CHAPTER 1
AN INTRODUCTION TO HIV AND ITS PATHOGENESIS
1.1. DISCOVERY OF THE AIDS VIRUS (HIV)

Acquired Immunodeficiency Syndrome (AIDS) is a pathological condition caused by Human immunodeficiency virus (HIV) where the host immune system is completely subdued by the virus. Clinically, a condition where the CD4+ T cell count becomes less than or equal to 200 cells/μL of blood is defined as AIDS. The discovery of the AIDS virus happened way back in 1983 when Dr. Francoise Barre-Sinoussie and Prof. Luc Montagnier (Pasteur Institute, Paris) reported the identification of a retrovirus which was recovered from a Caucasian patient with symptoms that generally precede AIDS and was named Lymphadenopathy Associated Virus (LAV) (Barre-Sinoussi et al., 1983). Concurrently, Dr. Robert C. Gallo and his colleagues (National Cancer Institute, USA) identified a retrovirus as the causative agent of AIDS which was named Human T cell Leukaemia Virus-III (HTLV-III) (Gallo et al., 1983). Later in the same year, Dr. Gallo and Prof. Montagnier announced in a joint press conference that LAV and HTLV-III are identical and is likely the cause of AIDS. In 1984, another group led by Prof. J. A. Levy also isolated the virus from AIDS patients which was named AIDS-associated retrovirus (ARV) (Levy et al., 1984). By the end of 1984, the molecular clone of LAV was characterized which was found to have a genome size of 9.1-9.2 kilobases, and was longer than other known retroviruses of that time (Alizon et al., 1984). By 1985, at least one case of HIV began to be reported from every part of the world. In 1985, it was proclaimed that LAV, HTLV-III and ARV are variants of the same AIDS virus (Ratner et al., 1985). An international virus taxonomy consortium in 1986, unanimously decided to name the AIDS virus as Human Immunodeficiency Virus (HIV) (Coffin et al., 1986). In honour of the discovery of HIV, Dr. Francoise Barre-Sinoussie and Prof. Luc Montagnier was awarded the Nobel Prize for physiology or medicine in 2009.

1.2. THE ORIGIN AND NATURAL HISTORY OF HIV

Acquired Immunodeficiency Syndrome (AIDS) in humans is caused primarily by two lentiviruses, namely Human immunodeficiency Viruses 1 and 2 (HIV-1 and
HIV-2). Although AIDS emerged as an epidemic in the early eighties, the oldest verified case dates back to 1959 in a stored serum biopsy specimen (Zhu et al., 1998) and in 1960 in a lymph node biopsy specimen (Worobey et al., 2008) both from Kinshasa, Zaire (previously known as Congo). Both HIV-1 and HIV-2 originated as an outcome of multiple cross-species (zoonotic) transmissions of simian immunodeficiency viruses (SIVs). African primates are the natural targets of SIVs. Many of the zoonotic transmission in these primates resulted in formation of viruses with only limited capability of spreading in humans. HIV-1 emerged as a consequence of four zoonotic transmissions of SIV involving chimpanzees and gorillas (Figure 1.1).

![Image of primate species and transmission pathways]

**Figure 1.1. The Origin of HIV** - Adapted from Cold Spring Harb Perspect Med. 2011 Sep; 1(1)

The HIV-1 group M, one of the major determinants of AIDS pandemic, involved a zoonotic transmission of SIVcpz found in chimpanzee *Pan troglodytes troglodytes*
in West-Central Africa which is the primary reservoir of HIV-1 (Gao et al., 1999; Keele et al., 2006). HIV-2 emerged as a result of eight zoonotic transmissions (Figure 1.1) and the primate reservoir is identified as the sooty mangabey (*Cercocebus atys*) in West Africa (Gao et al., 1999; Peeters et al., 2013).

1.3. **Epidemiology of AIDS**

HIV/AIDS is considered as a global pandemic as it is one of the most serious disease affecting humans worldwide. Since the beginning of the epidemic, almost 71 million people have been infected with HIV virus and about 34 million people have died of HIV. According to the estimates by the UNAIDS, 36.9 million [34.3 million – 41.4 million] people globally were living with HIV by the end of 2014. Among this, 15.8 million people are accessing antiretroviral therapy for AIDS. By the end of 2014, there were 2 million [1.9 million – 2.2 million] new infections and 1.2 million [980,000 – 1.6 million] people died from AIDS-related illnesses. Sub-Saharan Africa still remains the most severely affected region, with nearly 1 in every 20 adults (4.8%) living with HIV and accounting for nearly 70% of the people living with HIV worldwide. Figure 1.2 depicts an overview of the global HIV epidemiology.

![Global HIV Epidemiology - Adults and children estimated to be living with HIV, 2014](image_url)

**Total: 36.9 million** [34.3 million – 41.4 million]

*Figure 1.2. Global HIV Epidemiology - Adults and children estimated to be living with HIV, 2014 - Adapted from UNAIDS Report, 2015*
However, in spite of this huge number of infections worldwide, the new HIV infections have fallen by 35% globally since 2000. Worldwide, about 2 million [1.9 million – 2.2 million] people became newly infected with HIV in 2014, which came down from 3.1 million [3.0 million – 3.3 million] in 2000. There was also a drastic decline of new HIV infections among children by about 58% since 2000. According to UNAIDS report 2015, the total number of newly infected children were 2,20,000 [1,90,000 – 2,60,000] in 2014 which came down from 5,20,000 [4,70,000 – 5,80,000] in 2000. The AIDS-related deaths have also fallen by 42% since its peak in 2004. The number of AIDS-related deaths in 2014 have come down to 1.2 million [9,80,000 – 1.6 million] from 2 million [1.7 million – 2.7 million] in 2005 (Figure 1.3). The tuberculosis-related deaths in people living with HIV have also fallen by 32% since 2004.

Global summary of the AIDS epidemic 2014

| Number of people living with HIV in 2014 | Total 36.9 million [34.3 million – 41.4 million] |
| Adults 34.3 million [31.8 million – 38.5 million] |
| Women 17.4 million [16.1 million – 20.0 million] |
| Children (<15 years) 2.6 million [2.4 million – 2.8 million] |

| People newly infected with HIV in 2014 | Total 2.0 million [1.9 million – 2.2 million] |
| Adults 1.8 million [1.7 million – 2.0 million] |
| Children (<15 years) 220,000 [190,000 – 260,000] |

| AIDS deaths in 2014 | Total 1.2 million [980,000 – 1.6 million] |
| Adults 1.0 million [890,000 – 1.3 million] |
| Children (<15 years) 150,000 [140,000 – 170,000] |

Figure 1.3. Global Summary of AIDS Epidemic, 2014 - Adapted from UNAIDS Report, 2015

1.4. PREVALENCE OF AIDS IN INDIA

According to the UNAIDS Gap report 2014, India has 2.1 million people living with HIV, the third-largest population of people infected with the virus around the globe, after South Africa and Nigeria (Figure 1.4). India accounts for about 4 out of 10 people living with HIV in the Asia-Pacific region. An estimate of 0.35% of the country is accounted positive for HIV, the most affected state in the country being Nagaland. About 36% of adults living with the virus have access to
antiretroviral treatment. Between 2005 and 2013, India recorded a 38 per cent decline in AIDS-related deaths as there was a major scale up of access to HIV treatment during this period. In 2014, there were 810,339 patients on government-sponsored antiretroviral therapy, including children and transgender people, according to Indian government estimates. According to National AIDS Control Organization (NACO), India, the transgender population is emerging as a risk group with high vulnerability and high levels of HIV in India.

### HIV prevalence in India

India has the third largest number of people living with HIV in the world, according to the World Health Organisation.

<table>
<thead>
<tr>
<th>Country</th>
<th>HIV infected 2013, in millions of people</th>
<th>Deaths due to HIV/AIDS 2013, in thousands of people</th>
<th>Percentage HIV Positive in India</th>
</tr>
</thead>
<tbody>
<tr>
<td>South Africa</td>
<td>1</td>
<td>0.5</td>
<td>0.75 - 1.0%</td>
</tr>
<tr>
<td>Nigeria</td>
<td>0.5</td>
<td>0.3</td>
<td>0.5 - 0.74</td>
</tr>
<tr>
<td>India</td>
<td>0.4</td>
<td>0.2</td>
<td>0.25 - 0.49</td>
</tr>
<tr>
<td>Mozambique</td>
<td>0.3</td>
<td>0.2</td>
<td>&lt; 0.25</td>
</tr>
<tr>
<td>Uganda</td>
<td>0.2</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>0.1</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>0.1</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Tanzania</td>
<td>0.1</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Zambia</td>
<td>0.1</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Malawi</td>
<td>0.1</td>
<td>0.1</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1.4. Prevalence of HIV in India** - Adapted from UNAIDS Gap Report 2014

#### 1.5. TAXONOMY AND DISTRIBUTION OF HIV

HIV is a primate lentivirus and belongs to the genus Lentivirus, sub-family Orthovirinae and family Retroviridae. Retroviridae constitutes a family of enveloped viruses which contain two copies of the viral genome in the form of a non-covalently linked RNA dimer (Clever and Parslow, 1997). Retroviruses rely on the enzyme reverse transcriptase to transcribe their RNA genome into DNA. Lentiviruses ('lentus' meaning slow in Latin) are species specific, slow growing
viruses with a long asymptomatic incubation period. HIV is an enveloped, positive sense, single strand RNA virus and possesses 2 copies of RNA per virion. HIV is classified majorly into 2 groups: **HIV-1** and **HIV-2**. HIV-1 was first isolated in 1983 (Barre-Sinoussi et al., 1983; Gallo et al., 1983) as mentioned earlier. HIV-2 was isolated later in 1986 and was named as a second HIV isolate as it differed from HIV-1 in sequence by more than 55% and were also antigenically distinct (Clavel et al., 1986; Coffin et al., 1986). The HIV-1 group originated from a chimpanzee (SIVcpz) whereas the HIV-2 group comprises of virus originated from sooty mangabey monkeys (SIVsmm). The genome of HIV-1 and HIV-2 are quite similar differing in the presence of a single gene. HIV-2 possesses \textit{vpx} gene instead of \textit{vpu} gene of HIV-1 (Tristem et al., 1992). However, the envelope glycoprotein sequences of HIV-1 and HIV-2 have major differences (Clavel et al., 1987). Among both the groups, HIV-1 is more predominant worldwide and pathogenic as compared to HIV-2. HIV-2 is restricted mostly to West Africa with few cases being reported from Western Europe, United States and South America. The pathogenicity of HIV-2 is less as it is not easily transmitted like HIV-1. Also the incubation period of disease development is relatively longer for HIV-2 when compared to HIV-1.

HIV-1 and HIV-2 have been classified further into various strains based on their whole genome sequence. HIV-1 is sub-classified into 4 strains - **M**, the major/main group; **O**, the outlier group; **N**, the non-M, non-O group; and **P**, putative or pending group. These strains differ majorly in their envelope gene sequence. Group M, that originated from SIVcpz is responsible for more than 90% of the total HIV infections worldwide.

**HIV-1** Group M is further sub-grouped into nine genetically distinct **subtypes/clades**: A to D, F to H, J and K (Peeters et al., 2003; Robertson et al., 2000). These subtypes differ from each other in envelope and gag region (Korber et al., 2000). Subtype A is predominant in Central Africa, subtype B in North America and Europe, subtype C in South Africa and India, subtype D in Central
Africa, subtype F in Brazil, Romania, Cameroon and Democratic republic of Congo and the others majorly in Africa (Hemelaar et al., 2011) (Table 1.1).

<table>
<thead>
<tr>
<th><strong>HIV-1 Subtypes</strong></th>
<th><strong>Global prevalence</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtype A</td>
<td>Central and East Africa as well as East European countries that were formerly part of the Soviet Union</td>
</tr>
<tr>
<td>Subtype B</td>
<td>West and Central Europe, the Americas, Australia, South America, and several southeast Asian countries (Thailand, and Japan), as well as northern Africa and the Middle East</td>
</tr>
<tr>
<td>Subtype C</td>
<td>Sub-Saharan Africa, India, and Brazil</td>
</tr>
<tr>
<td>Subtype D</td>
<td>Central Africa, North Africa and the Middle East</td>
</tr>
<tr>
<td>Subtype F</td>
<td>Brazil, Romania, Cameroon, Democratic republic of Congo, South and southeast Asia</td>
</tr>
<tr>
<td>Subtype G</td>
<td>West and Central Africa</td>
</tr>
<tr>
<td>Subtypes H, J, and K</td>
<td>Africa and the Middle East</td>
</tr>
</tbody>
</table>

Table 1.1. Global Prevalence of HIV-1 Subtypes (Buonaguro et al., 2007; Hemelaar et al., 2011)

When co-circulation of multiple subtypes occurs in a geographical region, it leads to the emergence of recombinant viruses which are called circulating recombinant forms (CRFs). CRFs are generated as a result of recombination during replication (Hemelaar et al., 2011). Subtype B had been the most prevalent since long which has now been dominated by subtype C as it accounts for more than 51% of infection worldwide (Tebit and Arts, 2011). Subtype C is rapidly progressing in South America, China and Eastern Africa apart from India and South Africa and trends indicate that it might spread more rapidly than subtype B across the world (Arien et al., 2007; Tebit and Arts, 2011). Similar to HIV-1, HIV-2 is sub-classified into eight strains/groups: A to H (Damond et al., 2004). Group A is predominant in Senegal and Guinea Bisssau, group B in Ivory Coast and C to F in Sierra Leone and Liberia.
1.6. HIV-1 TROPISM

The natural targets for HIV-1 are cells expressing CD4 receptor. Since CD4+ T helper cells express CD4 receptor abundantly, these cells are the main targets of HIV-1. Other cell types like monocytes and macrophages that express lower levels of CD4 can also be infected by HIV-1. However, CD4+ T helper cells exhibit the highest levels of virus production. Macrophages, in general, function as virus reservoir, although some strains of HIV can kill them as well. The virus replication is usually at a low rate in monocytes and dendritic cells as they express low levels of CD4 (Lee et al., 1999).

Along with CD4 receptor, a major determinant of HIV-1 tropism (phenotype) lies at the level of virus entry into target cells, governed by the expression of the chemokine receptors CCR5 and CXCR4 which function as coreceptors for HIV-1 entry into CD4+ T cells/macrophages (Berger et al., 1998). Virus isolates that use CCR5 but not CXCR4 are termed \textbf{R5 tropic viruses} and those isolates that use CXCR4 but not CCR5 are designated \textbf{X4 tropic viruses}. Virus isolates which are able to use both co-receptors with comparable efficiency are known as \textbf{R5X4 tropic (dual tropic) viruses} (Berger et al., 1998; Lee et al., 1999). Figure 1.5 schematically shows the different viral tropisms.

\textbf{Figure 1.5. HIV-1 Tropism} (Source www.viivhxresource.com)
CCR5Δ32 is a 32 base pair genomic deletion in the CCR5 gene (CCR5Δ32) that leads to expression of a truncated gene product which provides resistance to HIV-1 infection in individuals homozygous for this mutation (Cornu et al., 2015). However, CXCR4-using viral strains can still lead to infection (Proudfoot, 2002). Heterozygote carriers of this truncated gene form are resistant to HIV-1 infection relative to wild types and when infected exhibit reduced viral loads and a 2-3 year slower progression to AIDS in comparison to wild types (Dean et al., 1996). Allogeneic hematopoietic stem cell (HSC) transplantation with CCR5Δ32 donor cells have been reported to confer HIV-1 resistance to the recipient (Cornu et al., 2015).

The co-receptors are differentially expressed on the surface of CD4+ T cells (Bleul et al., 1997). During the early stages of HIV infection, viral isolates tend to use CCR5 for viral entry, while as the infection progresses, the virus can evolve to enter the cells using CXCR4 (Wilen et al., 2012). In the peripheral blood, the CD4+ memory T cell subset expresses higher levels of only CCR5, whereas CXCR4 is expressed at relatively high levels on both memory and naive CD4+ T cells (Bleul et al., 1997; Lee et al., 1999; Nicholson et al., 2001). Hence, in vivo, the memory CD4+ T cells are the predominant cells that are infected (Douek et al., 2002; Sleasman et al., 1996). In cell culture, HIV-1 infects activated cells more efficiently than quiescent cells (Korin and Zack, 1998), where central and effector memory cells are the primary targets (Pfaff et al., 2010).

1.7. HIV-1 PATHOGENESIS

The HIV-1 infection begins with the entry of a single virus particle into its target cell. This virion then replicates in the host cell to generate viremia and persistent infection in the entire lymphoid tissue in the body. As mentioned earlier, HIV-1 preferentially infects T cells with high levels of CD4 and in particular, the memory T cell subsets (Swanstrom and Coffin, 2012). HIV infection is associated with the progressive loss of CD4+ T cells through their depletion by cell death as well as due to their decreased production (Douek et al., 2002). The CD4+ T cell
count normally ranges between 800-1200 cells per cubic millimeter (or μl) of blood. Clinically, a condition where this number drops down to less than 200 cells per cubic millimeter, is referred to as AIDS (Levy, 1988). The depletion of CD4+ T cells can occur by multiple processes, including cell death by apoptosis, necrosis and pyroptosis. The continuous exhaustion of CD4+ T cells leads to gradual failure of immune system leading to immunodeficiency. Along with the onset of immunodeficiency, the virus evolves to infect new cell types as well as develops the ability to enter cells with low levels of CD4 on their surface and this potentiates the ability to infect macrophages (Swanstrom and Coffin, 2012). This ultimately leads to the emergence of opportunistic infections and malignancies.

There are three major clinical stages of HIV-1 infection - acute phase (or the early period) where high virus production occurs, chronic phase, a period when virus is maintained at a relatively low threshold regulated mostly by immune system, and finally the symptomatic period when virus re-emerges and symptoms of development of disease (AIDS) becomes evident. Generally, the virus population is homogeneous during early and symptomatic period, while heterogeneous virus population is present during chronic phase (An and Winkler, 2010). A depiction of the different stages of HIV-1 infection from preliminary infection, progression through chronic phase and subsequent progression to AIDS is shown in Figure 1.6. The corresponding HIV RNA copies per millilitre of plasma, with respect to each stage are also shown alongside.
During the acute phase of HIV-1 infection, the virus tends to infect large number of CD4+ T cells and replicates rapidly to build up a high viral load in the body. Hence the risk of HIV-1 transmission is very high during this stage. Acute phase begins around one week after the entry of virus and may last for about 1-6 months before antiviral antibodies become detectable in the blood of the infected individual. However, from around 2-14 days the viral RNA can be detected (Fiebig et al., 2005) and peak viremia may be observed around 3 weeks post infection (Little et al., 1999). Clinical symptoms including a high temperature (fever), sore throat, a blotchy rash on the body, tiredness, swollen glands, joint and muscle pain. Flu-like illness and often maculopapular rash are generally shown by 50-90% of infected people (Daar et al., 2001).

The second phase or chronic phase of infection begins when plasma B cells start secreting neutralizing antibodies against HIV-1 (Kahn and Walker, 1998). As a result, a reduction in virus levels occurs in the plasma and CD4+ T cells also return back to their normal level. There is also a slowing down of viral replication
but a constant level may be maintained. For all these reasons, this period is sometimes called as 'asymptomatic HIV infection' and it can last for 5-10 years before the onset of symptomatic stage, AIDS.

The symptomatic or the late stage of HIV-1 infection is AIDS which is detected by a CD4+ T cell count lesser than 200 cells/µl, higher viral load and simultaneous reduction in CD8+ T antiviral response. AIDS is a disorder characterized by a compromised immune system and involves severe symptoms such as wasting, pneumonia, and other life threatening conditions where the affected individual becomes incapable of evading majority of the microbes/pathogens leading to various opportunistic infections and malignancies, which ultimately results in death.

1.8. STRUCTURE AND GENOMIC ORGANIZATION OF HIV-1

HIV-1 virion comprises of a viral envelope and associated matrix enclosing a capsid, composed of the viral protein p24, typical of lentiviruses. The capsid further encloses two copies of non-covalently linked, unspliced, positive-sense single-stranded RNA genome and several enzymes (Lu et al., 2011). In retroviruses like HIV-1, both RNA molecules are utilized for strand transfer-mediated recombination during the process of reverse transcription (Hu and Temin, 1990a) although only one DNA allele is generated. Retroviruses are therefore considered 'pseudodiploid'. The retroviral recombination rate is approximately 2% per kilobase per replication cycle (Hu and Temin, 1990b).

HIV-1 is an enveloped virion which is spherical in structure and is around 100-120 nm in diameter (Kuznetsov et al., 2003) in which the RNA genome of the virus and other viral proteins are embedded. The RNA genome is of 9.8 kb and encodes for 9 open reading frames (ORFs) and 15 different viral proteins. The viral particle is composed of structural, regulatory and accessory proteins. A schematic representation of the HIV-1 virion structure is given in Figure 1.7.
The viral genome codes for the different structural and non-structural proteins using the *gag* (group specific antigen), *pol* (polymerase) and *env* (envelope glycoprotein) genes giving rise to the major polyproteins Gag, Pol and Env. The *gag* gene codes for HIV-1 p55 Gag precursor polyprotein (Pr55\text{Gag}). Viral proteins, matrix (MA or p17), capsid (CA or p24) and nucleocapsid (NC or p7), are formed from this Gag precursor polyprotein. The matrix forms the inner shell of the virus and is present just below the viral envelope. Capsid is present beneath the matrix and surrounds the viral genome. Nucleocapsid interacts with viral RNA inside the capsid. Viral protein, protease (PR) is responsible for generating these viral proteins from Pr55\text{Gag} precursor polyprotein (Gomez and Hope, 2005).

The *pol* gene gives rise to a large polyprotein precursor (Pr160\text{GagPol}) which is cleaved into reverse transcriptase (RT), protease (PR) and integrase (IN). The *env* gene forms an Env precursor polyprotein known as gp160 which is heavily glycosylated and is processed by a cellular protease to be cleaved into surface (SU or gp120) and transmembrane glycoprotein (TM or gp41) (McCune et al., 1988). gp120 is an extracellular protein which primarily functions in the recognition of
HIV receptors CD4 and CCR5/CXCR4 on the target cells (Deng et al., 1996). gp41 on the other hand is an integral membrane protein which contains a transmembrane anchor domain that anchors Env into the lipid membrane of the host cell (Levy, 1993).

In addition to these structural proteins, HIV-1 possess two regulatory proteins - Tat and Rev as well as four accessory or auxiliary proteins - Nef, Vif, Vpr and Vpu (Frankel and Young, 1998). Tat plays critical role in HIV-1 LTR mediated transcription and Rev plays major role in export of viral RNAs from nucleus to cytoplasm. Each of the accessory proteins is involved in critical functions during viral replication to enhance the efficiency of virus production and infectivity. Vif and Nef remain in close association with the core of the virus (Camaur and Trono, 1996; Pandori et al., 1996). Vpr is found outside the viral core (Lu et al., 1993). The presence of these proteins in virion substantiates the importance of them in the early steps of viral life cycle. Presence of some lipid domains from host cell membrane has also been demonstrated to be incorporated in virion during budding and release of virions from the surface of host membrane (Aloia et al., 1993). The HIV-1 genomic organization is schematically depicted in Figure 1.8.

![Figure 1.8. Schematic representation of the genomic organization of HIV-1. Adapted from (Ayinde et al., 2010)](image)

1.9. HIV-1 LIFE CYCLE

HIV-1 persistently infects humans by subverting the host innate and adaptive immune systems, despite the fact that it encodes only 15 mature proteins. Viral replication at the cellular level proceeds through a series of steps. The life cycle
begins when a virus productively engages cell surface receptors to bind to its host and ends when nascent particles mature into infectious virions and buds off from the host cell.

These stages of HIV-1 viral life cycle can be broadly classified into **early** and **late phase** of viral replication. Early phase constitutes attachment of virus on host membrane, viral entry, uncoating of virus, reverse transcription, formation of pre-integration complex (PIC), nuclear import of PIC and integration of viral DNA into host genome. Late phase includes transcription, translation, assembly, budding, release and maturation of virus. The early gene products, *tat*, *rev*, and *nef* are first transcribed, followed later by the rest of the HIV genome. Assembly and budding of progeny virions takes place at the plasma membrane of the infected cell. An overview of the different stages of viral life cycle is shown in Figure 1.9.

**Figure 1.9. Schematic overview of HIV-1 Life Cycle - Adapted from (Engelman and Cherepanov, 2012)**

A detailed description of the different stages of viral life cycle is given below:
1.9.1 Attachment and fusion

The HIV-1 envelope spikes contain trimers of non-covalently linked heterodimers, which constitutes the surface glycoprotein gp120 and the transmembrane glycoprotein gp41. When triggered, these viral spikes initiate a cascade of conformational changes that culminates in fusion between the viral and host cell membranes and subsequent release of the viral core into the cytoplasm of the host cell. The primary targets of HIV-1 are CD4+ T cells and macrophages. Initially there occurs an interaction between gp120 and the surface receptor CD4 which induces the formation of a bridging sheet between the inner and outer domains of the gp120 monomer. This exposes the binding site for CCR5/CXCR4 (Chen et al., 2005; Kwong et al., 1998). Involvement of this co-receptor in the process leads to insertion of the fusion peptide into the cell membrane. The fusion peptide is located at the amino terminus region of gp41. This fusion event further triggers significant rearrangements of the N and C terminal heptad repeat sequences within gp41, resulting in the formation of a six-helix hairpin structure and ultimately fusion of the viral and host cell membranes (Buzon et al., 2010; Weissenhorn et al., 1997) (Figure 1.10).

![Figure 1.10. Binding and fusion of gp120/gp41 with the host cell - Adapted from (Didigu and Doms, 2012)](image)

1.9.2. Uncoating and Reverse Transcription

The post-entry events of the HIV-1 viral life cycle begin with the uncoating of the viral particle. Uncoating refers to the partial dissolution of the capsid core of the virus so as to enable the release of the viral genome enclosed within it, in order to
initiate the process of reverse transcription (Engelman and Cherepanov, 2012). Reverse transcription is the process of synthesizing of double stranded DNA from viral RNA genome. Uncoating is considered as a transition between reverse transcription and formation of pre-integration complex (PIC) (Arhel, 2010).

The loss of integrity of the intact capsid core is a cytoplasmic event as the capsid core is too large to remain intact as it traffics through the nuclear pore complex (NPC). However, recent observation suggests that some capsid (CA) protein remains associated with the RTC in the nucleus (Hulme et al., 2015; Peng et al., 2014). Hence it is possible that before nuclear import all the CA is not removed from the reverse-transcription complex (RTC) (Campbell and Hope, 2015). Experimental evidence suggests uncoating of the viral capsid core of HIV-1 might occur by three potential mechanisms. According to early biochemical studies, core disassembly occurred rapidly and relatively completely soon after HIV-1 fusion with the plasma membrane (immediate uncoating) (Fassati and Goff, 2001; Miller et al., 1997).

However, later studies, which included several imaging-based approaches, support a model in which only some core disassembly occurs in the cytoplasm. However, a measurable amount of the viral CA protein remains associated with the RTC, which enables association with critical host factors and nuclear import (cytoplasmic uncoating) (Hulme et al., 2011; Xu et al., 2013). In addition to this, other studies propose an alternative mechanism where the core remains intact until it arrives at the NPC, which allows the replicating viral genome from cytosolic DNA sensors (NPC uncoating) (Lahaye et al., 2013; Rasaiyaah et al., 2013). Both the cytoplasmic and NPC uncoating models support the finding that some amount of CA remains associated with the pre-integration complex (PIC) in the nucleus (Figure 1.11) (Matreyek et al., 2013; Peng et al., 2014; Zhou et al., 2011).
Figure 1.11. Different models of HIV-1 uncoating - Adapted from (Campbell and Hope, 2015)
1.9.3. Pre-integration complex (PIC) formation and nuclear import

The conversion of the viral RNA into cDNA is carried out by the reverse transcriptase enzyme in the reverse transcription complex (RTC), a nucleoprotein complex derived from the core of the infecting virion. Before integration into the host genome, this complex has to be transported to the nucleus and must cross the nuclear envelope (Suzuki and Craigie, 2007). Uncoating of the viral capsid core releases the pre-integration complex (PIC), which is then, recognized by the cellular nuclear-transport machinery to direct it to nucleopores along the microtubular network (Peterlin and Trono, 2003). It is widely accepted that PIC harbours determinants that promote active nuclear import because it is larger (~28 nm) than the passive diffusion limit for nuclear pores (~9 nm) (Mattaj and Englmeier, 1998; Woodward et al., 2009).

Many studies have documented the importance of HIV-1 capsid (CA) protein, importins, and nuclear transport proteins, including NUP153, NUP358, CPSF6, and/or TNPO3 in PIC nuclear import (Ao et al., 2010; Lee et al., 2010; Matreyek and Engelman, 2011; Schaller et al., 2011). As discussed previously, some of the CA protein remains associated with the RTC and has been shown to be important for the nuclear import of the PIC. CA binds to NUP358 and docks the PIC to the cytoplasmic surface of the nuclear pore complex (NPC). At the NPC, CA interacts with CPSF6 to facilitate transport of the PIC through the NPC (Matreyek et al., 2013).

1.9.4. Integration

Integration of the viral genome into the host cell chromosome is an essential step in the viral replication cycle. The integrated proviral DNA then replicates along with cellular DNA during cycles of cell division, like any other cellular gene. The integrated provirus functions as the template for transcription of viral RNAs. Some viral RNAs get translated to yield viral proteins. A portion of the full-length viral RNA is recruited which function as genomic RNA in progeny virions.
The viral protein integrase (IN) mediates the process of integration which is introduced into cells during infection along with reverse transcriptase, the viral RNA, and other proteins as a part of the viral core.

Evidences indicates that human immunodeficiency virus type 1 (HIV-1) preferentially integrates into a subset of transcriptionally active genes of the host cell genome (Schröder et al.; Wang et al., 2007). Recently it has been shown that HIV-1 integration occurs in the outer shell of the nucleus in close association with the nuclear pore. This region comprises of a series of cellular genes, which are preferential targets of the virus (Marini et al., 2015).

1.9.5. Transcription and Translation

HIV-1 gene expression is controlled by the two viral regulatory proteins, Tat and Rev. Tat facilitates transcription elongation by directing the cellular transcriptional elongation factor PTEF-B to nascent RNA polymerases. Rev is essential for the transport of the unspliced and incompletely spliced mRNAs from the nucleus to the cytoplasm. These mRNAs encode the structural proteins of the virus (Karn and Stoltzfus, 2012).

The integrated HIV-1 provirus functions as a transcription template that is regulated at the transcriptional and post-transcriptional levels. At the early stages of infection, HIV-1 produces only short completely spliced mRNAs encoding the viral regulatory proteins Tat and Rev. With the progression of infection, transcription increases sharply, resulting in the production of incompletely spliced mRNAs which are involved in the formation of Env and accessory proteins Vif, Vpr, and Vpu. Full-length unspliced transcripts which act both as the virion genomic RNA and the mRNA for the Gag-Pol polyprotein are also synthesized late (Kim et al., 1989; Pomerantz et al., 1990).
1.9.6. Assembly, release and maturation

Once the integrated viral DNA is transcribed and translated to form new viral RNA and viral proteins, these then translocate to the cell surface to assemble into new immature virus forms (Barre-Sinoussi et al., 2013). During the process of assembly, the viral Env glycoproteins traffic via the secretory pathway, from the rough endoplasmic reticulum (RER) to the golgi and into the vesicles until they arrive at the plasma membrane. Matrix (MA), capsid (CA), nucleocapsid (NC) and p6 domains are synthesized in the cytosol from full-length viral RNA from the Gag precursor polyprotein by means of a programmed frame shifting event during the translation of Gag-encoding viral RNA. Gag recruits the viral genomic RNA and then begins to multimerize. The multimerized Gag reaches the plasma membrane by a still-undefined pathway. Gag then anchors to the plasma membrane in lipid raft microdomains through the insertion of its N-terminal myristate into the lipid bilayer as well as by direct interactions with the phospholipid phosphatidylinositol-(4,5)-bisphosphate.

Into the assembling particle, Env gets incorporated and then recruits endosomal sorting complex required for transport I (ESCRT-I) by means of a direct association between the PTAP motif in p6 and the tumour susceptibility gene 101 (TSG101) subunit of ESCRT-I. With the progress of the budding process, the ESCRT-III and vacuolar protein sorting 4 (VPS4) complexes are recruited which initiates the membrane scission reaction that leads to particle release. Following assembly, immature virions bud from the cell and undergo maturation where formation of the conical capsid core occurs via proteolytic processing of viral polyproteins. This results in the formation of mature virions that are capable of infecting new cells (Laskey and Siliciano, 2014). Figure 1.12 gives an overview of the late stages in the viral replication cycle.
Antiviral drugs are presently available which can inhibit the different stages of the viral life cycle - Fusion inhibitors (entry and fusion step), reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs) (reverse transcription step), integrase strand transfer inhibitors (InSTIs) and allosteric integrase inhibitors (ALLINIs) (integration step) and protease inhibitors (maturation step) (Laskey and Siliciano, 2014).
1.10. HIV-1 TRANSCRIPTION

HIV-1 gene expression is controlled by two of its viral proteins – Tat and Rev as well as sequences present on both ends of the viral genome known as long terminal repeats (LTR) which acts as the viral promoter. HIV-1 Tat activates viral transcription as it stimulates the process of elongation from the viral LTR promoter. HIV-1 Rev is involved in the nuclear to cytoplasmic transport of unspliced and incompletely spliced mRNAs encoding the structural proteins of the virus (Karn and Stoltzfus, 2012). Once integrated, the HIV-1 provirus functions as a transcription template that can be regulated at transcriptional and post-transcriptional levels. Soon after infection, HIV-1 produces short completely spliced mRNAs that encodes for viral proteins Tat, Rev and Nef. With the infection progression, the transcription process increases steeply leading to formation of incompletely spliced mRNAs which encodes for Env and the HIV-1 accessory genes Vif, Vpr and Vpu. Full-length unspliced transcripts also function as virion genomic RNA as well as mRNA for the Gag-Pol polyprotein (Kim et al., 1989; Pomerantz et al., 1990).

1.10.1. HIV-1 Transactivation by Tat

As in all retroviruses, the 5’ LTR functions as the viral promoter in HIV-1 as well. HIV-1 Tat functions as a transactivating factor for transcription of genes under the control of LTR promoter. A regulatory element termed transactivation-responsive region (TAR), is located downstream of the transcription initiation site between the nucleotides +1 to +59. This TAR RNA sequence forms a highly stable and nuclease resistant stem-loop structure, the formation of which is essential for HIV-1 Tat mediated transcription (Berkhout et al., 1989; Selby et al., 1989) as Tat can specifically recognize this region. As mentioned before, Tat regulates transcription at the elongation step rather than the initiation step. In the absence of Tat, the RNA polymerases stall near the promoter (Kao et al., 1987).
Once bound to TAR RNA region Tat recruits host factors cyclin-dependent kinase 9 (CDK9) and Cyclin T1 which forms the PTEF-B complex (Wei et al., 1998; Zhu et al., 1997). The PTEF-B complex hyperphosphorylates the RNA polymerase II at its C-terminal domain which increases the processivity of transactivation and enhances it up to 10-100 fold (Isel and Karn, 1999). The PTEF-B complex thereby functions as a co-factor for HIV-1 Tat mediated transactivation (Figure 1.14).

![Diagram](image)

**Figure 1.13. Schematic representation of Tat-mediated transactivation - Adapted from (Isel and Karn, 1999)**

1.10.2 Rev-dependent nuclear export (Post-transcriptional regulation)

As unspliced and incompletely spliced transcripts from cellular genes that typically degrade within the nucleus, HIV-1 expresses regulatory factors to facilitate the nuclear export of intron-containing viral RNA. The most important of these viral encoded regulatory factors is the HIV-1 Rev protein, that interacts with an RNA element in the env gene known as the Rev-responsive element
An introduction to 7/9 V and its Pathogenesis (RRE) of 351 nucleotides (Malim et al., 1989; Sodroski et al., 1986). Rev binds to the RRE region through its arginine-rich domain (ARD). Rev-RRE complex then interacts with Crm1 (exportin 1) through a leucine-rich nuclear export signal (NES) near the C-terminus of Rev and facilitates nuclear export. The complex is then destabilized by means of RanGAP (Ran GTPase activating protein) and RanBP1 (Ran binding protein 1) leading to release of Rev which is then transported back to the nucleus by importin-β (Henderson and Percipalle, 1997) (Figure 1.14).

Figure 1.14. Rev-dependent nuclear export cycle - Adapted from (Karn and Stoltzfus, 2012)

1.11. HIV-1 VIRAL PROTEINS AND THEIR FUNCTIONS

1.11.1. Gag Polyprotein

There are three fundamental defining open-reading frames (ORFs) for all members of the retrovirus family. Among the major structural proteins are the Gag (group specific antigen) polyproteins which constitute the internal structure of all retroviruses including HIV-1. It comprises about 50% of the total mass of...
the viral particle (Bell and Lever, 2013). From the gag gene forms a 55kD Gag precursor protein (p55), synthesized from the unspliced viral mRNA. During the process of translation, the N terminus of p55 gets myristoylated (Bryant and Ratner, 1990). Gag polyprotein which associates with the membrane recruits two copies of the viral genomic RNA as well as other viral proteins that trigger viral budding from the cell surface of the infected cell.

Following budding, p55 Gag protein undergoes cleavage by the viral protease into its four constituent proteins – matrix protein (MA, p17), capsid protein (CA, p24), nucleocapsid protein (NC, p9) and p6 peptide (Gottlinger et al., 1989) during the process of virion maturation (Figure 1.15). At least 5000 Gag molecules in the immature virion are cleaved by viral protease during the process of maturation (Campbell and Hope, 2015) (Figure 1.16).
An Introduction to HIV and its Pathogenesis

Chapter 1

1.11.1.a. Matrix (p17 or MA)

The 128 amino acid long matrix (MA) polypeptide derived from the myristoylated, N-terminal end of p55, plays crucial role in viral life cycle. The process of myristoylation enables membrane targeting where viral assembly and budding occurs. The MA molecules are found attached to the inner region of the virion lipid bilayer by interacting with phosphatidylinositol-4,5-bis-phosphate.
(PIP2) to form higher order structures (trimers and hexamers) that helps in the stabilization of the viral particle and gives it a three-dimensional structure (Hill et al., 1996). Another subset of MA gets recruited within the deeper layers of the virion to become part of the nuclear import complex and assists the viral DNA to the nucleus (Gallay et al., 1995). MA proteins enable nuclear import of the viral genome using a nuclear localization signal on it which is recognized by the cellular nuclear import machinery. This phenomenon permits HIV in infecting non-dividing cells, which is an unusual property for a retrovirus (Lewis et al., 1992).

1.11.1.b. Capsid (p24 or CA)

During the process of maturation, the capsid (CA) protein assembles spontaneously into a fullerene structure to form the conical core of the HIV-1 virion (Campbell and Hope, 2015). It is within this capsid core structure that the HIV genome, the viral replicative enzymes - reverse transcriptase and integrase as well as some accessory proteins are enclosed. Approximately 1500 CA monomers assemble to form hexamers predominantly along with few pentamers to result in the formation of an enclosed core structure which is curved on the top and bottom (Figure 1.17) (Ganser et al., 1999; Li et al., 2000).

![Figure 1.17. Schematic representation of fullerene cone capsid core - formed from hexameric (orange) and pentameric (yellow) subunits - Adapted from (Campbell and Hope, 2015)](image-url)
The CA protein comprises of two domains – an N-terminal domain which is 150 amino acid long (CA_{NTD}) and a C-terminal domain of 80 amino acids in length (CA_{CTD}) (Campbell and Hope, 2015). During the process of assembly into hexamers and pentamers, the CA_{NTD} orients towards the outer surface of the capsid core and the CA_{CTD} towards the interior of the core (Figure 1.18) (Campbell and Hope, 2015). Each CA_{NTD} comprises of three $\alpha$-helices that helps in the stabilization of the CA (Ganser-Pornillos et al., 2007). The contacts between CA_{NTD} and CA_{CTD} of adjacent monomers further stabilize the hexameric and pentameric subunits (Byeon et al., 2009; Ganser-Pornillos et al., 2007). This CA_{NTD} and CA_{CTD} also form a binding pocket through which numerous cellular factors interact and play important functions during infection.

Figure 1.18. Structural organization of the capsid subunits – CA_{NTD} (Blue) and CA_{CTD} (Red) - Adapted from (Campbell and Hope, 2015)
1.11.1.c. Nucleocapsid (p9 or NC)

The virions released from the surface of an infected cell contain a genome which is composed of a dimer of two identical full length RNA molecules (Didierlaurent et al., 2011). Packaging of full length RNA is predominantly mediated by its 5'UTR. HIV-1 5'UTR can form a series of secondary structures which includes the transactivation response element which forms the TAR hairpin, a polyadenylation signal hairpin (polyA), a large folded structure which includes primer binding site (called PBS domain) as well as three stem loop structures (SL1, SL2 and SL3) which form the Psi (ψ) region (Berkhout and van Wamel, 2000; Paillart et al., 2004; Wilkinson et al., 2008). Among the three stem loop structures, SL1 and SL3 are the major packaging determinants (D’Souza and Summers, 2005).

Nucleocapsid (NC) domain of the Gag polyprotein recognizes the packageable full length RNA during the assembly of virions. NC is relatively a smaller basic protein derived from the Gag polyprotein which binds to SL2 and SL3 motifs with high affinity (Amarasinghe et al., 2000; De Guzman et al., 1998). Also, about 1500 NC molecules coat the full length RNA in virions (Chertova et al., 2006). HIV-1 NC possesses two copies of highly conserved CCHC Zn finger motifs flanked by basic amino acids (Muriaux and Darlix, 2010). A mutation in these highly conserved CCHC residues can impair the packaging of the full length RNA and result in the production of non-infectious viral particles (Darlix et al., 1995; Levin et al., 2005).

1.11.1.d. p6

The C-terminal region of Gag polyprotein gives rise to p6, which is rich in proline amino acids. P6, predominantly is a phosphoprotein (Muller et al., 2002) and is primarily involved in the virion assembly and budding stages (Gottlinger, 2001). During budding, the virus interacts with components of cellular ESCRT (endosomal sorting complexes required for transport) as well as other viral
determinants which mainly includes gag structural proteins (Votteler and Sundquist, 2013). Two p6 domains - domain 1 which interacts with Tsg101 (tumor susceptibility gene 101 protein) and domain 2 which interacts with Alix proteins, enables recruitment of ESCRT complex (Garrus et al., 2001; VerPlank et al., 2001).

1.11.2. Envelope polyprotein

The \textit{env} ORF gives rise to the Env polyprotein. Env protein can undergo extensive glycosylation and oligomerizes to form heavily glycosylated trimers of gp120 and gp41 (Doms, 1993). The virus-encoded envelope glycoprotein forms a gp160 precursor, which undergoes trimerization to form, trimeric envelope spike (gp160)_3. This is then cleaved by a furin-like protease into two non-covalently associated fragments – gp120, the receptor-binding fragment and gp41, the fusion fragment which can also exists as a trimer (g120/gp41)_3. This trimer mediates viral entry (Kovacs et al., 2014) by attachment and fusion to the surface of the host cell membrane. gp120 binds to the primary host receptor CD4 and then interacts with a co-receptor (CCR5 or CXCR4). This then triggers huge conformational changes which are followed by a cascade of processes for fusion of the virus to the cell surface. Since Env protein recognizes the cell surface receptors present on the susceptible cells, it is the determining factor for the cell type that a retrovirus can infect.

The incorporation of Env into new virions can occur in different ways (Figure 1.19):

a) Passive incorporation, where the Env expressed on cell surface gets passively acquired during assembly and budding

b) Gag-Env targeting mechanism where both Gag and Env gets targeted to a specific site on the plasma membrane which allows Env to be concentrated at sites of assembly
c) Direct interaction between MA domain of Gag and cytoplasmic tail of gp41 subunit of Env
d) Indirect interaction between MA domain of Gag and cytoplasmic tail of gp41 subunit of Env through a host protein which bridges the two viral proteins.

Figure 1.19. Different mechanisms of Env incorporation into new virions - Adapted from (Freed, 2015)

1.11.2.a. Glycoprotein 120 (gp120 or SU)

gp120 is a 550 amino acid surface glycoprotein which is heavily glycosylated. gp120 consists of five hyper variable regions (V1-V5) and five constant regions
An introduction to HOW and its Pathogenesis (C1-C5). V1-V4 leads to formation of surface exposed loops with disulphide bonds at their bases (Willey et al., 1986).

The constant region of gp120 which forms the core of the gp120 protein interacts with CD4 receptor of the host cell (Kwong et al., 2000). The HIV-1 gp120 core comprises of inner and outer domains as well as a bridging sheet. The inner domain interacts with gp41. The ‘proximal’ side of the gp120 core resides closer to the viral membrane whereas the ‘distal’ side of gp120 faces outside to the target cell surface after binding to CD4 (Figure 1.20).

![Figure 1.20. gp120 Core structure - Adapted from (Guttman et al., 2012)](image)

1.11.2.b. Glycoprotein 41 (gp41 or TM)

The gp41 is a trans-membrane protein which is 345 amino acids in length. gp41 enables fusion of viral lipid envelop with the cell membrane. It possesses three domains, namely, extracellular domain, trans-membrane domain (TMD) and cytoplasmic tail (CT). Extracellular domain is responsible for membrane fusion. It comprises of a hydrophobic region known as fusion peptide (Bosch et al., 1989), a polar region, hydrophobic heptad repeats (HR1 and HR2) and finally a tryptophan-rich membrane-proximal external region (MPER) (Chan et al., 1997).
Fusion peptide which is the major fusion determinant is normally buried in the gp120/gp41 complex. Binding of gp120-gp41 to host cell surface receptors triggers conformational changes in gp41. This elicits exposure of fusion peptide and induces its penetration into the host cell which leads to membrane destabilization and fusion pore formation (Brasseur et al., 1990). HR1 and HR2 regions facilitate the fusion process by holding viral and cell membrane in close proximity. MPER, is essential for fusogenicity and virus infectivity and is an important target of many neutralizing antibodies.

1.11.3. Reverse Transcriptase (RT, p51/66)

The pol gene of the virus encodes for reverse transcriptase. Autocleavage of pol polyprotein generates three proteins – Reverse transcriptase, integrase and protease. RT is active as a heterodimer form, comprising of p66 and p51 (Lightfoote et al., 1986). RT exhibits three different enzymatic activities namely-RNA dependent DNA polymerization, DNA dependent DNA polymerization and ribonuclease H (RNase H) activity (Hansen et al., 1987; Hansen et al., 1988). However, RT lacks 3'-5' exo-nuclease activity and hence it is an erroneous polymerase as there is no proofreading associated with it. It causes a mutation rate of 1 in every 5000 base pair (Preston et al., 1988; Roberts et al., 1988). Before integration into the host genome, the virus has to synthesize double stranded DNA from its RNA genome. Reverse transcription involves the following steps (Figure 1.21):
An Introduction to HIV and its Pathogenesis

Figure 1.21. Steps involved in HIV-1 reverse transcription - Adapted from Alan Cann: Principles of molecular virology, 2005
1. HIV-1 specific cellular tRNA hybridises with primer binding site (PBS) at the 5' end leader sequence of the viral RNA
2. RT, using its RNA dependent DNA polymerase activity, synthesizes DNA segment
3. RNaseH removes R and U5 region of the viral genome
4. DNA:tRNA complex is transferred from 5' end of the viral RNA template to 3' end
5. Hybridization of DNA with the remaining R sequence present on 3' LTR followed by DNA extension from 3' end
6. RNase H removes most of the viral RNA leaving a patch called polypurine tract (PPT)
7. PPT induces extension of second strand of DNA
8. Meanwhile RNase H removes all the remaining viral RNA and tRNA
9. The PBS region of both the DNA strands are hybridised with each other followed by extension of both the strands.

1.11.4. Integrase (IN, p32)

Integration is an obligatory step in the life cycle of HIV-1 infection. The integrase protein is responsible for mediating the integration of HIV-1 viral genome into the genome of its specific host cell. The protein exists and functions as a tetramer where each monomer comprises of three domains (Figure 1.22):

- N-terminal domain – involved in multimerization and has a His2Cys2 motif which chelates zinc
- Central catalytic domain – which mediates catalytic activity through the DDE motif of the enzyme by binding to the ends of viral DNA
- C-terminal domain – which non-specifically binds to the DNA through an SH3-like fold in order to make the complex more stable (Delelis et al., 2008)
Following nuclear entry of PIC, viral genome is integrated into host genome through three major steps - 3' end processing, strand transfer and gap repair. The process involves:

1. Interaction of two molecules of integrase with 3' end of the double stranded viral DNA
2. Two nucleotides from both 3' ends are removed which leads to generation of 2-base overhangs at both ends
3. This complex then interacts with host DNA by bringing the two 3' ends to close proximity and initiating strand transfer by insertion of the processed viral DNA into host genome by trans-esterification reaction (Delelis et al., 2008; Engelman et al., 1991)
4. Strand transfer introduces 'gaps' at viral host DNA interaction junctions which triggers 'gap repair' (mediated by host polymerases, nucleases and ligases) process in the cell, which is a DNA damage response
5. Polymerase fills the gap during the process of extension of the host DNA
6. Nuclease is involved in removal of the 5' dinucleotide flaps on viral DNA
7. Ligase ultimately joins the unbound viral and host DNA strands. The integrated virus, at this stage is called a provirus.

1.11.5. Protease (PR, p10)

HIV-1 protease, which is a 99 amino acid long retroviral aspartyl protease is also formed by autocleavage of Pol polyprotein. It functions in cleavage of proteins by
recognizing amino acid aspartate in proteins. Primarily, protease functions to mediate HIV-1 virion maturation events and is also essential for viral infectivity (Adamson, 2012). Protease exists and functions as a dimer, which comprises of monomers of approximately 11 kDa. The enzyme active site is present on the interface of the two monomers. Protease gets activated following budding of the virus by an autocatalytic mechanism. Immature virions containing Gag and Gag-Pol polyproteins undergoes protease-mediated cleavage of these polyproteins, ultimately resulting in the formation of mature infectious virions. The immature virions undergo twelve proteolytic cleavages to generate fully mature forms of infectious virions (Kohl et al., 1988; Ikuta et al., 2000). As previously discussed, Pr55gag undergoes cleavage to produce matrix, capsid, nucleocapsid, p1, p2 and p6gag. Pr160gag-pol precursor polyprotein produces integrase, protease, reverse transcriptase (p51 and p66) as well as p6pol. Protease also cleaves the Nef precursor to form functional Nef (Schorr et al., 1996).

### 1.11.6. Tat

Tat is abbreviated for Trans-activator protein of HIV-1, which is a 72-101 amino acid containing protein as well as one among the early expressed proteins of HIV-1. Tat is a basic protein and has affinity for binding to nucleic acids. Tat comprises of three domains:

- N-terminal activation domain
- Central TAR (Transactivation response element) binding and NLS containing domain
- C-terminal domain (Kuppuswamy et al., 1989)

On the basis of its amino acid composition, Tat structural domains can be further divided into six regions (Figure 1.23):

- Proline-rich acidic region
- Cysteine-rich region
- Core region
- Basic region
• Glutamine-rich region and
• RGD-motif containing region (which is encoded by second exon from amino acid 60-72).

Tat activation domain constitutes Proline-rich acidic region, Cys-rich and core regions (Reddy et al., 1992). Basic and Gln-rich regions form part of the TAR binding and NLS containing domain (Verhoef et al., 1997). As discussed previously, the major function of Tat is in the elongation step of transcription. Although transcription can be initiated from the LTR, in the absence of Tat, RNA polymerase II gets stalled leading to formation of truncated transcripts. Tat enables the process of elongation, resulting in formation of full-length transcripts.

![Figure 1.23. Structural domains of HIV-1 Tat - Adapted from (Bose et al., 2015)](image)

1.11.7. Rev

Rev, abbreviated for Regulator of Expression of Viral proteins, is one of the two regulatory proteins of HIV-1. It is a phosphoprotein of 19kDa and comprises of 116 amino acids. Rev is typically localized in the nucleus as well as nucleolus and functions in the nuclear export of unspliced transcripts and genomic RNA of HIV-1 as discussed previously. Rev contains four structural domains (Figure 1.24):

• N-terminal arginine rich region which contains nuclear localizing signal (NLS) and Rev response element (RRE) binding region. This region also contains a nuclear inhibitory sequence (NIS), which regulate the activity of Rev by constraining it to the cytoplasm.
• C-terminal region which possess an activation domain that is responsible for the effector function of Rev, a nuclear export signal
(NES). This region also interacts with various cellular proteins essential for nuclear export of mRNA (Meyer and Malim, 1994; Suhasini and Reddy, 2009). NLS and NES sequences enable nuclear-cytoplastic shuttling of Rev protein.

**Figure 1.24. Structural domains of HIV-1 Rev** - Adapted from (Suhasini and Reddy, 2009)

### 1.11.8. Nef

HIV-1 Nef abbreviated for ‘Negative factor’, is a 27kDa accessory protein of HIV-1 and is among the early expressing viral proteins. Nef had been previously designated as a negative factor based on initial studies which identified it as a negative regulator of HIV-1 transcription and replication (Ahmad and Venkatesan, 1988; Niederman et al., 1989). Later studies however have contradicted these earlier studies and demonstrated that Nef has positive roles on viral replication (de Ronde et al., 1992; Miller et al., 1994; Spina et al., 1994; Terwilliger et al., 1991; Zazopoulos and Haseltine, 1993). The term Nef is hence thought to be a misnomer now.

Nef possess a central globular core domain, which is flanked by flexible N-terminal and dis-ordered C-terminal domains (Figure 1.25). There are few conserved motifs present in Nef, which include:
• Myristoylation signal, which is essential for its localization and plasma membrane binding
• An internal methionine for translation initiation,
• Acidic charge region,
• PxxP repeats,
• Protein kinase C phosphorylation site and
• Arginine rich motifs (Geyer et al., 2001).

These motifs play important functional roles in CD4 and MHC-down-modulation mediated by Nef as well as is involved in interaction with various cellular proteins. The motifs responsible for myristoylation, CD4 and MHC-I down-regulation are present in N-terminal flexible region (Arolf and Baur, 2001; Geyer et al., 1999).

HIV-1 Nef is majorly involved in the down-modulation of CD4 receptor (Garcia and Miller, 1991) by endosomal/lysosomal degradation. Nef has also been
shown to down-modulate MHC-I (Schwartz et al., 1996). This enables protection of infected cells against cytotoxic T-cells mediated killing (Collins et al., 1998). HIV-1 Nef hijacks cellular signalling pathways to manipulate the activation of lymphocytes (Alexander et al., 1997; Simmons et al., 2001). Nef has been demonstrated to increase viral replication in both activated and un-activated primary T cells. However, the effect of Nef was found to be more profound in resting T-cells as compared to proliferating T cells (Spina et al., 1994). Nef has been long known to be involved in regulation of viral infectivity (Aiken and Trono, 1995; Goldsmith et al., 1995; Khan et al., 2001; Miller et al., 1995; Papkalla et al., 2002; Tobiume et al., 2001; Tokunaga et al., 1998). However, the precise mechanistic role of Nef in regulation of viral infectivity was deciphered only recently where Nef was shown to sequester cellular restriction factors SERINC5 and SERINC3 to maintain infectivity (Rosa et al., 2015; Usami et al., 2015).

1.11.9. Vpr

HIV-1 Vpr which is a virion-associated 14-kDa basic protein comprising of 96 amino acids and is expressed late during infection. Vpr is a regulatory protein and is abbreviated for Viral Protein R. During the process of virion assembly, Vpr gets packaged in the virion, which suggests its role in early stages of viral life cycle.

Vpr comprises of different functional domains (Figure 1.26):

- Central region with three amphiphilic helices (17-33, 38-50 and 55-77), flanked by unstructured N- and C-terminals (Morellet et al., 2003).
- Negatively charged N-terminal and positively charged C-terminal
- Proline rich residues required for interaction with cellular peptidyl-propyl isomerase; cyclophilin A. This interaction mediates Vpr folding
- Ariginine residues at C-terminal which are essential for the transducing property of Vpr
• Leucine residues of third helix important for oligomerization of Vpr as well as interaction with cellular proteins
• Hydrophobic amino-acids in helix-I essential for G2/M cell cycle arrest and induction of apoptosis by Vpr.
• Nuclear localizing signal (NLS) and nuclear export signal (NES) (Guenzel et al., 2014)

Figure 1.26. Structural domains of HIV-1 Vpr - Adapted from (Guenzel et al., 2014)

Vpr expression in dividing cells results in G2/M phase arrest of the cell cycle. Induction of apoptosis is also seen associated with Vpr expression. With Vpr expression, a condition similar to that of DNA damage is established as is evident from the phosphorylation status of proteins that regulate cell cycle as well as by the activation of ATR and Chk1 (Tachiwana et al., 2006). Vpr also modulates the activity of reverse transcriptase and nuclear import of PIC. Vpr also regulates cellular and viral gene expression (Guenzel et al., 2014). Vpr promotes proteasomal degradation of cellular enzyme uracil-N-glycosylase (UNG) by hijacking proteins of the ubiquitin machinery Cullin 1 and Cullin 4 (Schrofelbauer et al., 2005). Lately, Vpr has been shown to be essential and sufficient for redirecting the host ubiquitin pathways in a manner that is beneficial to the virus (Arora et al., 2014).

1.11.10. Vpu

HIV-1 Vpu, for Viral protein U, is an 81 amino-acid long accessory phosphoprotein of HIV-1. It comprises of a transmembrane N-terminal domain and long
An introduction to KQV and its Pathogenesis

Cytoplasmic C-terminal domain (Maldarelli et al., 1993; Strebel et al., 1988). Vpu is typically found in ER, golgi bodies as well as endosomes. Two major roles have been associated with Vpu. Primarily, Vpu down-modulates CD4 receptor (Willey et al., 1992) by interacting with cytoplasmic tail of CD4. Vpu recruits a cullin1-SKP1-Fbox ubiquitin ligase complex through its interaction with SKP1 and β-TrCP, which leads to cellular ER associated degradation (ERAD) of CD4 (Magadan et al., 2010; Margottin et al., 1998). Upon synthesis of gp160, interactions between gp160 and CD4 results in the retention of gp160 in the ER. HIV-1 Vpu later causes degradation of CD4 through the endoplasmic reticulum associated degradation (ERAD) by recruiting cullin1-SKP1- β-TrCP ubiquitin ligase complex which leads to liberation of gp120 from ER. This is then followed by maturation and trafficking to cell surface, which then gets incorporated in virions) (Figure 1.27). The second major function associated with Vpu is its role in budding of newly synthesized virions from the cell by inhibiting cellular factor, BST2 (tetherin) (Neil et al., 2008; Van Damme et al., 2008).

Figure 1.27. Vpu mediated CD4 receptor down-regulation - Adapted from (Roy et al., 2014)
1.11.11. Vif

HIV-1 Viral infectivity factor (Vif), a 23kDa accessory protein of the virus, was characterized nearly two decades after the discovery of the virus. The major function of HIV-1 Vif protein seems to be the ubiquitination and subsequent proteasomal degradation of host restriction factor APOBEC3G (Marin et al., 2003; Sheehy et al., 2003) as well as other APOBEC family members (Desimmie et al., 2014). APOBEC3G family of restriction factors are a group of deoxycytidine deaminases that suppresses HIV-1 infection by incorporation into progeny virions of retroviruses and inhibits viral replication by inducing mutagenesis and functional inactivation of the virus. Among the seven APOBEC3 enzymes APOBEC3D, F, G, and H offer restriction to HIV-1. When APOBEC3G incorporates into virions during replication of minus strand to plus strand of DNA, HIV-1 RT incorporates G to A hypermutations in the plus strand of viral DNA that can functionally inactivate HIV-1 (Feng et al., 2014). Vif hijacks the ubiquitin proteasome system comprising of Cul5, EloB, EloC (Sheehy et al., 2003; Stopak et al., 2003) and additionally recruits a transcription factor CBFβ (Jager et al., 2012; Zhang et al., 2012), to mediate ubiquitination and proteasomal degradation of APOBEC3G. Vif incorporates itself into the EloBC-Cul5-Rbx2 E3 core and then hijacks CBFβ from the RUNX transcription complexes (Figure 1.28).

![Figure 1.28. Vif-mediated counteraction of APOBEC3G - Adapted from (Jager et al., 2012)](image-url)
Vif binds to APOBEC3G/F through several motifs in its N-terminal region. The BC-box of Vif binds to EloC (Yu et al., 2003) and a zinc-finger motif upstream of BC-box shows interaction with Cul5 (Luo et al., 2005; Mehle et al., 2006). Vif also possesses a conserved PPLP (Pro-Pro-Leu-Pro) sequence which is present downstream of the BC-box (Simon et al., 1999) which has been attributed to the multimerization of Vif (Yang et al., 2001) (Figure 1.29). A stretch of 12 conserved residues follow the PPLP motif which has not yet been identified to have any specific function. However, it has been shown that these amino acid residues do not interfere with the function of Vif in degradation of APOBECs (Simon et al., 1999).

![Figure 1.29. Schematic diagram of functional domains of Vif - Adapted from (Salter et al., 2014)](image)

1.12. INTRINSIC HOST RESTRICTION FACTORS OF HIV-1

To replicate and successfully propagate in the host cells, HIV-1 evades the complexities and intrinsic resistance mechanisms by the mammalian cell. Restriction factors constitute an expanding group of cellular proteins which create powerful barriers to the inbound virus, the most well-characterized of them being APOBEC3 family of proteins (Goff, 2003; Sheehy et al., 2003), TRIM5 (Aiken and Joyce, 2011; Pertel et al., 2011), BST2/ Tetherin (Neil et al., 2008; Van Damme et al., 2008), SAMHD1 (Ryoo et al., 2014; Yang and Greene, 2014), Mx2/MxB (Goujon et al., 2013; Kane et al., 2013) and recently identified SERINC3 and SERINC5 (Rosa et al., 2015; Usami et al., 2015). HIV-1 impedes these restrictive mechanisms of the host, mostly with the help of its highly evolved accessory proteins Nef, Vif, Vpr and Vpu, by using them to harness the functions of various host cellular factors.
1.12.1. APOBEC3G

During HIV-1 infection, the cellular restriction factor, APOBEC3G causes deamination of the non-coding (minus) strand of the genome from deoxyC to deoxyU resulting in G to A hypermutations in the coding strand that restricts propagation of the virus. The anti-viral properties of APOBEC3G were first recognized in the process of virus-encoded antagonist, HIV Vif (Sheehy et al., 2002). Vif protein is an important regulator of viral infection, as it negatively regulates APOBEC3G expression by mediating its polyubiquitination and subsequent proteasomal degradation. Vif achieves this by acting as an adaptor to recruit APOBEC3G to the Cul5 elongin B/C-Rbx SCF ligase (Yu et al., 2003). Ubiquitination and proteasomal degradation of Vif has also been reported, but there is no clarity regarding this (Mehle et al., 2004). Vif therefore functions in the depletion of cytosolic pool of APOBEC3G and prevents its incorporation into assembling virus particles. Vif also targets APOBEC3F molecules which are ubiquitinated and degraded through a Cul5-dependent SCF E3 ligase complex (Liu et al., 2005). The mechanism by which APOBEC3G mediates host restriction is shown in Figure 1.30. If Vif functional activity fails to neutralize all APOBEC3F/G activity within a cell, Vpr protein comes into play where it forms a complex with Cul1 and Cul4 to target cellular uracil DNA glycosylase (UNG) for ubiquitination and proteasomal degradation (Schrofelbauer et al., 2005).
1.12.2. TRIM5α

Tripartite motif-containing protein 5 (TRIM5) which is also known as RING finger protein 88 is a protein encoded by the TRIM5 gene. TRIM5α is a potent HIV-1 restriction factor that was isolated initially from rhesus macaques (Stremlau et al., 2004). It provides potent intrinsic defense in mammalian cells against retroviral infections. TRIM5α functions by blocking viral capsid entry prior to reverse transcription (Black and Aiken, 2010) (Figure 1.31).
TRIM5α dependent restriction of HIV-1 infection is associated with the structural perturbation of the viral capsid core which leads to aberrant HIV-1 uncoating in target cells. This, in turn, prevents successful transcription and nuclear transport of the viral genome. Cyclophilin A (CypA) which is recruited into nascent HIV-1 virions as well as incoming HIV-1 capsids renders HIV-1 sensitive to TRIM5α mediated host restriction in African green monkey and rhesus macaques. Cyclosporine A is a competitor of Cyclophilin A which opposes TRIM5α dependent host restriction (Keckesova et al., 2006). However, the human TRIM5α do not restrict HIV-1 infection.

1.12.3. TETHERIN

Tetherin (BST2 or CD317), another host restriction factor, prevents release of virus particle by inserting its N-terminal transmembrane domain in the plasma membrane of producer cell and its GPI-linked C terminus in the virus envelope lipid bilayer. The virus will therefore be tethered and thereafter endocytosed. Tetherin comes in contact with cortical actin through an interaction with RICH2. Syk phosphorylates tetherin on its tyrosine residues near its N-terminus which
then activates TRAF2, TRAF6 and TAK1. This complex phosphorylates IKK, leading to the degradation of IκBα, thereby activating NF-κB, which ultimately causes induction of transcription of the proinflammatory cytokines CXCL10, IL-6 and IFN-β. Tetherin is counteracted by HIV-1 Vpu which sequesters tetherin in the endoplasmic reticulum, thereby preventing its transit to the plasma membrane. Tetherin is also proteasomally degraded by the SCF-β-TRCP complex. In case of SIV infection, Nef binds to tetherin at the plasma membrane which induces endocytosis of tetherin through an AP2-dependent pathway (Simon et al., 2015) (Figure 1.32).

Figure 1.32. Tetherin-mediated host restriction - Adapted from (Simon et al., 2015)

1.12.4. SAMHD1

The cellular factor SAM domain and HD domain containing protein 1 (SAMHD1), functions as a host restriction factor against HIV by depleting the
intracellular dNTP pool. Resting CD4+ T cells are highly resistant to HIV-1 infection in comparison to activated CD4+ T cells. SAMHD1, which is abundantly expressed in resting CD4+ T cells (dendritic and myeloid cells), restricts reverse transcription of HIV-1 RNA in these cells (Baldauf et al., 2012) by lowering the dNTP pool to concentrations below than the requirement by reverse transcriptase enzyme (Lahouassa et al., 2012). This function of SAMHD1 is regulated by the phosphorylation status of SAMHD1 at Thr592 by cyclin dependent kinases Cdk1, Cdk2 and Cdk6 (Ryoo et al., 2014). In resting cells, SAMHD1 activity is counteracted by viral protein Vpx in HIV-2 as well as SIV infections, by causing proteasomal degradation of SAMHD1 (Lahouassa et al., 2012) and thereby elevating the dNTP pool (Figure 1.33).

Host cellular factors, Cyclin A2 and Cyclin L2 are exploited by HIV-1 to evade the restriction function of SAMHD1 in different cell types. While Cyclin
A2/Cdk1 complex restricts SAMHD1 activity in cycling cells by phosphorylating SAMHD1 and thus inactivating it (Cribier et al., 2013). Cyclin L2 acts as an HIV-1 dependency factor in non-cycling macrophages as it causes the degradation of SAMHD1 in these cells (Kyei et al., 2015) thereby enabling the virus to replicate efficiently.

1.12.5. MxB/Mx2

The interferon-inducible myxovirus resistance (Mx) proteins are dynamin like large GTPases that play crucial roles in combating a wide range of virus infections. There are mainly two types of myxovirus resistance genes – Mx1 (codes for MxA protein) and Mx2 (codes for MxB protein) (Haller, 2013). MxA has been long known for its inhibitory activity against many RNA and DNA viruses, whereas the antiviral activity of MxB had been relatively less established. Lately, human MxB has been identified to inhibit HIV-1 infection by decreasing the level of integrated viral DNA (Goujon et al., 2013; Kane et al., 2013). HIV-1 can escape the MxB restriction by a mutation in the alanine residue at position 88 in the capsid protein of the virus. This leads to a consequent loss of CA interaction with the CypA, thereby suggesting a role for CypA in MxB mediated HIV-1 restriction (Haller, 2013). A schematic overview of Mx2-mediated host restriction is shown in Figure 1.34.

![Figure 1.34. MX2-mediated host restriction against HIV-1 - Adapted from (Goujon et al., 2013)](image-url)
1.12.6. SERINC3 AND SERINC5

Serine incorporator 3 (SERINC3) and SERINC5 are multi-pass transmembrane proteins that function as host restriction factors as they are incorporated into HIV-1 virions and are thereby involved in reducing viral infectivity to several folds. Among both these proteins, SERINC5 is a more potent inhibitor of HIV-1 particle infectivity. Both these proteins are highly expressed in HIV-1 target cells (Usami et al., 2015). Viral protein, HIV-1 Nef counters the restriction activity of these proteins during infection in a host cell dependent manner by increasing the intrinsic infectivity of HIV-1 virions (Rosa et al., 2015). In the host cells, SERINC5 is found localized to the plasma membrane where it gets efficiently incorporated into the virions that bud from the surface of the host cell. Incorporation of these proteins in the virions impairs subsequent penetration of virions into new target cells (Figure 1.35). HIV-1 Nef functions by redirecting SERINC5 to a Rab7-positive endosomal compartment, thereby excluding it from HIV-1 particles (Rosa et al., 2015; Usami et al., 2015).

![Figure 1.35. SERINC3/C5 mediated host restriction and counteraction by Nef - Adapted from (Aiken, 2015)](image-url)
1.13. HIV-1 INDUCED APOPTOSIS AND CD4+ T CELL DEPLETION

HIV-1 infection is characterized by the progressive depletion of CD4+ T cells from the circulation as well as from lymphoid organs as the infected person progresses towards development of AIDS. Typically, a healthy adult harbours around \(2 \times 10^{11}\) mature CD4+ T cells. Clinically, a condition where the CD4+ T cell count drops to less than 200 cells/mm\(^3\) of peripheral blood is considered as AIDS (Haase, 1999). HIV-1 infection can lead to depletion of CD4+ T cells through different mechanisms. It could either occur by direct cytotoxicity of infected cells or by programmed cell death (either apoptotic or non-apoptotic) triggered in infected cells. Uninfected or the ‘bystander’ cells may also undergo programmed cell death which could be triggered by soluble or membrane-bound viral or host immune factors (Cummins and Badley, 2010). Some HIV proteins have been identified to be involved in induction of apoptosis in CD4+ T cells. Viral proteins gp120, Tat, Nef, Vpr, Vpu and HIV protease have been attributed to have pro- or anti-apoptotic properties. A brief outline of the association of some of the viral proteins with apoptosis is outlined in Figure 1.36.

The influence of different viral proteins on HIV-1 induced CD4+ T cell depletion:

- The surface protein, gp120 is proposed to have a pro-apoptotic effect with respect to HIV-1 infection. Some of the pro-apoptotic cellular mechanisms induced by gp120 include up regulation of Fas, FasL and TNF\(\alpha\) expression (Oyaizu et al., 1994), induction of G2 phase cell cycle arrest (Kolesnitchenko et al., 1995), generation of reactive oxygen intermediates (Radrizzani et al., 1997), reduced expression of Bcl-2 (Hashimoto et al., 1997), mTOR and p53 phosphorylation (Castedo et al., 2001), increased expression of PUMA (Perfettini et al., 2004), a pro-apoptotic protein and activation of p38 (Trushin et al., 2007).
An introduction to TdOV and its Pathogenesis

Thymus

Apoptosis of immature thymocytes

Unmrt'rd ThMs

A eocrtx byundr*

^  ApofMMir mIkwi r rrt

Apoptosis induced by Death Receptors

FasL

TNF

TRAIL

O c

Massive apoptosis in GALT

Circulating microbial products

Activation-induced cell death

Pro-apoptotic HIV proteins

Gp120

Vpr

Nef

Vpu

Protease

Figure 1.36. Influence of viral proteins in CD4+ T cell depletion – Adapted from (Cummins and Badley, 2010)

• HIV-1 Tat has been demonstrated to have both pro- and anti-apoptotic activities. Picomolar levels of Tat has been shown to have anti-apoptotic activities like apoptotic resistance to TNF, Fas (Gibellini et al., 1995) and TRAIL (Gibellini et al., 2001) with decreased expression of Caspase 10 and increased expression anti-apoptotic genes like Bcl-2 and c-FLIP (Gibellini et al., 2005). However, at nanomolar concentrations, Tat has been shown to function as a pro-apoptotic factor whereby it induces the expression of pro-apoptotic cellular genes FasL (Westendorp et al., 1995), caspase 8 (Bartz and Emerman, 1999) and Bax (Sastry et al., 1996).

• HIV-1 Nef exhibit up regulated expression of Fas, FasL (Zauli et al., 1999) and PD-1 (Muthumani et al., 2008) and decreased expression of Bcl-2 and Bcl-xl (Rasola et al., 2001). With Nef expression, cells undergo apoptosis by both
Caspase-dependent and independent mechanisms. Nef, expressed endogenously in infected cells can lead to lysosomal permeabilization which releases Cathepsin-D into the cytosol leading to mitochondrial membrane rupture (Laforge et al., 2007). HIV-1 Nef associates with Ask1, a member of the MAP3K family, which can activate the Jnk pathway and lead to apoptosis (Xu and Screaton, 2001).

- HIV-1 Vpr influences viral pathogenesis by affecting viral replication as well as transcription and proliferation of the host cell. Vpr induced cell cycle arrest at G2/M phase is widely studied. Apart from this, Vpr is also involved in modulation of T cell receptor triggered apoptosis. Vpr induces apoptosis in the absence of T cell receptor (TCR)-mediated activation whereas this activity is interrupted in the presence of T cell receptor mediated activation. The regulation of apoptosis by Vpr is attributed to suppression of NF-kB activity through the induction of IkB, an inhibitor of NF-kB. Vpr also suppresses the expression of pro-inflammatory cytokines IL-2, IL-10, IL-12, TNFα and IL-4 which are all NFkB-dependent (Ayyavoo et al., 1997). Vpr also causes an up regulation of Bcl2 and down regulation of Bax (Conti et al., 1998). Vpr, after induction of G2/M arrest, induces apoptosis by binding to either Bax or ANT and VDAC in the mitochondrial membrane, which causes release of cytochrome c and activation of caspases 9 and 3 (Jacotot et al., 2000). Vpr expression also leads to increased expression levels of NKG2D ligands, which renders infected CD4+ T cells susceptible to NK-cell mediated killing (Richard et al., 2010).

- Exogenous expression of Vpu in Jurkat cells has shown to increase susceptibility to Fas-mediated apoptosis (Casella et al., 1999). A possible explanation to this is that Vpu expression inhibits NF-kB-mediated expression of anti-apoptotic genes (Akari et al., 2001). Vpu deletion from pNL4-3 proviral construct leads to significant reduction in CD4+ T cell depletion (Rucker et al., 2004).
• Apart from cleaving of the Gag Polyprotein precursor, HIV-1 protease also possess pro-apoptotic functions and can cleave procaspase 8 that generates a 41 kDa fragment – Casp8p41 (Nie et al., 2007) – which can induce apoptosis in infected CD4⁺ T cells via a mitochondria dependent pathway (Nie et al., 2008). However, the exact molecular target of protease in this pathway has not yet been identified.

However, apoptosis of CD4⁺ T cells have not consistently correlated with viral load (Rothen et al., 1997) which raises possibilities that not all of the apoptosis occurring in CD4⁺ T cells is driven by active viral replication or as a result of the immune response against the virus (Cummins and Badley, 2010). With the progress of research, other cell death pathways which include autophagy (Dinkins et al., 2015) as well as pyroptosis (Doitsh et al., 2014) have recently gained relevance in HIV-1 pathogenesis.

CD4⁺ T cells function as important immune effector cells during any pathogenic infection, including HIV-1 infection as they provide optimal help for cytotoxic CD8⁺ T cell function (Bennett et al., 1998; Schoenberger et al., 1998). However, the unfortunate situation in HIV-1 infection is that these helper CD4⁺ T cells themselves are being targeted by the virus. Individuals with strong HIV-1 specific CD4⁺ T cell proliferative responses against HIV-1 p24 antigen have been shown to be able to control viremia in a better way when compared to those with diminished proliferative responses. However, in most cases, infected individuals exhibit very poor or rather no proliferative responses to HIV-1 even under antiretroviral therapy. The underlying mechanism of defective proliferation in HIV-1 infected cells could be associated with defects in regulation of cell cycle machinery induced by the virus.
1.14. THE HOST CELL CYCLE AND ITS REGULATION

Fundamentally, the process of cell cycle forms the basis of all eukaryotic systems from unicellular to complex multicellular organisms and is well-conserved in these organisms. The cell cycle is an orchestrated series of events in which the genetic material is being distributed from the parent cell to the daughter cells. The process of cell division consists of two consecutive processes – DNA replication and segregation of replicated chromosomes into two separate daughter cells. Cell division can be broadly divided into two phases – the process of nuclear division (mitosis) and interphase, which is the phase between two mitotic phases. Mitosis is subdivided into prophase, metaphase, anaphase and telophase; where the actual division takes place.

Figure 1.37. An overview of the mammalian cell cycle. The Cyclin-Cdk complexes specific to each phase, cell cycle check points and the inhibitors of CDKs at different stages are also shown - Adapted from (Asghar et al., 2015)
During interphase, the cell grows in size to enable it for the next mitosis. The interphase is subdivided into G1, S and G2 phases (Norbury and Nurse, 1992). The DNA replication occurs during S phase which is preceded by a gap called G1 during which the cell prepares for DNA synthesis. The S phase is succeeded by another gap phase called G2 where the cell prepares for mitosis. However, there are cells which do not commit for DNA replication and before G1 enters a resting or quiescent stage called G0 (Figure 1.37). Most of the non-proliferating cells in the human body are in this quiescent stage (Vermeulen et al., 2003).

The mammalian cell cycle is regulated by cyclin-dependent kinases (CDKs), whose activity is modulated by either cyclins which function as CDK activators or by CDK inhibitors (Ink4, Cip and Kip inhibitors) (Malumbres and Barbacid, 2009). Hence, the activities of CDKs are stringently regulated during cell cycle transitions to ensure successful cell division. Each phase of the cell cycle is regulated by specific cyclins and its associated CDK partners. An overview of the specific functions of different cyclin-CDK complexes is illustrated in Figure 1.38.
An introduction to 7/9 V
and its Pathogenesis

CDK-Cyc complex

Function

G1-S progression:
- Phosphorylation of RB stimulates E2F
- Accumulation of FOXM1

G1-S progression (DNA replication):
- Hyperphosphorylation of RB
- Centrosome duplication
- Induction of histone synthesis
- Phosphorylation of replication factors

G2-M progression (mitotic entry):
- Nuclear envelope breakdown
- Mitotic condensation
- Spindle assembly

Neuronal viability (G1-S control):
- Phosphorylation of RB

Basal transcriptional processes:
- Transcriptional initiation
- Transcriptional elongation
- RNA processing

1.14.1. HIV-1 induced cell cycle dysregulation

HIV being a virus that replicates in the nucleus of the host cells has evolved multiple strategies for promoting induction of its proliferative capacity by competing with host cell cycle regulatory pathways. CD4+ T cells from HIV-
infected patients have shown poor proliferative tendency on stimulation of T cell receptor (Clerici et al., 1989; Musey et al., 1999). The defect in proliferation occurs probably due to limited expansion of antigen-reactive cells in response to HIV-1 infection. The T cell proliferation function and HIV immunity can be correlated to proliferation responses of CD4+ T cells to HIV antigens and viral replication control (Rosenberg et al., 1997) as well as prolonged activity of cytotoxic T cells (Kalams et al., 1999). Hence it is important to uncover mechanisms that underlie the CD4+ T cell proliferation and associated cellular modulations during HIV-1 infection. The inability of patient cells to proliferate has been associated with decreased production of IL-2 (Fukumori et al., 2000) as well as enhanced susceptibility to apoptosis (Groux et al., 1992; Meyaard et al., 1992).

Despite the presence of highly integrated and complex cell-cycle regulatory mechanisms employed by the host, successful viruses can exploit multiple targets that can outrun several counter measures by the host. Dysregulation of the cell cycle has been observed in gut mucosa cells of both long term non-progressors and viremic patients (Miedema et al., 1988). Thereby, this reveals that a certain replicative state in the host cell is favourable for productive HIV-1 infection (Clerici et al., 1989).

Mainly two hypotheses have been put forth regarding cell cycle progression and CD4+ T cell proliferation in the context of HIV-1 infection. The first suggests that HIV-1 infected cells could be receiving defective early signalling events which ultimately lead to proliferation failure. An alternative hypothesis suggests that HIV-1 infected cells can generate early response signals but fail to complete the division. Experimental evidence suggests that cells from HIV-1 infected patients spontaneously enter S phase of the cell cycle upon culture, however without completing mitosis these cells tend to undergo apoptosis because of irreparable DNA damage (Patki et al., 2000; Sieg et al., 2001).

Cell cycle arrest in G2 has been characterized in detail in the context of DNA damage. Cells have evolved a variety of response pathways to protect the integrity
of their genomes and have coordinated these pathways with cell cycle progression and apoptosis. The objectives of these pathways are to excise damaged DNA, rejoin DNA strand breaks, or directly reverse lesions with the ultimate goal of preventing propagation of DNA mutations through cell division mainly in the S phase. Accordingly, checkpoint activation leads to cell cycle arrest prior to DNA replication (G1/S arrest) or prior to mitosis (G2/M arrest) to allow time for repair. If the damage is irreparable, the cell dies by apoptosis (Andersen et al., 2008). This cell cycle regulation is impaired during HIV-1 infection, thus leading to extensive apoptosis of the infected cells as well as the bystander cells.

HIV-1 manipulates the DNA damage responses as well as cell cycle check points to promote viral replication which includes primary infection, latency and pathogenesis of HIV. Infection by HIV-1 induces cell cycle arrest at G2/M phase as it is essential for optimal expression of the viral genome. HIV-1 long terminal repeat (LTR) promoter is reported to be most active in this phase (Goh et al., 1998). One of the major determinants of HIV-1 induced G2/M arrest is the viral protein Vpr (Belzile et al., 2007; Goh et al., 1998). Vpr indulges with the host cell cycle to bring about G2/M arrest by engaging itself with a DDB1 and cullin4A containing ubiquitin ligase complex through the Vpr binding protein (VprBP), DDB1 and cullin4A associated factor 1 (DCAF1) (Belzile et al., 2007). DDB1 targets proteins for degradation through the DDB1-CUL4A-RBX1 E3 ubiquitin ligase complex (Groisman et al., 2003). Here DDB1 acts as a scaffold which presents substrates to the E3 ligase. Vpr interacts with this E3 ligase through the intermediate VprBP, DCAF1 to enable Vpr mediated G2/M arrest by activating an ATR-mediated check point signalling (Belzile et al., 2007) (Figure 1.39). As with the case of induction of G2/M arrest, the HIV-1 pathogenic activity of Vpr protein has also been attributed to induction of apoptosis in HIV-1 infection.
An introduction to 7-tOVandi proteasomal G2 cell cycle arrest virus replication/pathogenesis

Figure 1.39. Mechanism of Vpr-mediated G2/M arrest. Vpr exploits the CUL4A–DDB1–DCAF1 system to induce proteasomal degradation cellular factor, the absence of which causes cell cycle arrest in the G2 phase, promoting viral replication and pathogenesis in vivo. Adapted from (Schwefel et al., 2014).

Similarly another viral accessory protein, Vif, has also been shown to be involved in HIV-1 induced G2/M arrest (Izumi et al., 2010). Vif-mediated G2/M arrest occurs through the TP53 pathway. Vif blocks the MDM2-mediated ubiquitination and nuclear export of TP53 thereby enhancing the stability and transcriptional activity of P53. When TP53 is stabilized and activated by Vif, it leads to an up-regulation of p21 and simultaneous down-regulation of Cdc2 and CyclinB1 levels which culminates in G2/M arrest (Izumi et al., 2010) (Figure 1.40). Hence Vif-mediated G2/M arrest is distinct from that mediated by Vpr. The requirement of two different viral proteins for the purpose of induction of G2/M arrest substantiates the importance of this cellular modulation by HIV-1. Association of Vif with the MDM2 E3 ligase also results in the ubiquitination of Vif (Izumi et al., 2009).
1.14.2. Regulation of cell cycle factors during HIV-1 infection

The major cell types infected in vivo by HIV-1 or HIV-2 as well as other primate immunodeficiency viruses are CD4+ T lymphocytes and cells of the myeloid lineage which constitutes macrophages and dendritic cells. For successful replication in these cells, the virus has to outrun the antiviral action by multiple proteins of the host innate immune system. Although HIV-1 possess a relatively small genome of only 9.8 kb, the accessory proteins encoded by the virus counteracts many of these cellular factors. Many of these cellular factors are regulated by viral accessory proteins Nef, Vif, Vpr, Vpu and Vpx (for HIV-2 and SIV) (Rice and Kimata, 2015).

Apart from induction of G2/M arrest, some cell cycle associated genes have also been identified to play crucial role in HIV-1 pathogenesis. Cyclin T1 and its CDK partner CDK9, is one of the foremost Cyclin-CDK complex that was found to be involved in HIV-1 transcription regulation as they form a part of the PTEF-B
complex and act as a cofactor for HIV-1 Tat induced transcription elongation of viral genome (Imai et al., 2009; Wei et al., 1998). On the contrary, Cyclin K, acts as a competitor of Cyclin T1-CDK9 complex as it binds to CDK9 and restricts its nuclear translocation in a Nef-dependent manner thereby inhibiting HIV-1 gene expression and replication (Khan and Mitra, 2011). p21/CDKN1A acts as an intrinsic inhibitor of HIV-1 reverse transcription in CD4+ T cells of elite controllers. The phosphorylation of reverse transcriptase by CDK2 at Thr216 increases the efficacy and stability of reverse transcriptase as well as increases viral fitness (Figure 1.41). p21 levels are up-regulated in CD4+ T cells from elite controllers (Leng et al., 2014).

Cell-cycle associated genes; Cyclin A2 and Cyclin L2 have been identified to be involved in the evasion of the restriction function of SAMHD1 by HIV-1 in different cell types. Cyclin A2/CDK1 complex functions in cycling cells and is involved in the restriction of SAMHD1 activity in these cells by phosphorylating SAMHD1 and thereby inactivating it. Here, SAMHD1 interacts with cyclin A2/ckd1 and phosphorylates SAMHD1 at Threonine 592 residue (Figure 1.42).

Figure 1.41. p21 functions as an intrinsic inhibitor of HIV-1 - Adapted from (Leng et al., 2014)
This phosphorylation of SAMHD1 results in the loss of its ability to restrict HIV-1 (Cribier et al., 2013).

Figure 1.42. Cyclin A2 functions as HIV-1 dependency factor in cycling cells - Adapted from (Cribier et al., 2013)

Cyclin L2 functions in non-cycling macrophages as an HIV-1 dependency factor as it causes the degradation of SAMHD1 in these cells (Kyei et al., 2015) thereby enabling the virus to replicate efficiently. Both SIV and HIV-2 counteracts SAMHD1 activity using Vpx protein. In case of macrophages, despite the presence of SAMHD1, HIV-1 is able to propagate itself due to the rescuing activity of Cyclin L2 where Cyclin L2 and SAMHD1 form a molecular complex which is partially dependent on the presence of DCAF1. This complex then proteasomally degrades SAMHD1 in a DCAF1-dependent manner (Kyei et al., 2015; Wei and Yu, 2015) (Figure 1.43)
1.15. RATIONALE AND OBJECTIVES OF THE PRESENT STUDY

Science has progressed tremendously as compared to the period of initial discovery of HIV virus as the cause of AIDS. However, AIDS still continues to be a global pandemic leading to millions of death worldwide. Although a complete eradication of the disease has not been possible in the last three decades, research on HIV and its pathogenesis has significantly improved the life expectancy of AIDS patients with the development and use of antiretroviral drugs and HAART therapy. Many of these drugs inhibit the activity of different viral proteins. Antiretroviral drugs in clinical use, target the viral enzymes – reverse transcriptase, integrase or protease, which are essential for HIV replication. However, the low fidelity of HIV-1 reverse transcriptase results in high incidence rate of viral mutations leading to drug resistance. Hence, these therapies targeting viral proteins become less effective. Figure 1.44 depicts the increasing viral diversity with infection progression.
To overcome this limitation, current therapeutic approaches tend to exploit host cellular factors, those that have been identified to be involved in the regulation of viral pathogenesis. This, when used in association with the existing antiretroviral drug therapies can function synergistically to combat or limit HIV-1 replication/pathogenesis. Hence, this raises the need for identification of more cellular factors that could influence HIV pathogenesis. This approach is more helpful in combating HIV, considering the fact that cellular proteins do not undergo frequent mutations as compared to viral proteins.

As HIV is completely dependent on the host replication machinery for its own replication and propagation, it is highly likely that virus encounters several cellular factors associated with the cell cycle pathway which aids or obstructs viral replication. As a consequence of these associations, the virus induces cell death through different pathways including apoptosis. Few independent and isolated research has identified some factors associated with the cell cycle as important elements associated with viral pathogenesis. However, a more comprehensive study related to this might pave way for identification of more such genes that could be important determinants of HIV-1 pathogenesis. Considering this...
scenario, the current study aims at deciphering the functions of novel cellular factors associated with cell cycle and apoptotic pathway which may be involved in the regulation of viral pathogenesis. To execute these primary concerns, the following objectives have been laid for the present work:

1. Differential expression profile of cell cycle and apoptosis associated genes during HIV-1 infection.

2. Analysis of the expression profile and validation of selected differentially expressed genes involved in cell cycle and apoptosis.

3. Functional characterization of the selected cell cycle/apoptosis associated protein during HIV-1 infection.