1.0 INTRODUCTION

1.1 The human and simian immunodeficiency viruses

1.1.1 Origin and similarities, infection and pathogenesis

Simian immunodeficiency virus (SIV) belongs to a group of lentiviruses that infect non-human primates. Though distinct, SIV is closely related to the human immunodeficiency virus type-1 (HIV-1) in its genetic structure. Figure 1.1a shows the evolutionary divergence of various lentivirus strains. The Chimpanzee species *Pan troglodytes troglodytes* is the natural and primary reservoir for HIV-1 (Gao *et al.*, 1999). Therefore, this primate species is thought to be the origin of both pandemic (group M) and non-pandemic (group N) HIV-1 (Keele *et al.*, 2006). HIV-1 group O-like viruses have also been linked to wild gorillas of West central African origin (Van Heuverswyn *et al.*, 2006), confirming that the HIV-1 pandemic may have followed the cross-species transmission route to infect their human host (Nerrienet *et al.*, 2005). SIV has also been isolated from other monkey species including Sooty mangabeys (Fultz *et al.*, 1986), African green monkeys (Fukasawa *et al.*, 1988) and Mandrills (Tsujimoto *et al.*, 1988), and shown to infect macrophages and CD4+ T-lymphocytes. The virus is associated with asymptomatic infection in its natural hosts. The SIV strains infecting these natural hosts are closely related to the HIV-2 (Anderson *et al.*, 2004), also prevalent in West Africa and from which the former is thought to have evolved. In other monkey species like the Asian macaques, infection by SIV (earlier called the Simian T-lymphotropic virus type III/STLV-III), induces acquired immune deficiency similar to human AIDS (Daniel *et al.*, 1985; Letvin *et al.*, 1985). AIDS-like symptoms can be experimentally induced in these animals, (Mann *et al.*, 1998), making them good candidates for HIV-1 infection studies.

In humans, progression of HIV-1 disease is determined by pathophysiological events occurring during infection, and also by age, immunological competence and other co-infections (Kahn & Walker, 1998; Ochieng *et al.*, 2006). While treatment leads to sustained suppression of plasma viral loads and immune recovery, (Berrey *et al.*, 2001), long-term therapy may

reverse treatment benefits (Lwembe *et al.*, 2007). Progression to simian AIDS also varies between animals (Ten Haaft *et al.*, 1998). This variation is influenced by among others, host genetic factors (Sauermann *et al.*, 1997).

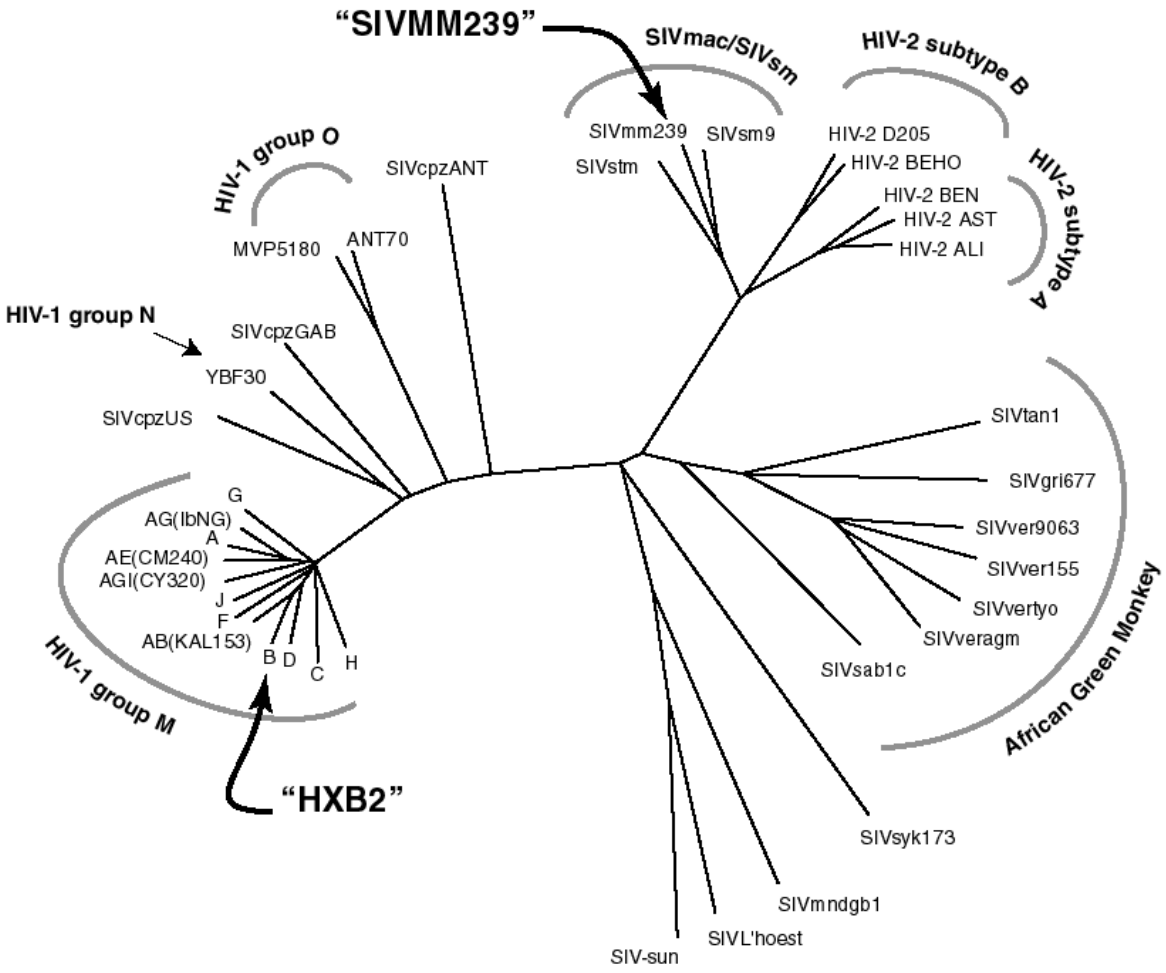

Figure 1.1a: Phylogenetic tree of the primate lentiviruses
Note the large distance between the SIVmac group and the HIV-1 M group and the wide divergence of SIVmac from other SIVs.

1.1.2 Gene structure and functions

SIV and HIV-1 are both lentiviruses that belong to the retrovirus family. The two viruses are closely related in gene structure, function and biology. They have a diploid genome that consists of two plus-stranded RNA copies, approximately 9

kilo base (kb) long. The RNA genome encodes at least 9 different proteins; Gag, Pol, Env, Tat, Rev, Nef, Vif, Vpu and Vpr.

Gag, Gag-Pol and the Protease genes

The *gag* gene product is a 55-kilodalton (kD) precursor protein known as p55. It is expressed from unspliced viral mRNA. The p55 N terminus is myristoylated during translation (Bryant & Ratner, 1990), a process that triggers its association with the cell membrane and which is essential for efficient viral assembly. The membrane-associated Gag polyprotein recruits two copies of the viral genomic RNA along with other viral and cellular proteins that trigger budding of the viral particle from the surface of an infected cell. During viral maturation and after budding, p55 is cleaved by the virally encoded protease into four smaller proteins; (i) MA or matrix (p17), CA or capsid (p24), NC or nucleocapsid (p9) and p6 (Gottlinger *et al.*, 1989). The MA polypeptide stabilizes the virion particle and also constitutes the complex required to transport proviral DNA to the nucleus to facilitate HIV infection (Gallay *et al.*, 1995; Lewis *et al.*, 1992). The CA also includes the p24 protein (p27 for SIV) which forms the viral core. The NC facilitates reverse transcription and is also required for viral packaging (Lapadat-Tapolsky *et al.*, 1993). It recognizes the packaging signal that mediates incorporation of a heterologous RNA into the virions. The p6 polypeptide mediates interactions between p55 Gag and the Vpr protein that are required for the incorporation of Vpr into the assembling virions (Paxton *et al.*, 1993). It also contains a 'late domain' needed for the release of budding virions. The viral protease (Pro), integrase (IN), RNase H, and reverse transcriptase (RT) are expressed as a Gag-Pol fusion protein (Jacks *et al.*, 1988). The Gag-Pol precursor (p160) is a product of a ribosomal frame shift triggered by a specific cis-acting RNA motif (Parkin *et al.*, 1992). Upon encounter with this motif, the ribosomes shift occasionally (5% of the times) to the *pol* reading frame. Although this shift does not interrupt translation, it leads to 20-fold more gag than Gag-Pol precursor products. The Pro is an aspartyl protease dimer that cleaves the Gag and Gag-Pol polyprotein precursors during virion maturation (Ashorn *et al.*,

1990). It accomplishes this role by cleaving the Pol polypeptide away from Gag and further digesting it to separate the protease (p10), RT (p50), RNase H (p15), and integrase (p31) activities. In the process of cleavage, about 50% of the RT protein remains linked to RNase H as a single polypeptide, the p65. Figure 1.1b shows the genome landmarks of SIVmac239 relative to HIV-1.

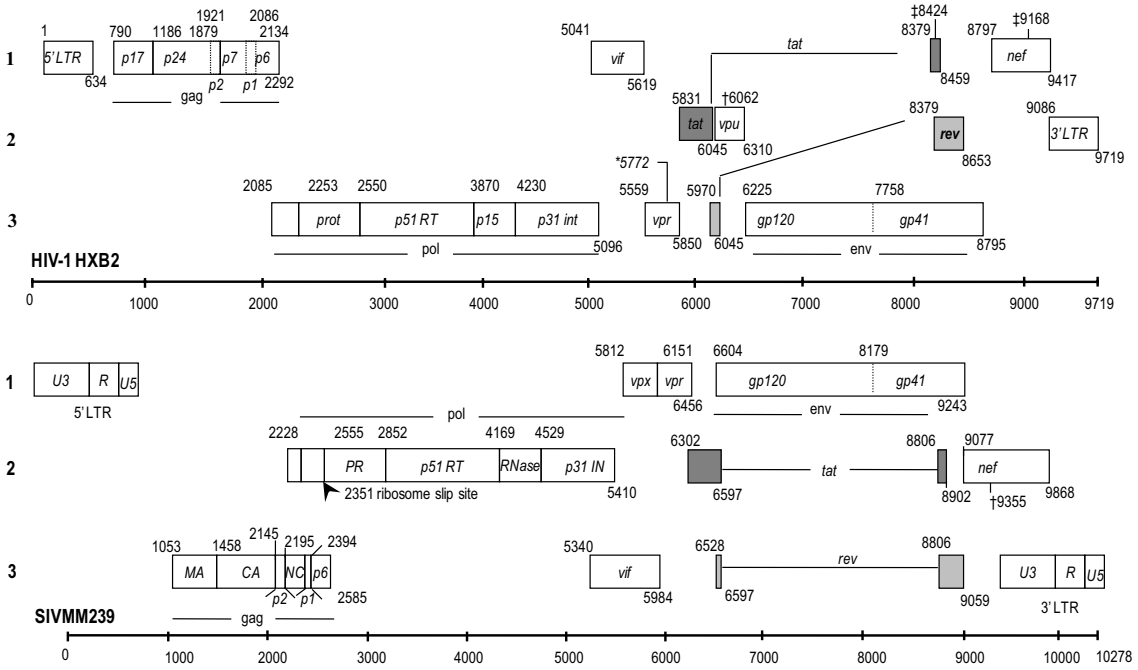

Figure 1.1b: Structural organizations of the SIVmac239 genome

Gene maps for HIV-1 (HXB2 strain, top) and SIVmac239 showing locations of the various genome landmarks. Rectangles show open reading frames (also labelled 1 to 3). The small number in the upper left corner of each rectangle records the position of the 'a' in the *ATG* start codon for the respective gene. The number in the lower right records the last position of the stop codon. For *pol*, the 5' end is the start of the open reading frame, taken to be the first 'T' in the *TTTTTTAG* sequence constituting part of the stem loop and potentiating ribosomal slippage on the RNA leading to a -1 frameshift and the translation of the gag-pol polyprotein. The *tat* and *rev* spliced exons are shown as shaded rectangles. In HXB2, *5772 marks the position of frameshift in the *vpr* gene caused by an extra "T" relative to most other subtype B viruses; †6062 indicates a defective *ACG* start codon in *vpu*; ‡8424, ‡9168 and †9355 mark premature stop codons in *tat* and *nef*. Abbreviations: MA matrix, CA capsid, NC nucleocapsid, PR protease, RT reverse transcriptase, IN integrase. Adapted from; Korber et al., Numbering Positions in HIV Relative to HXB2CG: in 'the database compendium', Human Retroviruses and AIDS, 1998. http://hiv-web.lanl.gov/HTML/reviews/HXB2.html.

Reverse Transcriptase (RT) and Integrase

The retroviral RT is encoded by the *pol* gene, which has both RNA- and DNA-dependent polymerase activities. All *pol* gene products are located within the capsid of free SIV/ HIV-1 virions. Using the dimer of single-stranded viral genomic RNA template, the polymerase synthesises a double-stranded DNA copy of the viral genome during reverse transcription. RNase H degrades the parent RNA template used to make the first DNA strand to allow synthesis of a complementary DNA. This process can be completed within 6 hours of viral entry (Zack *et al.*, 1990). Successful synthesis of proviral DNA requires several cis-acting elements found within the viral RNA. Required to initiate reverse transcription is the TAR element, a small RNA stem-loop structure located at the 5' end of viral RNAs and containing the binding site for Tat (Harrich *et al.*, 1996). The synthesised viral DNA is then integrated into the host genome or may remain in an un-integrated form for prolonged durations. The functional species of the polymerase is a heterodimer of p65 and p50. HIV-1 RT does not possess 3'- to 5'-exonuclease or proofreading activity (Laakso & Sutton, 2006). Consequently, the enzyme introduces several point mutations into each new copy of the viral genome synthesised during replication (O'Neil *et al.*, 2002).

Efficient expression of viral genes requires proviral DNA in its integrated form (Wiskerchen & Muesing, 1995). The virally encoded integrase (In) mediates the insertion of the proviral DNA into the infected host chromosome (Bushman *et al.*, 1990). This process requires three activities of this viral protein; (i) the exonuclease activity trims two nucleotides from each 3' end of the linear viral DNA duplex, (ii) the endonuclease cleaves the host DNA at the site integration and (iii) the ligase generates a single covalent linkage at each end of the proviral DNA. The site of integration is then repaired by cellular enzymes. The choice of integration sites is affected by accessibility of the chromosomal DNA within chromatin rather than by specific DNA sequences (Pryciak & Varmus, 1992).

Envelope (env) and Regulatory Proteins

The HIV envelope is a 160 kD glycoprotein (gp160) expressed from singly spliced mRNA as a precursor protein. It is synthesized in the endoplasmic

reticulum and transported to the Golgi complex where it undergoes glycosylation, a step required for virus infectivity (Capon & Ward, 1991). The gp160 is cleaved by a cellular protease into the gp41 transmembrane glycoprotein and the gp120 surface glycoprotein. The two Env proteins interact through non-covalent linkages. It is the gp120 that binds CD4 receptors on target T-cells to mediate viral internalisation and infection via gp41 (Landau *et al.*, 1988).

Tat is a 14 kD protein with three distinct functional domains that affect localisation and trans-activation of viral genes (Ruben *et al.*, 1989). It is expressed in two long forms; (i) a 72 amino acid HIV-1 Tat is expressed by early fully spliced mRNA, and (ii) the 101 amino acid form from late incompletely spliced viral mRNA. Both forms are located within the nuclei and nucleoli of infected cells and function as transcriptional activators. The trans-activation requires a cis-acting site called TAR, located on the R region of the retroviral long terminal repeat (LTR) (Feng & Holland, 1988; Roy *et al.*, 1990). The HIV-1 *tat* gene products stimulate in trans the expression of all viral genes and ensure generation of full-length transcripts during elongation phase of HIV-1. Rev is a 13-kD sequence-specific RNA binding protein with a binding site in the 'so-called' rev-responsive elements (RRE) of gag-pol and env mRNAs (Zapp & Green, 1989). It is produced from fully spliced mRNAs and acts to increase the amount of gag-pol and env messenger RNA by inducing transition from an early to a late phase of HIV gene expression (Kim *et al.*, 1989).

Accessory Proteins

HIV-1 contains four additional genes: *nef*, *vif*, *vpr* and *vpu* that code for accessory proteins. Instead of the *vpu*, SIV and HIV-2 have the *vpx*, which is a nucleocytoplasmic shuttling protein involved in nuclear import and targeting (Singhal *et al.*, 2006). These accessory proteins represent critical virulence factors (Nielsen *et al.*, 2005). Nef is a 27-kD myristoylated early gene product that is encoded by a single exon extending into the 3' LTR (Kim *et al.*, 1989). It is expressed from a multiply spliced mRNA and is therefore Rev-independent. It is required for optimal replication of both SIV and HIV during primary infection (Fackler *et al.*, 2007; Kestler *et al.*, 1991), as well as for the maintenance of

productive infection in CD4(+) T lymphocytes (Mahlknecht *et al.*, 2000). *In vitro* studies showed that exogenous Nef proteins stimulate HIV-1 LTR via *NF-kappaB* activation (Varin *et al.*, 2003), confirming that these gene products could exacerbate disease progression and increase virion infectivity (Miller *et al.*, 1994). Nef down-modulates CD4 and MHC-I and II surface expression, and polymorphisms in this gene can lead to different disease outcomes in both HIV and SIV infection (Munch *et al.*, 2005; Schindler *et al.*, 2007).

Vpr, Vpu, and Vif are products of incomplete mRNA splicing normally expressed from singly spliced mRNAs during the late Rev-dependent infection phase. The Vpr protein is incorporated into viral particles via interactions with the c-terminal of p55 Gag (Cohen *et al.*, 1990). It is needed for infection of non-dividing cells (Heinzinger *et al.*, 1994) and also acts as a nucleocytoplasmic transport factor connecting the viral genome to the nuclear pore (Vodicka *et al.*, 1998). Vpu is a 16-kD integral membrane phosphoprotein polypeptide localized in the internal cell membranes (Sato *et al.*, 1990). It functions to down-modulate CD4 and enhance virion release (Schubert *et al.*, 1996). The vpx is also incorporated into budding virions at the plasma membrane. Vif is a 23-kD polypeptide essential for the replication of HIV in peripheral blood lymphocytes, macrophages, and other target cells (Strebel *et al.*, 1987).

1.1.3 Receptor-virus interactions

Cellular tropism

HIV-1 infection is the primary cause of the acquired immunodeficiency syndrome (AIDS), a slow but progressive, degenerative and wasting disease that attacks and compromises the host immune system. Receptor and coreceptor usage by both HIV-1 and SIV plays an important role in virus tropism and pathogenesis. The primary targets for HIV-1 and SIV infections are the CD4+ T-lymphocytes and monocytes/macrophages (Dittmar *et al.*, 1997), although they can also infect astrocytes and microglial cells. Infection is initiated by binding of the virion gp120 surface glycoprotein to the CD4 receptor on T-cells (Nielsen *et al.*, 2005).