PREDICTION OF CO-RECEPTOR BINDING OF HIV-1 VIRUS OF KELANTAN BASED ON SEQUENCE ANALYSIS OF V3 LOOP STRUCTURE

by

ONG SI MIN

Dissertation submitted in partial fulfillment of the requirement for the degree of Bachelor of Health Sciences (Biomedicine)

May 2013

CERTIFICATE

This is to certify that the dissertation entitled "PREDICTION OF CO-RECEPTOR BINDING OF HIV-1 VIRUS OF KELANTAN BASED ON SEQUENCE ANALYSIS OF V3 LOOP STRUCTURE" is the bona fide record of research work done by MISS ONG SI MIN during the period from September 2012 to May 2013 under my supervision.

CERTIFICATE

This is to certify that the dissertation entitled "PREDICTION OF CO-RECEPTOR BINDING OF HIV-1 VIRUS OF KELANTAN BASED ON SEQUENCE ANALYSIS OF V3 LOOP STRUCTURE" is the bona fide record of research work done by MISS ONG SI MIN during the period from September 2012 to May 2013 under my supervision.

Supervisor,

.....

Dr. Maizan Mohamed
Lecturer
Institute for Research in Molecular Medicine (INFORMM)
Health Campus
Universiti Sains Malaysia
16150 Kubang Kerian
Kelantan, Malaysia

Date: .....

ACKNOWLEDGEMENT

I would like to acknowledge and extend my heartfelt gratitude to the following persons who have made the completion of this project possible. First, I would like to take this opportunity to offer my profoundest gratitude to my supervisor, Dr. Maizan bt Haji Mohamed for all her guidance, valuable and constructive suggestions and advices during planning and development of this project. Without her kind and patient instruction, it is impossible for me to finish this project successfully. I would also like to give my gratitude to Dr. Wan Nazrah Wan Yusoff for her kindness in providing the samples and not to forget Miss Che Wan Salma Che Wan Zalati for her assistance in completing this project.

I would also like to express my deepest gratitude to the members and staffs of INFORMM Health Campus for their great support and cooperation in allowing me to use their equipment and performing lab work at INFORMM. Great deals appreciation to Dr. See Too Wei Cun, our Final Year Project Course Coordinator for his suggestion, guidance and discussion about completing this dissertation. My thanks and appreciations also go to my colleague Miss Hii Siew Ching and Miss Noor Hamizah bt Minal in developing the project and people who have willingly helped me out with their abilities.

Most thanks especially to my parents and sibling for their unconditional love, support, blessings, wishes and encouragement for the successful of this final year project. Last but not least, thanks to all of those who have supported me in any aspects during the completion of this dissertation especially all my fellow course mates. Above all, I would like to thank God, who made all things possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMNS

°C	Degree Celsius
bp	Base pair
cDNA	Complementary DNA
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
EDTA	Ethylene Diamine Tetra Acetic Acid
g	Gravity
H_2O	Water
Min	Minute
M	Molar
mM	Milimolar
µL	Microlitre
mL	Mililitre
mg	Miligram
µg	Microgram
ng	Nanogram
NFW	Nuclease free water
OD	Optical density
PCR	Polymerase chain reaction
ρM	Picomolar
ssRNA	Single stranded ribonucleic acid
RNA	Ribonucleic acid
s	Second
U	Unit
UV	Ultraviolet
TBE	Tris borate EDTA
rpm	Revolution per minute
RT	Reverse transcriptase
VL	Viral load
NRTIs	Nucleotide reverse transcriptase inhibitors
NNRTIs	Non-nucleoside reverse transcriptase inhibitors
BLAST	Basic Local Alignment Search Tools
NCBI	National Center for Biotechnology Information
HIV	Human Immunodeficiency Virus
AIDS	Acquired Immunodeficiency Disease Syndrome
IDU	Intravenous drug user
CRFs	Circulating Recombinant Forms
WHO	World Health Organization
USA	United States of America
EU	European Union
CDC	Center for Disease Control
SIV	Simian immunodeficiency virus
CD	Cluster of differentiation
Gp	Glycoprotein
P	Protein
V3	Third hypervariable

STI	Sexually transmitted infection
HSV	Herpes simplex virus
ARV	Antiretroviral agents
HAART	Highly active antiretroviral therapy
ART	Antiretroviral therapy
UNICEF	United nation children fund
mm^3	Millimeter cube
cm	Centimeter
env	Envelope
pol	Polymerase
gag	Group specific antigen
AZT	Zidovudine
PIs	Protease inhibitor
FIs	Integrase inhibitor
CRAs	Chemokine receptor antagonists
Tm	Melting temperature
Ta	Annealing temperature

ABSTRAK

Lingkaran V3 virus human immunodeficiency 1 (HIV-1) adalah penting untuk pengikatan ko-reseptor dan ia merupakan penentu utama penggunaan sellular ko-reseptor, CCR5 atau CXCR4. Virus menggunakan lingkaran V3 ini untuk memasuki sel. Ia juga merupakan penentuan untuk pembentukan syncytium, infectiviti virus dan peneutralan antibodi. Ramalan penggunaan ko-reseptor virus oleh pesakit adalah penting kerana ia membolehkan pemilihan ubat-ubatan peribadi yang berkesan dan prognosis perkembangan penyakit. Kajian ini bertujuan untuk menentukan variasi jujukan struktur lingkaran V3 dan meramalkan penggunaan ko-reseptor oleh virus HIV-1 dari Kelantan dengan menggunakan alat bioinformatik untuk memantau perkembangan penyakit pesakit. Ia juga bertujuan untuk menentukan asal-usul virus HIV-1 dari Kelantan dengan membina pokok filogenetik berdasarkan rantau V3 virus. Dalam kajian ini, lingkaran V3 30 sample virus HIV-1 dari Kelantan telah berjaya diamplifikasikan dan amplicons ini telah dijujukan dan diselaraskan dengan menggunakan perisian Bioedit. Motif hujung virus HIV-1 telah ditentukan dengan menterjemahkan urutan nukleotida ke asid amino. Pengunaan ko-reseptor oleh virus berkenaan akan diramal dengan menggunakan dua alat bioinformatik, Geno2pheno (Max Planck Institut Informatik) dan Profil berasakan Rententan Kernel Software (Corbeil Research Group). Keputusan menunjukkan bahawa dalam kalangan 30 sampel; 29 sampel telah menggunakan CCR5 sebagai ko-reseptor dengan motif hujung GPGQVFYRTGDETGDI manakala hanya satu virus yang menggunakan CXCR4 sebagai ko-reseptor dengan motif hujung GPGQVFYRTGDITGDL. Berdasarkan keputusan filogenetik dan BLAST, 30 virus telah dikenalpasti sebagai Circulating Recombinant Forms (CRFs) dengan 10 sampel dikalsifikasi kepada CRF01_AE dan 20 adalah rekombinan baru subjenis AE/B. Virus HIV-1 berkait

rapat dengan virus HIV-1 yang diterbit dari Thailand, Singapura dan Malaysia sendiri. Alat bioinformatik boleh dicadangkan sebagai ramalan awal penggunaan ko-reseptor oleh virus HIV-1 untuk memantau perkembangan penyakit.

ABSTRACT

The V3 loop of human immunodeficiency virus type 1 (HIV-1) is critical for co-receptor binding and it is the main determinant of which cellular co-receptor is used, CCR5 or CXCR4. The virus uses this V3 loop for cell entry. It is also the determinant for syncytium formation, virus infectivity and antibody neutralization. Accurate prediction of the co-receptor used by the virus in the patient is important as it allows personalized selection of effective drugs and prognosis of disease progression. This study aims to determine the sequence variation of V3 loop structure and predict the co-receptor used by HIV-1 viruses from Kelantan isolate using bioinformatics tools to monitor the disease progression of the patients. It is also aims to determine the origin of HIV-1 viruses from Kelantan by constructing a phylogenetic tree based on V3 region of the virus. In this study, the V3 loop of thirty HIV-1 viruses from Kelantan was successfully amplified and these amplicons were sequenced and aligned using Bioedit software. The tip motif of the HIV-1 virus was determined by translating the nucleotide to amino acid sequence. The co-receptor used by the virus was predicted using two bioinformatics tool, Geno2pheno (Max Planck Institut Informatik) and Profile-based String Kernel Software (Corbeil Research Group). The results reveal that among the thirty samples; twenty-nine samples were using CCR5 as co-receptor with tip motif GPGQVFYRTGDETGDI while only one virus that utilized CXCR4 as co-receptor with tip motif of GPGQVFYRTGDITGDIL. Based on phylogenetic and BLAST results, 30 viruses were identified as Circulating Recombinant Forms (CRFs) virus with 10 samples were classified into CRF01_AE and 20 were the new recombinant of AE/B subtype. The HIV-1 viruses were closely related to the published HIV-1 viruses from Thailand, Singapore and Malaysia itself. The bioinformatics tools could be suggested

as preliminary prediction of co-receptor used by HIV-1 virus for prognosis of disease progression.

CHAPTER 1 INTRODUCTION

1.1 Background of Study

Human Immunodeficiency Virus (HIV) was identified in 1983 with the first reported cases of Acquired Immunodeficiency Syndrome (AIDS) in 1981 and 1982. It is a disease caused by a virus called Human Immunodeficiency Virus (HIV). The illness alters the immune system, causing people to be more vulnerable to infections and this susceptibility worsens as the disease progresses

(Adapted from: Medical News Today, Updated 5 April 2013).

According to Center of disease control and prevention (CDC), there are two types of HIV, which are HIV-1 and HIV-2 that circulating among human population. Both types of HIV are able to damage a person's body by destroying specific blood cells, called CD4+ T cells, which are crucial in helping the body to fight diseases.

There are approximately 34 million people currently living with HIV and nearly 30 million people have died of AIDS-related causes since the beginning of the epidemic (UNAIDS, 2012). According to the latest estimates from UNAIDS there were approximately 34.0 million people living with HIV in 2011, rising from 29.4 million in 2001, the result of continuing new infections, people living longer with HIV, and general population growth.

Most of the new infections are transmitted heterosexually, although the risk factors may vary. In some countries, homosexual, injecting drug users, and sex workers are at a significant risk (WHO/UNAIDS/UNICEF, 2011). HIV not only affects the health of individuals, but it also brings impacts on households, communities, and the

development and economic growth of nations. Many countries that had hardest hit by HIV also suffer from other infectious diseases, food insecurity, and other serious problems (UNAIDS, 2010).

The essence of HIV-1 infection is a slow decline in CD4+ T-cells over time, such that once a threshold of approximately 200 x 10^9 CD4+ cells/L is passed, immune deficiency and virally-induced tumours are increasingly liable to occur (Weber. J., 2001). After the virus successfully attach to the host cells, fusion of the viral and host cell membranes takes place (D'Souza *et al.*, 1996; Berger *et al.*, 1999). The co-receptor selectivity of the viral host cell attachment is the central of pathologic and clinical importance of HIV infection.

HIV-1 enters human cells in a process that comprises of several steps, which includes the binding of the viral gp120 protein to the cellular receptor protein CD4+ T cells and a co-receptor protein, usually one of the two chemokine receptors CCR5 and CXCR4 (D'Souza *et al.*, 1996). The type of co-receptor used by the virus, the so-called co-receptor tropism, has a prognostic value, since patients with a CXCR4-tropic virus (''X4 virus'') progress faster to Acquired Immunodeficiency Syndrome (AIDS) compared to patients with a CCR5-tropic virus (''R5 virus'') (Koot *et al.*, 1993).

In addition to the purely X4 and R5 tropic viruses, there are also a ''dual-tropic'' strains where the viruses are able to use both co-receptors (''R5X4 virus''). After binding to CD4+ cells, a conformational switch in the surface protein gp120 of HIV reveals the co-receptor binding site, most notably the third hypervariable (V3) loop

region. Hence, the V3 loop is considered as the major viral determinant for co-receptor specificity (Jensen *et al.*, 2003).

In the usage of co-receptor of HIV-1 virus, it has been shown that viruses binding to CCR5 are almost exclusively present during the early asymptomatic stage of the infection whereas CXCR4-binding viruses may emerge in later phases of the infection and are associated with a CD4+ T-cell decline and progression towards AIDS (Miedema *et al.*, 1994). The specificity of the virus to use one of the co-receptors is often termed tropism. Understanding the co-receptor tropism and genetic diversity of circulating HIV-1 strains is crucial for AIDS treatment and vaccine development (Zhang *et al.*, 2009).

Predictive methods for inferring co-receptor usage based on V3 loop region of the envelope protein gp120 using genotypic methods can provide us with these monitoring facilities while avoiding expensive phenotypic tests. Although phenotypic assays for monitoring co-receptor usage are commercially available, they are time-consuming and costly. In order to become a routine part of clinical diagnosis, a cheaper and faster genotypic analysis is desired (Sing *et al.*, 2006). Hence, through the prediction of co-receptor based on analysis of V3 loop structure will be very helpful in the development of antiviral therapy or vaccine for the treatment of HIV infection.

1.2 Problem Statement and Rationale of Study

In this era of globalization and with the advancement of medical technologies, human nowadays are able to predict the stages of HIV infection using bioinformatics tools. The prediction of HIV virus tropism carries significant aids in the developing of new and effective vaccines or anti-viral therapy in the efforts to prevent or delay the progression of disease or even to cure HIV infections among human population.

Up to date, phenotypic and genotypic assays have been developed to assess HIV-1 co-receptor usage that may be used to determine HIV-1 tropism, and consequent debate as to which is the most appropriate for the routine clinical use (Rose *et al.*, 2009; Poveda *et al.*, 2006, 2007). Most phenotypic assays such as PhenX-R method, Virco platform, Trofile™, and HIV-1 Phenoscript Env™ are based on recombinant viruses that bear the challenged envelope glycoprotein. These are the gold standard method for predicting HIV-1 co-receptor usage (Raymond et al., 2009; Trouplin *et al.*, 2001; Whitcomb *et al.*, 2007).

Phenotypic assays based on recombinant viruses appear to give the most accurate results (Weber *et al.*, 2006) and Trofile™ (Monogram Biosciences, San Francisco, Ca) is the most widely used test. Yet, these methods are quite laborious, expensive, and feasible only in sophisticated facilities. In addition, comparisons between distinct phenotypic tests have displayed substantial disagreement (Skrabal *et al.*, 2007).

In some studies related to the structure of HIV virus, HIV-1 envelope V3 region has been shown to be the major determinant of viral tropism (Cormier and Dragic, 2002), prediction of co-receptor usage based on the interpretation of V3 sequences using

bioinformatics tools could be a good alternative to infer HIV-1 tropism in clinical settings when phenotypic test are not available. Therefore, more rapid and easier tools for estimating HIV-1 tropism are desirable (Low *et al.*, 2007). The genotypic determination of co-receptor usage presents advantages such as shorter turnover time, lower costs, possibility to adapt the results to the patients' needs and possibility of analyzing clinical samples with very low or even undetectable viral load (VL) (Sierra *et al.*, 2011).

Computational methods for predicting HIV tropism are based on sequence and structure of the third variable region (V3 loop) of the viral gp120 protein which appears to be the major determinant of the HIV tropism. Hence, to become a routine part of clinical diagnosis, substituting the phenotype from cheaper and faster genotypic analysis is desired (Sing *et al.*, 2006). As an alternative to costly phenotypic assays, computational methods aim to predict virus tropism based on the sequence and structure of the V3 loop of the virus gp120 protein that are used.

1.3 Objectives

1.3.1 General Objectives

- To predict disease progression of HIV patient and perform phylogenetic analysis of HIV viruses from Kelantan isolate, Malaysia using bioinformatics approaches.

1.3.2 Specific Objectives

- To determine the variation of V3 loop sequence on HIV-1 gp120.
- To predict the co-receptor used by the viruses using Geno2pheno co-receptor (Max Planck Institut Informatik) and Profile-based String Kernel bioinformatics software (Corbeil Research Group).
- To determine the origin of the Kelantan HIV-1 viruses based on envelope glycoprotein gene of the virus by constructing phylogenetic tree.

CHAPTER 2 LITERATURE REVIEW

2.1 History of HIV

In July 1981, the New York Times had reported an outbreak of a rare form of cancer among gay men in New York and California, which first referred as the "gay cancer"; but from the perspective of medicine the cancer is known as Kaposi Sarcoma. About the same time, Emergency Rooms in New York City began to see a rash of seemingly healthy young men presenting with fevers, flu like symptoms, and a pneumonia called Pneumocystis. About a year later, the Centers for Disease Control (CDC) had linked the illness to blood and coins the term Acquired Immune Deficiency Syndrome (AIDS). In that first year over 1600 cases were diagnosed with close to 700 deaths (Adapted from: AIDS About.com The History of HIV. An HIV Timeline Updated June 20, 2007).

As the number of deaths soared, medical experts scrambled to find a cause and more importantly a cure. In 1984, Institute Pasteur of France discovered what they called as the HIV virus, but it wasn't until a year later a US scientist, Dr. Robert Gallo confirmed that HIV was the cause of AIDS. (Adapted from: AIDS About.com The History of HIV. An HIV Timeline Updated June 20, 2007).).

Following the discovery by Dr. Gallo, the first test for HIV was approved in 1985. Over the next several years, medications to combat the virus were developed as well as medicines to prevent infections that burgeon when the immune system is damaged by HIV and AIDS. By the end of 1987, there were 71,000 confirmed cases of AIDS, resulting in over 40,000 deaths. (Adapted from: AIDS About.com The History of HIV. An HIV Timeline Updated June 20, 2007).).

2.2 Etiology of HIV

HIV is the causative agent of AIDS. It is one of the members of the *Lentivirus* genus of *Retroviridae* family and is grouped in two groups namely HIV-1 and HIV-2 (Requejo, 2006). Infections with *Lentiviruses* typically show a chronic course of the disease, with a long period of clinical latency, persistent viral replication and involvement of the central nervous system. The retrovirus genome is composed of two identical copies of single-stranded RNA (ssRNA) molecules with the presence of structural genes *gag, pol,* and *env* (Fanales-Belasio *et al.*, 2010).

HIV-1 and HIV-2 viruses differ in the organization of their genomes, although the basic structure for example the presence of the three structural genes, *gag, pol* and *env* are the same for all retroviruses (Fanales-Belasio et al., 2010) (Figure 2.1). Both the viruses have potential to cause AIDS, although disease of the central nervous system may be more frequently occur in HIV-2 infection (Lucas *et al.*, 1993). In addition, HIV-2 appears less virulent than HIV-1 and infection course takes longer to progress to AIDS (Whittle *et al.*, 1994).

2.3 Classification of HIV

2.3.1 HIV-1

HIV-1 infection, the main cause to AIDS was first recognized in 1981 when a common pattern of symptoms were observed among a small number of homosexual men in the USA (Brennan & Durack, 1981; Gottlieb *et al.*, 1981). HIV is thought to have originated from zoonotic transmissions from Simian immunodeficiency virus (SVI) - infected non-human primates (Gao *et al.*, 1992, 1999; Hirsch *et al.*, 1989).

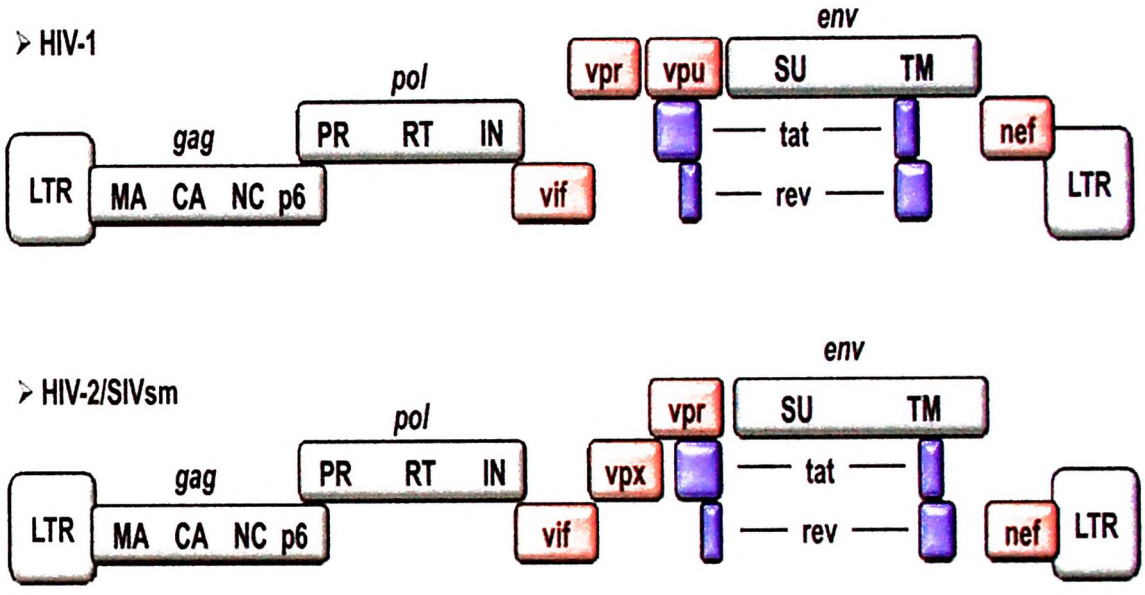

Figure 2.1: Schematic representations of HIV-1 and HIV-2 genomes. Grey boxes represent structural genes; blue boxes indicate regulatory genes; and pink boxes indicate accessory genes. (Modified from Ayinde *et al.*, 2010)

SIVs from chimpanzees are clustered phylogenetically with HIV-1 (Gao *et al.*, 1999); hence, the HIV-1 epidemic is likely to have originated from SIV.

HIV-1 viruses comprises of four distinct lineages, termed groups M, N, O, and P (Figure 2.2). Group M was among the first group to be discovered and represents the pandemic form of HIV-1; it has infected millions of people worldwide and found in all countries worldwide. The M subgroup, which is associated with the global pandemic, is further subdivided into subtypes (A–D, F–H, J and K) and circulating recombinant forms (CRFs) (McCutchan, 2000; Peeters and Sharp, 2000).

These HIV-1 strains were introduced independently into different risk groups, with B subtype being the most frequently observed among intravenous drug users (IDU) and CRF01_AE being more prevalent among commercial sex workers (Ou *et al.*, 1993). However, in recent years, there has been a dramatic shift with CRF01_AE now disseminating faster in all risk groups (Kalish *et al.*, 1994; Weniger *et al.*, 1994), together with the emergence and spread of new CRFs and unique recombinant forms (URFs) involving both CRF01_AE and B subtypes (Ramos *et al.*, 2003; Tee *et al.*, 2005; Tovanabutra *et al.*, 2003).

Group O was discovered in 1990 and is much less prevalent than group M (De Leys *et al.*, 1990; Gurtler *et al.*, 1994). It represents less than 1% of global HIV-1 infections, and is only restricted to Cameroon, Gabon, and their neighboring countries (Mauclere *et al.*, 1997; Peeters *et al.*, 1997). Group N was identified in 1998 (Simon *et al*,. 1998), and it is even less prevalent than group O; up to date, only 13 cases of group N infection have been documented, all in individuals from Cameroon (Vallari *et al.*, 2010).

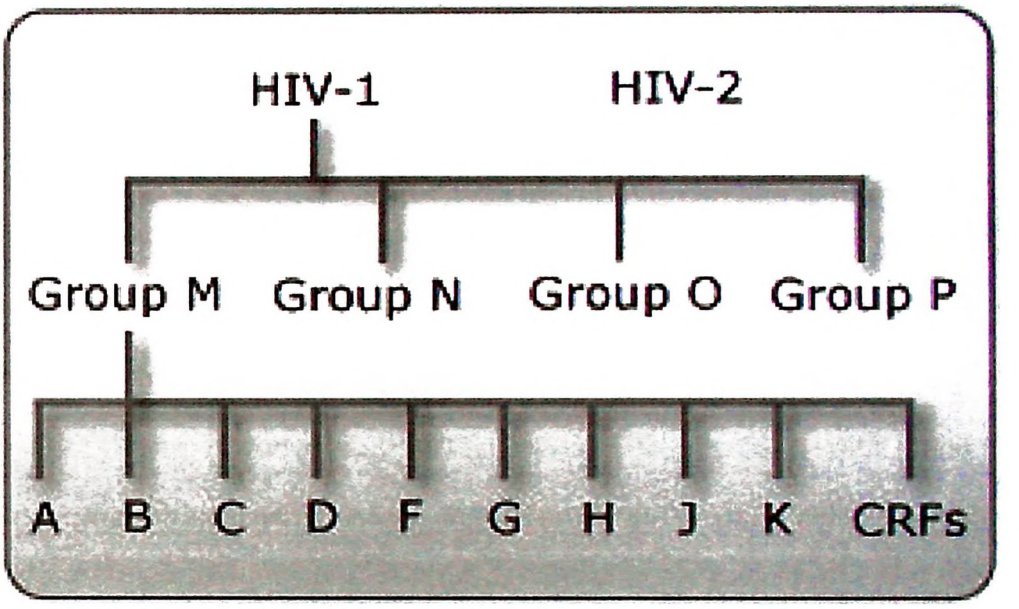

Figure 2.2: The different levels of HIV-1 classification (Source: Averting HIV and AIDS. Updated 2013).

Finally, group P was lastly been discovered in 2009 in a Cameroonian woman living in France (Plantier *et al.*, 2009).

2.3.2 HIV-2

A sooty mangabey which is an Old World monkey origin of HIV-2 was first proposed in 1989 (Hirsch *et al.*, 1989) and subsequently confirmed by demonstrating that humans in West Africa harbored HIV-2 strains that resembled locally circulating sooty mangabeys which is infected with a strain of Simian Immunodeficiency Virus (SIV_{smm}) infections (Gao *et al.*, 1992; Chen *et al.*, 1996). Since the first discovery of SIV_{smm}, HIV-2 has remained largely restricted to West Africa, with its highest prevalence rates recorded in Guinea-Bissau and Senegal (de Silva *et al.*, 2008). HIV-2, together with HIV-1 and SIV, comprise the subgenus `primate lentiviruses'. Genetic variability between HIV-2 strains is comparable to that within HIV-1 groups, with up to 25% divergence in *gag, pol* and *env* (Gao *et al.*, 1994; Schulz *et al.*, 1990; Zagury *et al.*, 1988).

There are at least eight distinct lineages of HIV-2 have been identified so far with each of the lineages appears to represent an independent host transfer. By analogy with HIV-1, these lineages have been termed as groups A–H, with only groups A and B have spread within human population. Group A has been found throughout the Western Africa area (Damond *et al.*, 2001; Peeters *et al.*, 2003), whereas group B predominates in Republic of Cote d'Ivoire, a country in West Africa (Pieniazek *et al.*, 1999; Ishikawa *et al.*, 2001).

All other HIV-2 "groups" were initially identified only in single individuals, suggesting that they represent only incidental infection with very limited or no secondary spread. Of these, groups C, G, and H have been linked to SIVsmm strains from Republic of Cote d'Ivoire, group D of the lineage appears to most closely related to a SIVsmm strain from Liberia, and groups E and F resemble SIVsmm strains from Sierra Leone (Gao *et al.*, 1992; Chen *et al.*, 1996, 1997; Santiago *et al.*, 2005).

2.4 Structure of HIV

HIV and SIV are genetically related members of *Lentivirus* genus of the *Retroviridae* family. Genomes of these viruses are characterized by three main structural genes *gag, pol, env*, and a complex combination of other additional genes (Requejo, 2006) (Figure 2.3). Similarly to other retroviruses, the *gag* gene encodes the structural proteins of the core (p24, p7, and p6) and matrix (p17) while the *env* gene encodes the viral envelope glycoproteins gp120 and gp41, which recognize cell surface receptors. The *pol* gene encodes for enzymes crucial for viral replication, which are the reverse transcriptase (RT) that converts viral RNA into DNA, the integrase that incorporates the viral DNA into host chromosomal DNA, the provirus and the protease that cleaves large *gag* and *pol* protein precursors into their components (Fanales-Belasio *et al.*, 2010). Glycoproteins gp120 and gp41, are linked together to allow gp 120 protrudes from the surface of the virus. Inside this envelope is a matrix (p17), which surrounds a central core of protein, p24. Within this core, there are two copies of single-stranded RNA which is the virus genome. Proteins p7 and p 9, are bound to the RNA and are believed to be involved in regulation of gene expression. RT enzyme is also found in the core which is responsible for converting the viral RNA into proviral DNA (Abbas, 2000).

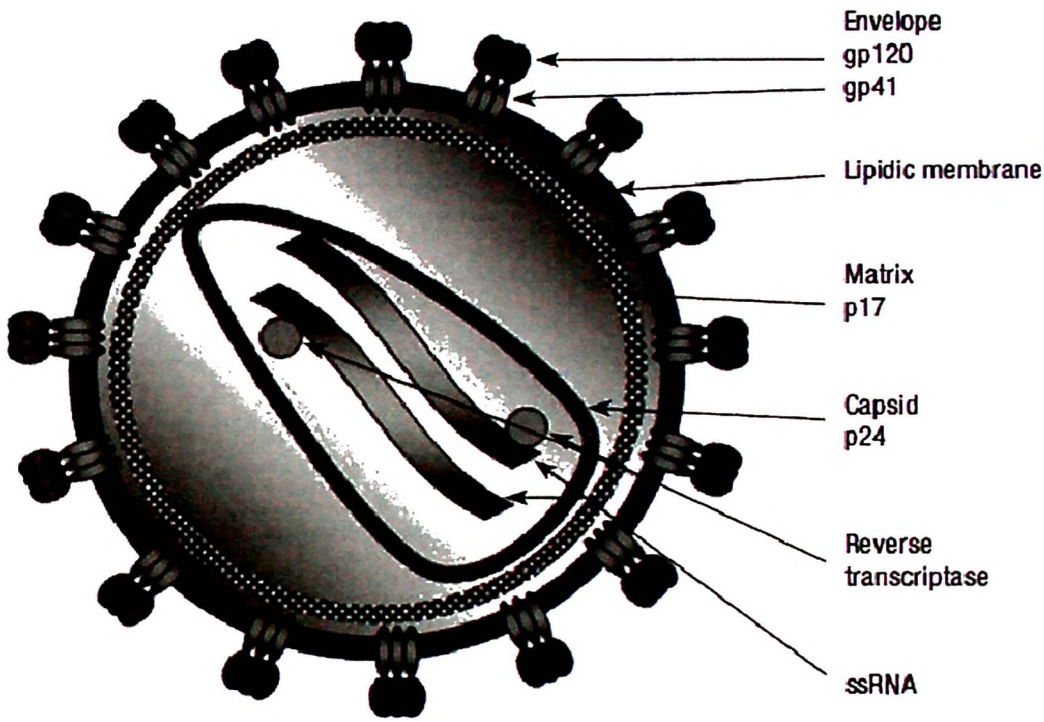

Figure 2.3: Structure of HIV-1 particles. *ssRNA*: single strand RNA. (Modified from Fanales-Belasio et al., 2010)

2.5 Third Hypervariable Region (V3 loop)

The *env* gene of HIV-1 consists of several predicted antigenic determinants, majority of the determinants are found in the five hypervariable regions (V1-V5) of the portion which encode gp120 (Stracich *et al.*, 1986; Modrow *et al.*, 1987). The third hypervariable region of *env*, V3, carries the principle neutralization-specific epitope (Palker *et al.*, 1988; Javaherian *et al.*, 1989) and also a cytotoxic T cell epitope (Hoffenbach *et al.*, 1989; Takahashi *et al.*, 1989).

The third hypervariable region (V3 loop) is located on the surface of most HIV-1 virus (Gorny *et al.*, 2004) (Figure 2.4) (Figure 2.5) and it elicits one of the earliest antiviral antibody responses in HIV-1 infection. The V3 loop of gp120 is 35 amino acids long, highly variable and critical in the determinant of co-receptor usage of virus, although other regions such as V1 or V2 and the conserved region 4 (C4) have also been implicated (Hartley *et al.*, 2005; Pastore *et al.*, 2006).

The V3 loop region is fundamental to various aspects of HIV-1 viral infectivity. Changes in cell tropism is closely linked to the variation in the V3 loop sequence (Cann *et al.*, 1992; Chavda *et al.*,1994; Chesebro *et al.*, 1992, 1996; Harrowe and Cheng-Mayer, 1995; Hwang *et al.*, 1991; Mammano *et al.*, 1995; Stamatatos and Cheng-Mayer, 1993), the ability to induce syncitia (Bhattacharyya *et al.*, 1996; de Jong *et al.*, 1992; Fouchier *et al.*, 1992; Hogervorst *et al.*, 1995; Milich *et al.*,1997), the ability to be neutralized by soluble CD4 (Hwang *et al.*, 1992), and the progression from initial infection to full-blown AIDS (Distler *et al.*, 1995). The V3 region also plays an important role in the process of viral entry (Tian *et al.*, 2002) and is responsible for viral

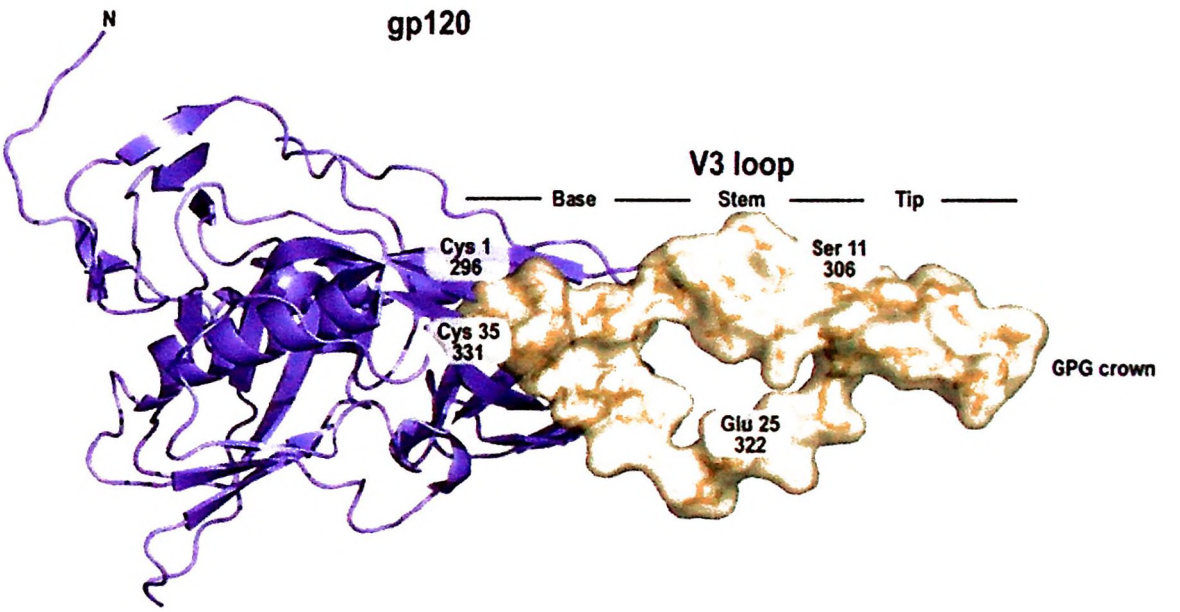

Figure 2.4: Crystal structure of gp120 with the V3 colored in Amber. (Modified from Sander *et al.*, 2007)

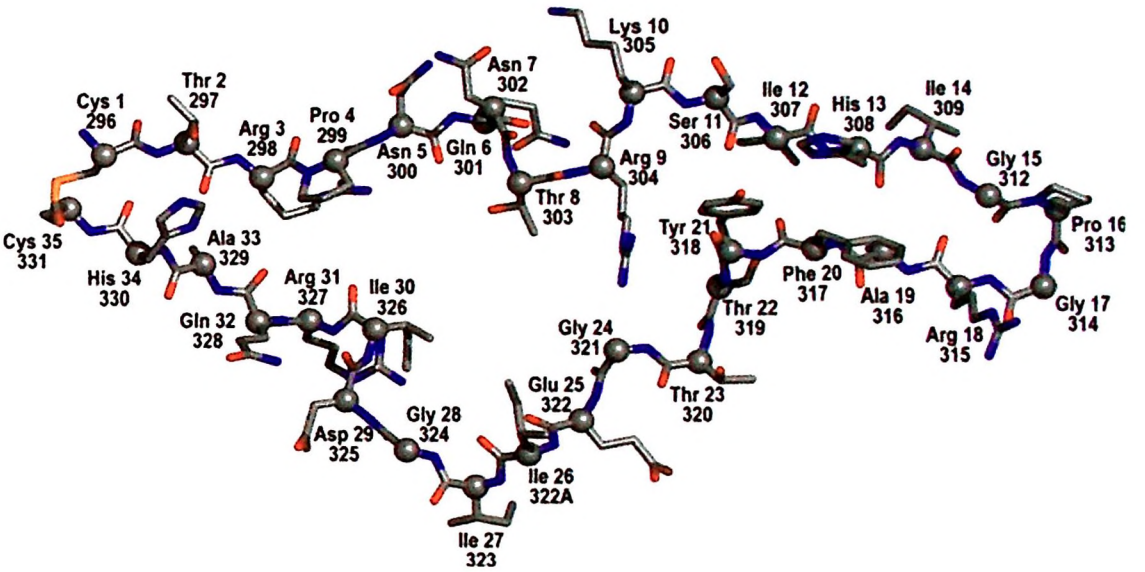

Figure 2.5: V3 of gp120 with residual labels. (Modified from Sander *et al.*, 2007)

infectivity. The tertiary structure of V3 loop is responsible in influencing the co-receptor usage of the virus, which is an important determinant of HIV pathogenesis (Watabe *et al.*, 2005).

The ability of the virus to infect normal CD4+ cells of a living organism depends mostly on the tertiary structure of the receptor-binding site for the virus (Kwong *et al.*, 2002; Gamblin *et al.* 2004). The gp120 glycoprotein experiences several different conformational states. Among the changes is a variation in V3 shape or exposure (Stamatatos, Cheng-Mayer, 1995). R5 strains are generally responsible for the establishment of the initial infection, while the use of the CXCR4 co-receptor is generally seen in more advanced stage of the disease. The emergence of X4 variants has been associated with a more rapid CD4+ cells decline and progression to AIDS (Schuitemaker *et al.*, 1992; Connor *et al.*, 1997).

2. 6 HIV Co-receptor

HIV-1 co-receptors belong to the family of seven transmembrane G-protein coupled chemokine receptor family. In order to trigger the membrane fusion process that leads to viral entry, HIV-1 must first interact with CD4+ cells then with a co-receptor of the host cells. A cluster of residues in the CCR5 amino terminal domain (Nt) participates in gp120-binding and is essential during the process of fusion and entry for both R5 and R5X4 isolates. The binding of HIV-1 to CD4+ cells and co-receptor promotes the entry of the virus into the host target cells.

The two major co-receptors involved are CCR5 and CXCR4 (Deng *et al.*, 1996; Dragic *et al.*, 1996; Feng *et al.*, 1996), with CCR5 being the most commonly used co-receptor

during transmission and early infection of the disease (Michael *et al.*, 1997). Based on the ability of virus to use CCR5 or CXCR4, HIV-1 variants are classified as CCR5 tropic (R5), CXCR4-tropic (X4), or dual-mixed tropic (X4R5) (Berger *et al.*, 1998; Goodenow and Collman, 2006).

2.6.1 CXCR4

Among the co-receptors, the first of these to be identified was CXCR4, or known as fusin previously, which is expressed on T-cells of the host (Feng *et al.*, 1996). Co-expression of CXCR4 and CD4+ on a cell allow T-tropic HIV virus to fuse with the cell and infect it. The gp120 of HIV will interact with both CD4 and CXCR4 to adhere to the cell and to introduce a conformational change in the gp120/gp41 complex that allows the membrane fusion of gp41. CXCR4 is expressed on many T cells, but are seldom expressed on macrophages and it does not allow fusion by macrophage-tropic (M-tropic) HIV virus (Feng *et al.*, 1996). It is interesting to note that the stimulation with some bacterial cell wall products will up-regulate CXCR4 expression on macrophages and allows infection by T-tropic strains of HIV (Moriuchi *et al.*, 1998).

2.6.2 CCR5

CCR5, which is mainly expressed on macrophages and some populations of T cells can also function in concert with CD4+ to allow HIV membrane fusion (Deng *et al.*, 1996; Dragic *et al.*, 1996; Alkhatib *et al.*, 1996). The gp120 of HIV binding to CCR5 is CD4-dependent, as an antibody inhibition of CD4+ can reduce the binding of CD4+ to CCR5 by 87% (Trkola *et al.*, 1996). M-tropic HIV isolates appear to use CCR5 as their main co-receptor for both macrophages and some T cells. Individuals with certain mutations

in CCR5 are resistant to HIV infection (Liu *et al.*, 1996; Samson *et al.*, 1996; Dean *et al.*, 1996).

For HIV-1, the importance of CCR5 co-receptor in virus infection was shown by the discovery on individuals who are lack of CCR5 appears to be more highly resistant to virus infection (Liu *et al.*, 1996; Samson *et al.*, 1996). R5 virus strains are predominantly transmitted by the major virus population in asymptomatic individuals and usually remain present throughout the course of infection (Connor *et al.*, 1997; de Roda Husman *et al.*, 1999; Huang *et al.*, 1996; Li et al., 1999; van'tWout *et al.*, 1998).

2.6.3 Other Co-receptors

CCR5 and CXCR4 often appear to be the two major co-receptors for the entry of HIV virus into cells, but it appears that they are not the only such chemokine co-receptors that are been used. CCR3, a chemokine which is expressed on eosinophils and microglia, is also used by some strains of HIV for infection of the microglia which results in central nervous system pathology (He *et al.*, 1997). It is also possible that binding of HIV gp120 to other such chemokine receptors and be used for the entry of HIV virus into target host cells.

2.7 The HIV Replication Cycle

The HIV replication cycle can be summarized in six steps which includes binding and entry, uncoating, reverse transcription, provirus integration, virus protein synthesis and assembly, and lastly budding of new HIV particles (Fanales-Belasio *et al.*, 2010). The entry of the HIV into human cells is initiated by binding of the viral envelope glycoprotein gp120 to the cellular CD4 receptor of the host cell (Chan *et al.*, 1998). Figure 2.6 illustrates the replication of HIV virus.

2.7.1 Binding, Entry and Uncoating

When HIV first comes into contact with T-lymphocytes (CD4 cells), the spikes of gp120 will 'slot into' the CD4 receptors which function like a lock and key to allow the binding of virus to the host cell. Once the virus is inside the cell, it loses its outer envelope which is termed as the uncoating of the virus and its contents are released into the host cell's cytoplasm including viral RNA and reverse transcriptase (Abbas, 2000).

2.7.2 Reverse Transcription

By using the viral RNA as a template, reverse transcriptase enzyme catalyses the production of a single, complementary strand of DNA from nucleotides of the host cell (Abbas, 2000). DNA which is constructed of units is called nucleotides. Each nucleotide is made up of a base composed of adenine, guanine, thymine, or cytosine, a sugar and a phosphate molecule. The sequence in which the bases occur determines the genetic code of the virus. RNA differs from DNA in that uracil replaces thymine as a base and different type of sugar molecules. Once the single strand of DNA has been produced, it acts as a template in the production of a second strand of DNA. This replication step is also catalyzed by reverse transcriptase enzyme and this results in a double-stranded DNA which is known as the proviral DNA (Abbas, 2000).

2.7.3 Provirus Integration

This phase occurs when the provirus DNA is incorporated into the DNA of the host cell by the viral enzyme, integrase. It is concerning the integration of the viral genetic material into the host cell's own genetic material that makes eradication of the virus, without damage to the host cell, a formidable goal (Hughes, 2002).

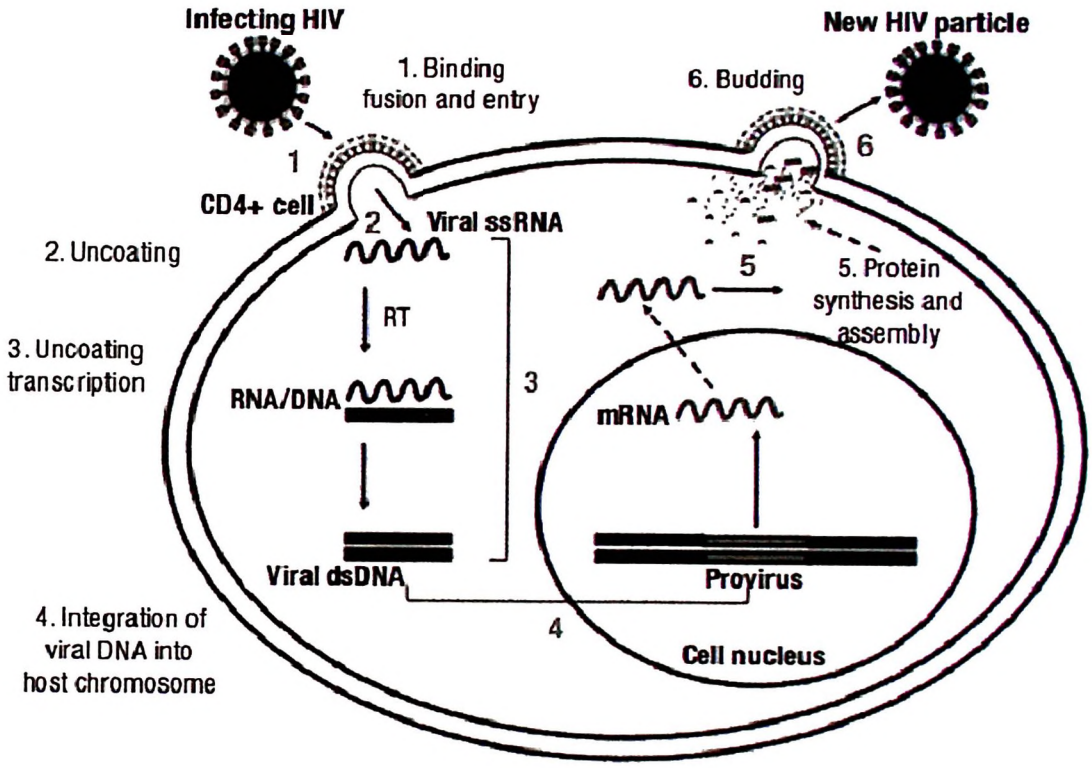

Figure 2.6: HIV replication cycle (Modified from Fanales-Belasio *et al.*, 2010).

Effectively, the virus has now hijacked the host cell's own replication system. As a result, when the cellular DNA is transcribed, same goes to the viral DNA to form an RNA transcript. Further processing of this RNA into messenger RNA (mRNA) and genomic viral RNA occurs (Hughes, 2002).

2.7.4 Virus Protein Synthesis, Assembly and Budding

After the integration of the viral genetic material, the viral mRNA is then translated into viral proteins, which along with the genomic RNA, are assembled into new virus particles. This last stage requires the viral enzyme, protease (Marr, 1998). Finally, the new viral particles are released from the infected cell and continue to infect other cells in the body.

2.8 Genetic Diversity of HIV

One of the major unique characteristics of HIV virus is that it has an extremely high genetic variability. This extensive heterogeneity is a result of two factors. First, there have been multiple introductions of genetically diverse simian viruses into humans (Hahn *et al.*, 2000). Second, once these simian viruses have entered the human population, they will rapidly accumulated more genetic diversity because of the high error rate of reverse transcriptase (Preston *et al.*, 1988) and the fast turnover of virions in HIV-infected individuals (Wei *et al.*, 1995; Ho *et al.*, 1995). On top of this, reverse transcriptase is also known to be highly recombinogenic (Hu *et al.*, 1990), so that radically different genomic combinations may be generated in individuals infected by genetically diverse viruses.

HIV virus has presents a remarkable ability to mutate and adapt to the new conditions of the human environment. A large incidence of errors at transcriptional level results in changes on the genetic bases during the reproductive cycle of HIV (Luciw, 1996; Peeters *et al.*, 1999). The extensive genetic variability of HIV is primarily due to the high error rates of the viral reverse transcriptase, which results in approximately 10 genomic base changes per replication cycle.

Reverse transcriptase plays a major role in provenance of diversity of retroviruses through high iteration of mutations occurs which includes genetic substitutions, deletions, recombination, repetitions, and also insertions. Each of these events may involve more than one nucleotide. Due to these reasons, reverse transcriptase has shown to play a distinguished role for producing HIV sequence diversity in infected individuals (Luciw, 1996; Peeters *et al.*, 1999).

The *env* gene of HIV virus appears likely to be subject to the most extensive genetic variation, although alterations also occur in other genes (Levy, 1994, Sharp *et al.*, 1994; Myers *et al.*, 1994; Korber *et al.*, 1995; Louwangie *et al.*, 1993). HIV-1 virus will continue to undergo continuous genetic variability within individual patients, who usually harbor a swarm of highly related but individually distinguishable viral variants, which are referred to as quasispecies, with a heterogeneity usually not exceeding 2-5% in the *env* gene (Wain-Hobson, 1993; Delwart *et al.*, 1993).

This elevated genomic variability of HIV gives raise to important implications for laboratory diagnosis, treatment, prevention, and also for epidemiological investigation (Luciw, 1996; Peeters *et al.*, 1999). Potentially, each provirus is a new mutant strain and