CHAPTER 1

INTRODUCTION
General background of HIV

Acquired immunodeficiency syndrome (AIDS) is one of the major death causes in the world. Approximately 39.5 million (range: 34.1 to 47.1 million) individuals were living with HIV and 2.9 million (2.5-3.5 million) died due to HIV/AIDS in year 2006 (UNAIDS global estimates) (Figure 1.1). Out of these deaths, 380,000 were children below 15 years of age. In some of the African countries (e.g. Kenya, Malawi, Zimbabwe, Swaziland) prevalence of HIV is >25%, which has significantly reduced the life expectancy. Since majority of individuals infected with HIV fall in the age group of 15 to 50, which is prime time of working life, HIV/AIDS has devastated several millions of families.

First cases of HIV infection in New York and Los Angeles were described in form of Kaposi’s sarcoma (KS), a benign cancer and by March 1981 a more aggressive form of the Kaposi’s sarcoma was reported among 8 young gay men (Hymes, Cheung et al. 1981) (An, Hymes et al. 1981). At the same time rare infections of *Pneumocystis Carinii* Pneumonia (PCP), a rare lung infection was also reported among young gay men in these cities (Hymes, Blum et al. 1981, Hymes and Warshaw 1981). Later on in December 1981 and 1982, it was clear that disease affected other population groups as well. These included injecting drug users (IDUs) (Masur, Michels et al. 1981), haemophiliacs and Haitian people (Deschamps, Fitzgerald et al. 2000). The acronym AIDS was suggested at a meeting in Washington, D.C. in July (Time 2003). By August, this name was being used in newspapers and scientific journals (Marx 1983). However, the word AIDS (Acquired Immune Deficiency Syndrome) was properly defined by CDC in September and the case definition for AIDS was implemented, henceforth (1988, 1988, 1993). Isolation of etiological agent from these patients was first reported in 1983 by Barre-Sinoussi et al., at the Pasteur Institute (Barre-Sinoussi, Chermann et al. 1983) as a reverse transcriptase containing virus (retrovirus) from lymph node of a man with persistent lymphadenopathy. They termed the virus as Lymphadenopathy Associated Virus (LAV). In 1984, Gallo et al., from NIH, USA reported isolation of retrovirus from an AIDS patient called _Human T cell Lymphotropic Virus III_(HTLV-III) (Gallo, 1984).
Few other investigators also reported isolation of AIDS associated Retroviruses (ARV). Subsequently, the three prototype viruses (LAV, HTLV-III and ARV) were recognized as members of same group of retroviruses. In 1986, the International Committee on Taxonomy of Viruses (ICTV) recommended giving the AIDS virus a separate name, the Human Immunodeficiency Virus (HIV) (Coffin, Haase et al. 1986). Subsequently, two types of HIV (type 1 and type 2) were identified. HIV-1 and HIV-2 evolved independently from each other, and originate from respectively a chimpanzee virus (SIVcpz) and an SIV variant in Sooty mangabeys. Both viruses may have crossed the monkey-human species barrier at several independent occasions (Fletcher et al., 2004; Weller and Davis, 2002).

HIV-1 belongs to the lentivirus genus, a subfamily of the Retroviridae. Lentiviruses generally cause a slow, degenerative disease consisting of 3 phases: acute infection, latency and disease progression (Pool et al., 2000). The retroviral life cycle is characterized by reverse transcription of the single stranded RNA genome into DNA by the viral reverse transcriptase enzyme. A schematic representation of the HIV-1 replication cycle is depicted in Fig. 1.7, HIV-1 attaches to the target cell via its Envelope protein gp120, which binds to CD4 and a co-receptor. These interactions induce fusion of the viral and cellular membranes, followed by release of the viral core into the cytoplasm. In the core, viral RNA is reverse transcribed into DNA, which subsequently is inserted into the cellular genome by the viral integrase enzyme. The inserted viral DNA serves as a template for the production of viral mRNAs and new viral RNA genomes. Viral structural proteins and two RNA genome copies assemble into a virion particle that leaves the cell by budding. A detailed description of the HIV-1 replication cycle is given in reference (Weber, Desai, and Darbyshire, 2005).

By infecting cells of the immune system, HIV-1 causes a gradual collapse of this defense system, leading to opportunistic infections, various forms of cancer, and finally death (Ackermann and de, 2002). HIV-1 can infect cells of the immune system that bear the appropriate receptors: CD4 and an additional co-receptor (Ackermann and de, 2002; McMichael and Hanke, 2003). Cells that are susceptible to HIV-1 infection are CD4+ T cells, monocytes, macrophages, dendritic cells (DC), thymocytes and microglia (Frank and Pope, 2002; Lu et al., 2004; Lu et al., 2003;
Song et al., 2006; Trumpfheller et al., 2006; Wille-Reece et al., 2005). CCR5 is the major co-receptor used by HIV-1, and in some patients the virus evolves to use CXCR4 as co-receptor (Ackermann and de, 2002; McMichael and Hanke, 2003). The CCR5 co-receptor is primarily expressed on the memory T cell subset and macrophages (Desrosiers, 2004; Lee et al., 1999; Pantaleo and Koup, 2004; Pope, 2003; Weissman et al., 2000). Within 3 weeks after transmission, the virus becomes established in a lymphatic tissue reservoir, which is the principal site of virus production, storage, persistence and pathology (Bentley et al., 2004; Carballo-Dieguez et al., 2000; Corey et al., 2004; Galvin and Cohen, 2004; Madan, Keller, and Herold, 2006; Severy et al., 2005; Sewankambo et al., 1997) when HIV-1 starts using CXCR4, the virus expands its target cell repertoire to naive T cells, which coincides with faster disease progression (Dhawan and Mayer, 2006; Hicks et al., 1985). Sexual transmission of HIV-1 is complicated by a mucosal barrier that has to be crossed, a process that is not yet entirely understood (Kreiss J et al., 1992; Roddy et al., 1998; Van Damme et al., 2002).

**HIV-1 Classification:**

HIV is a member of genus Lentivirus, belonging to the family *Retroviridae*. There are two types of retrovirus (Figure 1.2): 1) oncogenic or transforming retroviruses, which lead to neoplasms; and 2) cytopathic or lentiviruses, of which HIV is an example. Lentiviruses are transmitted as single-stranded, positive-sense, enveloped RNA viruses. Upon entry into the target cell, the viral RNA genome is converted (reverse transcribed) into double-stranded DNA by a virally encoded reverse transcriptase (RT) that is transported along with the viral genome in the virus particle. Two types of HIV have been characterized: HIV-1 and HIV-2, on the basis of serologic properties and sequence analysis of molecularly cloned viral genomes. HIV-1 is the virus that was initially discovered and termed both LAV and HTLV-III. It is more virulent, more infective, and is the cause of the majority of HIV infections globally.
The strains of HIV-1 can be classified into four groups: the "major" group M, the "outlier" group O and two new groups, N and P. The M group is most prevalent and consists of nine subtypes (A-D, F-H, J and K). These group M viruses frequently recombine and these inter subtype recombinants are classified as either circulating recombinant forms (CRF) or unique recombinant forms (URF) (Figure 1.3).
Fig 1.2: Phylogenetic tree of 19 "prototypic" members of the retrovirus family. Modified from (Griffiths, Venables et al. 1997)

Fig 1.3: Schematic representation of circulating HIV-1 groups and subtypes (Source: Fields Virology, 2005)
Different HIV-1 subtypes predominate in different geographic regions; subtype A in Africa and Eastern Europe, subtype B in America, Europe and Australia; subtype C in India, China and South Africa. Apart from these subtypes there are HIV-1 strains that have mosaic genomes. They have been reported in geographic areas where more than one HIV-1 subtypes are circulating. These strains are product of recombination between two different HIV-1 subtypes that infect a single cell. The recombination occurs due to template switching during reverse transcription by reverse transcriptase enzyme. Recombination between different groups (Peeters, Toure-Kane et al. 2003) different subtypes (Lal, Chakrabarti et al. 2005) within subtype (Rousseau, Learn et al. 2007) as well as among recombinant strains of HIV-1 (Yang, Li et al. 2005) has been reported. To date, 19 CRFs have been identified and several have played a significant role in the establishment of certain regional epidemics including CRF01_AE and CRF02_AG. (Casado, Thomson et al. 2005) Approximately 8% of HIV-1 genomic sequences available in Los Alamos HIV sequence database (http://www.hiv.lanl.gov) display mosaic genome structures (Kothe, Li et al. 2006). When HIV-1 strains with similar mosaic structure are obtained from three or more epidemiologically unlinked individuals, they are referred as Circulating Recombinant Forms.

Origin of HIV-1

The origin of HIV is a topic of much controversy. However, the human AIDS virus type 1 (HIV-1) and type 2 (HIV-2) represents cross species (zoonotic) infections (Gao, Bailes et al. 1999). Both HIV viruses are similar to SIV commonly found in African monkeys. Although primate reservoir of HIV-2 has been clearly identified as the sooty mangabey (Cerco cebusatsys) (Gao, Bailes et al. 1999), the origin of HIV-1 remains uncertain, but evidence points to HIV-1 being a descendent of the common chimpanzee (Pan troglodytes) (Sharp and Hahn 2010).
Epidemiology: Current Global Distribution of HIV

According to the latest report by UNAID, new HIV infections are declining. Following the discovery of HIV in 1983, HIV-1 has caused a worldwide pandemic resulting in more than 25 million deaths and the current estimate of 33 million people living with HIV-1 infections (Global distribution of HIV-1 infections shown in Figure 1.1; Statistics provided by the Joint United Nations Program on HIV/AIDS (UNAIDS) and the World Health Organization (WHO)). According to the 2009 UNAIDS/WHO report on the global AIDS epidemic, there were an estimated 2.6 million [2.3 million–2.8 million] people who became newly infected with HIV. This is nearly one fifth (19%) fewer than the 3.1 million [2.9 million–3.4 million] people newly infected in 1999, and more than one fifth (21%) fewer than the estimated 3.2 million [3.0 million–3.5 million] in 1997, the year in which annual new infections peaked. In 33 countries, the HIV incidence has fallen by more than 25% between 2001 and 2009; 22 of these countries are in sub-Saharan Africa. In sub-Saharan Africa, where the majority of new HIV infections continue to occur, an estimated 1.8 million [1.6 million–2.0 million] people became infected in 2009; considerably lower than the estimated 2.2 million [1.9 million–2.4 million] people in sub-Saharan Africa newly infected with HIV in 2001. This trend reflects a combination of factors, including the impact of HIV prevention efforts and the natural course of HIV epidemics. In Western, Central, and Eastern Europe, Central Asia, and North America, the rates of annual new HIV infections have been stable for at least the past five years. However, evidence is increasing of a resurgence of HIV in several high income countries among men who have sex with men. In Eastern Europe and Central Asia, high rates of HIV transmission continue to occur in networks of people who inject drugs and their sexual partners.

The number of average AIDS related deaths is steadily decreasing from the peak of 2.1 million (1.9 million – 2.3 million) in 2004 to an estimated 1.8 million (1.6 million – 2.1 million) in 2009. The decline reflects increased availability of antiretroviral therapy, as well as care and support, to people living with AIDS.
particularly in middle and low income countries; it is also a result of decreasing incidence starting in late 1990s.

**HIV Transmission**

HIV is transmitted by sexual contact, contaminated blood and from mother to child during pregnancy, birth or breast-feeding.

**HIV Pathogenesis**

Infection with HIV-1 is associated with a progressive decrease of the CD4+ T cell count and an increase in viral load (Lane and Fauci 1985, Phillips and Lundgren 2006). The stage of infection can be determined by measuring the patient's CD4+ T cell count, and the level of HIV in the blood.

HIV infection has four basic stages: incubation period, acute infection, latency stage and AIDS (Figure 1.4). The initial incubation period upon infection is asymptomatic and usually lasts between two and four weeks. The second stage, acute infection, lasts an average of 28 days and can include symptoms such as fever, lymph adenopathy (swollen lymph nodes), pharyngitis (sore throat), rash, myalgia (muscle pain), malaise, and mouth and esophageal sores. The latency stage, which occurs third, shows few or no symptoms and can last anywhere from two weeks to twenty years and beyond. AIDS, the fourth and final stage of HIV infection shows as symptoms of various opportunistic infections (Rowland-Jones 2003).
Acute phase:

Within 2-3 weeks after acquisition of HIV-1 infection, virus becomes well established in the lymphoid tissue. Plasma viral RNA levels increase exponentially with a doubling time of 10 to 20 hrs (Fiebig, Wright et al. 2003). Plasma viral RNA level is highest at the start (primary stage) and end of infection (AIDS), whereas low and relatively steady level is maintained in between for a variable number of years (Figure 1.4) (Veazey, DeMaria et al. 1998). This steady state of viraemia (viral set point) is achieved within 6 to 12 months of infection and may depend on various host (immune response and genetic factors such as HLA) (Mattapallil, Douek et al. 2005) and viral factors as yet not fully understood (Brenchley, Schacker et al. 2004). There is massive infection and loss of CD4+ T cells predominantly in lymphoid tissues of gastro-intestinal tract (Mattapallil, Smit-McBride et al. 1998).

Clinically latent phase:

Acute primary HIV-1 infection is followed by a long period of clinical latency (usually 7 to 10 years). Although patients in this stage of infection do not show clinical disease and plasma viral RNA level is often low, virus multiplication continues in the lymphoid tissue. Ho et al. showed that there is daily production and clearance of 0.05
x 10^9 to 2 x 10^9 HIV-1 virions and daily turnover of CD4+ T lymphocytes ranges between 0.2 x 10^9 to 5.4 x 10^9 cells (Ho, Neumann et al. 1995). This turnover in the CD4 cells leads to steady decline in peripheral CD4+ T cell count leading to immune suppression. The rate of CD4+ T cell decline may vary leading to slow, average or rapid disease progression. There are three general patterns of disease progression that occur following infection: 1) typical progressor; 2) rapid progressor; and 3) long-term non-progressor. About 10% of the HIV-infected individuals progress to AIDS within 2 to 3 years after infection and are referred as rapid progressors. About 5-10% individuals do not show clinical disease even after 10 years of infection (Liu, de Vlas et al. 2010); (Lefrere, Morand-Joubert et al. 1997). These individuals show less than 5% decline in CD4+ T cell count annually.

With plasma viral RNA level undetectable or very low without anti-retroviral therapy. These individuals are called Long-Term Non-Progressors (LTNPs). Majority of infected persons may lead to AIDS stage within 7 to 10 years. Clinical AIDS is established when an infected individual's circulating CD4+ T cell level drops below 200 cells/μl or if they exhibit one of the AIDS defining clinical conditions as defined by the Centre for Disease Control and Prevention (CDC). The patients acquire various opportunistic infections and malignancies such as, Candidiasis of respiratory tract & lungs, Coccidioidomycosis (disseminated or extra pulmonary), Cryptococcosis (extra pulmonary), Cytomegalovirus disease (other than liver, spleen, or nodes), Herpes simplex infection (with chronic ulcer and greater than 1 month's duration), Histoplasmosis (disseminated or extra pulmonary), Isosporiasis (chronic intestinal with greater than 1 month's duration), Kaposi's sarcoma, Burkitt's Lymphoma, Lymphoma of brain, *Mycobacterium avium* complex, *Mycobacterium tuberculosis* (pulmonary & extra pulmonary) and wasting syndrome. Patients in this stage show highest risk of death.

**Structure of HIV**

HIV is roughly spherical with a diameter of about 1/10,000 of a millimeter with a bilayer membrane or envelope surrounding the cone shaped nucleocapsid (Figure 1.5). These lipid bilayers are taken from the membrane of the host cell during
budding of the newly formed virus particles. The nucleocapsid is composed of two copies of positive single-stranded RNA, about 9.2 kb long that have positive polarity with respect to translation. Nine genes are coded by the viral RNA. The RNA is additionally enclosed by a conical capsid composed of 2,000 copies of the viral protein p24. The single-stranded RNA is tightly bound to nucleocapsid proteins, p7, and enzymes needed for the development of the virion such as reverse transcriptase (RT), proteases, ribonuclease and integrase. A matrix composed of the viral protein p17 surrounds the capsid ensuring the integrity of the virion particle.

The capsid, in turn, is surrounded by the viral envelope derived from host cell along with 70-80 copies of a complex HIV-1 protein called Env. Within the viral envelope are encoded two glycoproteins, gp120 and gp41. Structurally, Env consists of a cap made of three molecules of gp120, and a stem consisting of three gp41 molecules that anchor the structure into the viral envelope. This glycoprotein complex enables the virus to attach to and fuse with target cells to initiate the infectious cycle.
Genome organization of HIV-1

HIV has several major genes coding for structural proteins that are found in all retroviruses, and several nonstructural ("accessory") genes that are unique to HIV. The HIV-1 proviral RNA genome consists of about 9200 nucleotides, which encodes for nine genes (gag, pol, and env, tat, rev, nef, vif, vpr, vpu, and sometimes a tenth vev, which is a fusion of tat, env and rev), encoding 19 proteins (Figure 1.6).

The gag gene and the gag and pol genes together are translated into large polyproteins which are then cleaved by a virus-encoded protease that is part of the pol polyprotein. Gag polyprotein is cleaved to into four proteins that are found in the mature virus: MA (matrix), CA (capsid), NC (nucleocapsid), p6. Pol polyprotein is cleaved to three proteins: PR (protease), RT (reverse transcriptase), IN (integrate).

Env gene is translated to a polyprotein (Gp160) which is then cleaved by a host cell protease (called furin) that is found in the Golgi body. It is not cleaved by the virus-encoded protease. Gp160 is cleaved to: SU (Gp120) and TM (Gp41). The latter retains the trans-membrane part of Gp160 while Gp120 remains attached to Gp41 via non-covalent bonds.

The remaining six genes, tat, rev, nef, vif, vpr, and vpu (or vpx in the case of HIV-2), are regulatory genes for proteins that control the ability of HIV to infect cells, produce new copies of virus (replicate), or cause disease (Table 1.1). HIV-1 uses three reading frames for transcription of mRNAs and HIV-1 proteins are synthesized after differential splicing of primary mRNA transcripts.
HIV-1 life cycle:
HIV-1 life cycle is interplay of viral and host proteins and begins with the attachment of virion at the cell surface. The life cycle can be split into an early and late phase. The late phase initiates following integration of the proviral DNA into the host genome and extends till virus budding (Figure 1.7). Target cells for HIV-1 are mostly resting or activated CD4 T lymphocytes and macrophage.

<table>
<thead>
<tr>
<th>Serial Number</th>
<th>HIV gene Products</th>
<th>Role and Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>gag</td>
<td>p55 myristoylated protein precursor, which is processed to p17 (Matrix), p24 (Capsid), p7 (Nucleocapsid), and p6 proteins, by the viral protease. Gag associates with the plasma membrane, where virus assembly takes place.</td>
</tr>
<tr>
<td>2</td>
<td>pol</td>
<td>The genomic region encoding the viral enzymes protease, reverse transcriptase, and integrase. These enzymes are produced as a Gag-Pol precursor polyprotein, which is processed by the viral protease.</td>
</tr>
<tr>
<td>3</td>
<td>env</td>
<td>Env is synthesized as precursor (gp160), which is processed to give surface glycoprotein gp120 and the trans-membrane glycoprotein gp41. The mature gp120-gp41 proteins are bound by non-covalent interactions and are associated as a trimer on the cell surface. Gp120 contains the binding site for the CD4 receptor and co-receptors for HIV-1. Essential for virus entry and attachment to host cell.</td>
</tr>
<tr>
<td>4</td>
<td>tat</td>
<td>Trans-activator of HIV gene expression. It is a regulatory gene, localized in the nucleus and acts by binding to the TAR RNA element and activating transcription initiation and elongation from the LTR promoter, preventing the 5' LTR AATAAA polyadenylation signal from causing premature termination of transcription and polyadenylation</td>
</tr>
<tr>
<td></td>
<td>Protein</td>
<td>Description</td>
</tr>
<tr>
<td>---</td>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>5</td>
<td>Rev</td>
<td>The second necessary regulatory factor for HIV expression. A 19-Kd phosphor protein, localized primarily in the nucleolus/nucleus, Rev acts by binding to RRE and promoting the nuclear export, stabilization, and utilization of the viral mRNAs containing RRE.</td>
</tr>
<tr>
<td>6</td>
<td>Vif</td>
<td>A 23-Kd cytoplasmic protein that promotes the infectivity but not the production of viral particles. Vif prevents the action of the cellular APOBEC-3G protein, which deaminates DNA: RNA heteroduplexes in the cytoplasm.</td>
</tr>
<tr>
<td>7</td>
<td>Vpr</td>
<td>Vpr is a 96-amino acid (14-Kd) protein localized in nucleus and is incorporated into the virion. Proposed functions for Vpr include the targeting the nuclear import of pre-integration complexes, cell growth arrest, trans-activation of cellular genes, and induction of cellular differentiation.</td>
</tr>
<tr>
<td>8</td>
<td>Vpu</td>
<td>Vpu is unique to HIV-1, SIVcpz (the closest SIV relative of HIV-1), SIV-GSN, SIV-MUS, SIV-MON and SIV-DEN. Vpu is a 16-kd (81-amino acid) type I integral membrane protein with at least two different biological functions: degradation of CD4 in the ER, and enhancement of virion release from the plasma membrane of HIV-1-infected cells.</td>
</tr>
<tr>
<td>9</td>
<td>Nef</td>
<td>A multifunctional 27-Kd myristoylated protein present predominantly in the cytoplasm and associated with the plasma membrane via the myristoyl residue linked to the conserved second amino acid (Gly). Nef down regulates CD4, the primary viral receptor, and MHC class I molecules, and these functions map to different parts of the protein. Nef interacts with components of host cell signal transduction and clathrin-dependent protein sorting pathways. It increases viral infectivity.</td>
</tr>
</tbody>
</table>
Virus attachment to the host cell and entry:

Attachment of HIV is mediated by interaction between the extracellular domain (gp120) of Env glycoprotein and the CD4 antigen present on the surface of susceptible cells followed by interaction with co-receptors (members of the seven membrane-spanning CC or CXC families of chemokine receptors). The two major co-receptors for HIV infection are CXCR4 and CCR5. Once the gp120 of virus envelope binds to CD4 receptor and co-receptor of target cell, it induces a conformational change in gp41 that leads to fusion of viral and cell membranes. The membrane fusion event enables the viral core to gain entry into the host cytoplasm.

HIV-1 uses CCR5 as preferential co-receptor for entry into the target cell at the time of transmission but in some cases HIV-1 isolated from patients during late stage of infection uses CXCR4 as co-receptor. Therefore based on phenotype, HIV-1
is isolates are classified as CCR5 tropic (R5 phenotype) and CXCR4 tropic (X4 phenotype). HIV-1 isolates that use both i.e. CCR5 & CXCR4 co-receptors (R5/X4 phenotype) as well as co-receptors other than CCR5 and CXCR4 have been also reported.

**Viral uncoating and viral DNA synthesis by reverse transcription:**

Uncoating of viral capsid occurs within the cytoplasm of infected cell. It is believed that uncoating is promoted in response to multiple successive changes in the cellular environment, sequential contact with different cellular factors, and through the molecular rearrangements that accompany reverse transcription, thus triggering progressive or stepwise conformational changes and disassembly. After uncoating, there is formation of virus reverse transcription complexes (RTCs) and pre-integration complexes (PICs). RTCs are simply defined as HIV-1 complexes that undergo reverse transcription, during which they convert their single-stranded positive RNA viral genome into double-stranded DNA. The RTC genomes are thus either RNA or RNA-DNA intermediates of reverse transcription. In contrast, PICs no longer contain any RNA but only the double-stranded DNA. PICs are per definition integration-competent HIV-1 complexes and can integrate efficiently into a target DNA *in vitro*. Reverse transcription involves firstly the formation of the minus strand strong-stop DNA, a strand transfer event, and the synthesis of the minus strand DNA with concomitant degradation of the RNA template. In the HIV-1 genome, two poly purine tracts (PPT), the central PPT (Cppt) and 3' PPT, resist degradation by Rnase H and serve as primers for synthesis of plus-strand DNA. Reverse transcription proceeds with synthesis of plus-strand DNA, involves a second strand transfer event, and terminates at a central termination sequence (CTS) in the centre of the genome. The initiation of plus strand synthesis at the Cppt, as well as the 3' PPT, leads to a discrete plus-strand displacement of 100 nucleotides in the centre of the genome. The final product of HIV-1 reverse transcription is therefore a linear double-stranded (ds) DNA with a central DNA Flap. Upon DNA Flap formation and completion of reverse transcription, the viral complex becomes a PIC, competent for import into the nucleus and integration within the host cell chromatin (Arhel 2010).
Integration of viral DNA into host cell chromosome:

The PIC moves towards nucleus using microfilaments and microtubules and enters into the nucleus using nuclear import pathway. After translocation into the nucleus, the integrase cleaves the 3' termini of the viral double-stranded DNA to generate two nucleotide 5' overhangs at each end. Subsequently the integrase triggers trans-esterification reaction in which the 3' hydroxyl group attacks phosphodiester bonds of chromosomal DNA and joins viral DNA to host DNA. The viral DNA is randomly integrated at many chromosomal locations. The integrated form of the virus is called provirus. The non-integrated linear DNA is circularized, which reduces signal for apoptosis by reducing number of linear DNA molecules.

Transcription of viral RNA:

The provirus remains latent or is actively transcribed depending on chromatin structure around the integration site and metabolic status of the host cell. The virus may remain latent due to integration into areas of repressed heterochromatin or due to absence of factors such as nuclear factor κB (NF-κB) and Nuclear Factor of Activated T cells (NFAT), which act as transcriptional enhancers. In an activated cell, NF-κB and NFAT bind to enhancer sequence of LTR and promote viral transcription. Host cell RNA Polymerase II binds to transcription initiation site and begins transcription but fails to elongate efficiently in absence of viral Tat protein. Tat protein binds to TAR (first 45 nucleotides of viral mRNA, which is target sequence for viral trans-activation) and prevent premature termination of transcription.

Processing of viral RNA transcripts, nuclear export and expression of viral proteins:

Viral RNA transcripts are completely spliced incompletely spliced or remains un-spliced. The viral Rev protein plays an important role in controlling the splicing mechanism and transport of viral RNA transcripts to cytoplasm. In the cytoplasm these RNA molecules are translated into viral proteins. Multiply spliced mRNA molecules encode Nef, Tat and Rev whereas incompletely spliced RNA transcripts
encode Env, Vif, Vpr and Vpu proteins. The un-spliced RNA transcripts encode Gag (p55) and Gag-Pol (p160) precursor proteins. The Gag-Pol precursor is produced by ribosomal frame shifting near the 3' end of gag.

**Assembly, budding and maturation of HIV virus particles:**

The final step of the viral cycle, assembly of new HIV-1 virions, begins at the plasma membrane of the host cell. The Env polyprotein (gp160) goes through the endoplasmic reticulum and is transported to the Golgi complex where it is cleaved by protease and processed into the two HIV envelope glycoproteins gp41 and gp120. These are transported to the plasma membrane of the host cell where gp41 anchors the gp120 to the membrane of the infected cell. The enzyme protease plays a vital role at this stage of the HIV life cycle by chopping up long strands of protein into smaller pieces, which are used to construct mature viral cores. Assembly of HIV-1 is directed by the Gag protein and it contains all of the determinants necessary for assembly as Gag alone is capable of forming non-infectious viral like particles. In addition, Gag recruits other HIV-1 proteins, especially Env and viral RNA into nascent virions. Gag is synthesized as a polyprotein precursor, Pr55gag, and is cleaved into its component subunits by HIV-1 protease. Approximately 2000 Gag proteins, 200 Gag-Pol proteins, two un-spliced viral RNA and other proteins (Vif, Vpr and Nef) assemble below the cell membrane (Wilk, Gross et al. 2001). These assembled viral components form immature virion which buds out of the host cell using cellular ESCRT (Endosomal Sorting Complex required for Transport) that mediates outward vesiculation (Marsh and Thali 2003). In this process, host cell plasma membrane embedded with gp120 and gp41 proteins form HIV-1 envelope and encloses the nucleocapsid. After budding the viral protease enzyme cleaves Gag and Gag-Pol precursor proteins. The Gag precursor protein is cleaved into p24, p17 and other subunits whereas Gag-Pol precursor protein is cleaved into reverse transcriptase, protease and integrase. This results in formation of mature infectious virion (Wilk, Gross et al. 2001).