The 3' LTR has an almost complete deletion of the U3 region, which includes the TATA box (from nucleotides 2418 to 218 relative to the U3/R border. The fourth vector, pMD.G, encodes a heterologous envelope to pseudotype the vector, here shown coding for VSV-G. Only the relevant parts of the constructs are shown.

1.3. Objectives

1. To design and construct ZFP which bind to the integrase recognition sequence at 2-LTR-circle junctions
2. To study binding activity of designed ZFP with its target DNA by surface plasmon resonance (SPR) and an electrophoretic mobility shift assay (EMSA)
3. To construct the lentiviral vector and produce viral carrying designed ZFP gene for gene delivery
4. To produce the SupT1 stable cell line expressing designed ZFP for further evaluating the intracellular inhibition
5. To investigate intracellular inhibition of viral integration by designed ZFP