# Literature Review

## CHAPTER 1

### Overview of HIV/AIDS

The United Nations Acquired Immune Deficiency Syndrome (UNAIDS) global report 2012 estimates that by the end of 2011 a total of 34 million people worldwide had been infected with Human Immune deficiency Virus (HIV) [UNAIDS 2012]. Though the report shows that the global trend of new HIV infections and HIV-related deaths per year is declining, the current number of HIV infections is the highest since 1990 (figure 2). There is a significant difference in HIV prevalence among the countries around the globe (Figure 1). The sub-Saharan region of Africa is the region most aggravated by the virus with 23.5 million people living with HIV. UNAIDS estimates that approximately 1 in every 20 adults are HIV infected in this region [UNAIDS 2012]. This is 25 or more times the HIV prevalence in any other region of the world. Countries in Sub-Saharan Africa also have varying HIV prevalence with South Africa at the top followed by Nigeria (UNAIDS 2011). The next severely affected regions besides African continent are Asia (China, Thailand, Indonesia), Caribbean and Eastern Europe, North America, western and central Europe.

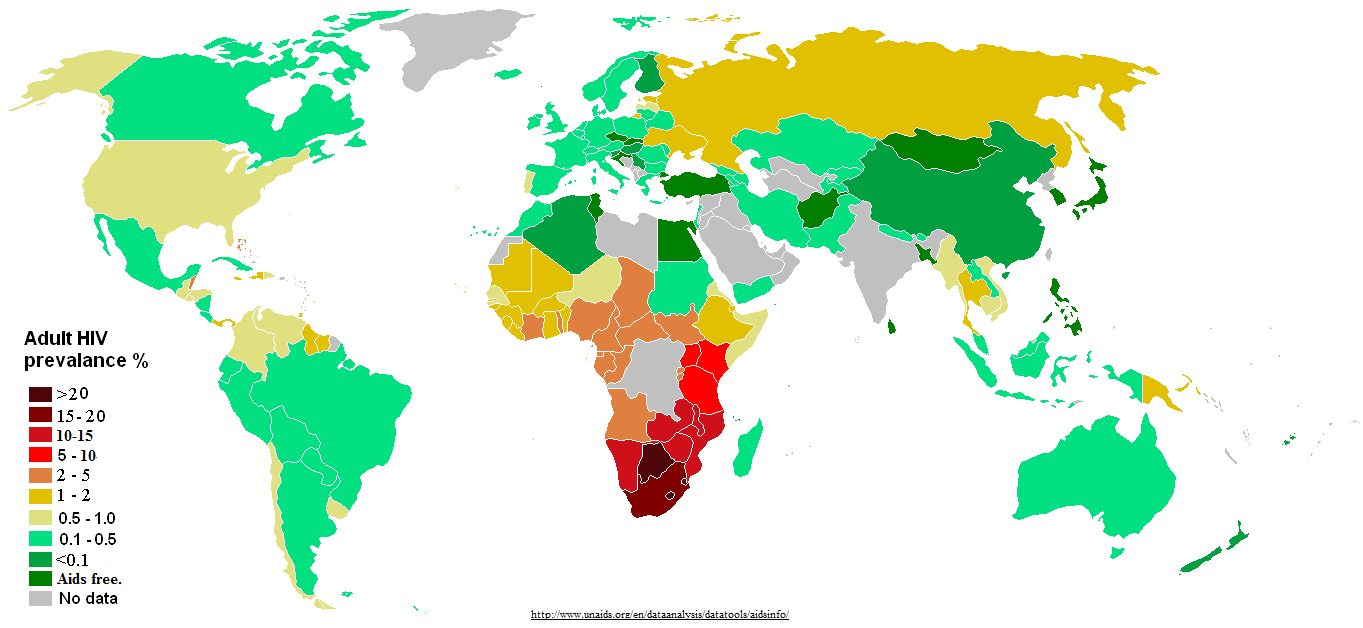


Figure 1: Distribution of HIV prevalence in adults around the world.

Source: http://www.unadis.org/en/dataanalysis/datatools/aidsinfo/

::Figures_for_thesis:hiv global trend.pdf

Figure 2: Global trend of new HIV infections from 1990 to 2011. A) Number of people living with HIV globally B) Number of people newly infected with HIV globally C) Number of Adults and child deaths due to HIV globally. Source: UNAIDS 2012

### Discovery of Human Immunodeficiency Virus

As early as 1959, HIV infection cases had been documented but were unreported. Curious doctors at that time collected patient blood samples and kept frozen, which were later were later shown to have HIV antibodies [Hedge 1983]. In June 1981, the first case of HIV infection was reported in the USA (Friedan-Kien et al 1981) when some homosexuals were showing opportunistic infections coupled with no lymphocyte proliferation [Gottlieb et al 1981]. Until 1983, the causative agent responsible for the severe immune depletion, named AIDS, was unknown [Francis et al 1983], when Luc Montagnier’s group at “Institut Pasteur” in Paris isolated the virus, which was initially named Human T-cell Leucamia Virus (HTLC) and later became Human Immunodeficiency Virus (HIV) [Hedge 1983]. Levy’s group in San Francisco, USA also subsequently found the virus confirming the discovery in Paris [Levy et al 1984]. They found HIV was a retrovirus; unique to any other previously isolated viruses and the viral transmissions were possible from infected to healthy people [wofsy et al 1986, Rogers et al 1987], mother to child through umbilical cord [Gallo et al 1983, Ziegler et al 1985].

Very soon, scientists around the world focused on this transmissible retrovirus. Complete sequencing of HIV genome [Ratner et al 1985] led scientists to know more insights of HIV including its origin, genes/proteins and life cycle [Wain-Hobson 1985]. HIV strain HXB2 (GenBank AC: K03455) became the reference sequence to aid in the viral research.

## Origin and Evolution of HIV

Exploration of the retrovirus led researchers to identify similarities between HIV and a retrovirus in African non-human primates that were then called as Simian Immunodeficiency Virus (SIV) [Gao et al 1999]. About 40 different primates, in Africa, are infected with SIV and some are harboring more than one strain of SIVs REF. Phylogenetic analysis of SIV from African non-human primates and HIV in human provided remarkable understanding of viral transmission as zoonotic (Bailes et al 2002) and evolution of the virus as HIV in human after transmission in to new host [Gao et al 1999].

Depending upon the source of non-human primate, from where SIV is derived, HIV is divided into two groups – HIV-1 and HIV-2 and that HIV-1 and HIV-2 transmission to human are independent [Sharp and Hahn 2010]. Discovered in 1986, HIV-2 is transmitted from Sooty mangabey monkeys (Cercocebus atys) [Hirsch et al 1989] and its prevalence is also high in the geographical location of these monkeys in West Africa (Santiago et al., 2005). Sooty mangabey monkeys are naturally infected by a strain of SIV (Hirsch et al. 1989). The phylogenetic analysis of all HIV-2 strains show that it closely groups with the SIVsmm strain (Geo et al 1992, 1994) that are non-pathogentic to its host monkeys. SIVsmm modified to produce multiple strain and subsequent multiple zoonotic transmissions from sooty mangabey monkeys to human (Hahn et al 2000) giving rise to different subtypes of HIV-2. Although HIV-2 subtypes A to G are identified in human, it is assumed that more subtypes were introduced into human (Gurtler 2004) but are lost for low adaptation fitness (Damond et.al 2004).

HIV-1 is the result of at least three cross species transmission events from chimpanzees (Pan troglodytes troglodytes (Ptt)) to human (Peeters et al. 1989; Huet et al. 1990). Phylogenetic analysis of HIV-1 sequences show that there was either divergent evolution of HIV-1 within human population or at three independent cross transmission of the virus in to the human population, each giving rise to three sub groups: group M (Major), group O (Outlier) and group N (Non M or Non O) [De Leys et al 1990, Simon et al 1998, Hu et al 1996, McCutchan et al 1999, Gao et al 1999, Hahn et al 2000, Santiago et al 2002].

Group M is the most prevalent and accounts for 98% of all infections [UNAIDS]. Its epicenter is thought to be Kinshasha of present day Democratic Republic of Congo [Sharp and Hahn 2008]. Molecular clock analysis of group M shows that its evolution dates back to 1920s (Korber et al 2000, Worobey et al 2008). By 1960, long before human came to know about its presence, it had already diversified substantially [Worobey et al 2008].

Group O and group N are rare and geographically confined to West African regions such as Cameroon and neighboring countries. It is still not understood about the non-pandemic characteristics of group O and N HIV-1 virus after the first zoonotic transmission [Arien et al 2005]. It has been suggested that reduced replication capacity and transmission fitness are keys to contributing their very low prevalence [Arien et al 2005]. Group O strain has 50% identity at env region of group M [Gurtler et al 1994, Vanden Haesevelde et al 1994] and the molecular clock model of this group also showed that its origins date back to 1920s [Charneau et al 1994, Lemey et al 2004]. Group N was supposed to be introduced into human population in 1960s [Simon et al 1998]. Phylogenetic analysis shows its close grouping with SIV from Chimpanzee. This indicates that group N might be a recombinant strain of two or more SIV in its host [Geo et al 1999, Corbet 2000].

Recently a new HIV -1 strain distinct from group M, N and O, classified as group P, has been discovered in a patient in Cameroon [Plantier et al 2009]. Group P is transmitted from gorilla as it is closely related to its SIV. Population level study of this new group virus shows its prevalence at 0.06% in the area but can still be pandemic as it can adapt in human [Vallari et al 2011].

### Structure of HIV

HIV is spherical in shape and the size is approximately 145nm [Briggs 2003]. The viral structure can be divided into outer viral envelope and the inner viral core (Figure 3).

The viral envelope consists of the outermost lipid bilayer membrane, derived from the host immune cell membrane during budding out from the cell. Several host protein molecules are embedded on the membrane. Besides these proteins, the viral envelope protein is also present on the surface. The envelope proteins form spike like structure that emerges out from the viral lipid membrane. Env consists of glycoprotein-120 (gp-120) that make trimeric structure with three gp41 stems that anchor the viral envelope (Zhu et al. 2006, Subramaniam et al. 2006, Zanetti et al. 2006, Zhu et al. 2008).

The inner core consists of matrix, capsid (p24 viral proteins) and nucleocapsid. Inside the capsid, there are two copies of the viral RNA molecules that codes for all the viral proteins and structures [Benjamin et al 2005].

### HIV genome and proteins – structures and functions

HIV has nine genes and produces 15 proteins in total [Frankel and Young 1998]. The genes are broadly categorized as accessory (vif, vpr, vpu and nef), structural (pol, gag and env) and regulatory (tat and rev). The higher number of proteins than genes is a result of post-transcriptional proteolysis of the products of the structural genes [Frankel and Young 1998].

**Accessory genes:**

Virion Infectivity Factor (Vif) is a protein of 220 amino acids and weighs 23 kiloDalton. This protein promotes the viral infectivity to the host, but has no role in viral production.

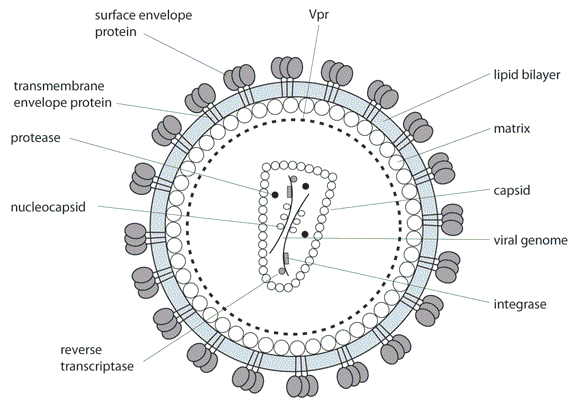


Figure 3: Structure of HIV

Vif is produced in the late stage of viral production [shreehy et al. 2002, Von Schwedler et al 1993] to suppress the innate antiviral activity of human immune cells [simon et al. 1998, Madani et al 1998]. Human APOBEC3 family members – APOBEC3G and APOBEC3F are the two potent cytidine deaminases that attributes to introduction of excessive G → A mutations in the minus strand reverse transcripts [Harris et al 2003, Mangeat et al 2003, Lecossier et al 2003, Zhang et al 2003]. Vif protein prevents APOBEC3 members from hypermutating HIV reverse transcripts [Conticello et al 2003, Marin et al 2003, Sheehy et al 2003, Stopak et al 2003, Mariani et al 2003, Mehle et al 2004, Wiegand et al 2004] and therefore the protein is only expressed while infecting APOBEC3G expressing immune cells [Navarro and Landau 2004].

Viral Protein R (Vpr) is 96 amino acids long protein (14 kiloDaltons) and is packed into the nascent virions during budding out [Connor et al 1995]. Vpr is essential for viral core to enter the host cell nucleus and thus localizes in the host cell nucleus after infecting the cell. Vpr arrests the cell cycle during the transfer from G2 to M phase [Jowett et al 1995, Rogel et all 1995] by preventing the activation of the p34cdc2/cyclin B complex [He et al 1995]. Vpr is also important for efficient viral replication in monocyte or macrophage cells, but T-cells [Connor et al 1995].

Viral protein unit or vpu is a protein unique to HIV-1 [Marassi et al 1999, Klimkeit et al 1990] and is 16 kiloDalton, 81 amino acids long protein. The Env precursor protein, gp160, makes stable intracellular complex with host cell CD4. Vpr prevents the formation of env-CD4 by degrading CD4 at cell surface, which allows gp160 processing to gp120 and gp41 [Willey et al 1992]. The vpu action of CD4 degradation could be blocked with peptide aldehyde or lactacystin, by interfering proteasome function of the protein [Schubert et al 1998]. Another role of vpu is interference of host immune cell MHC class II antigen presentation on the cell surface allowing the virus for host immune escape [Hussain et al 2008, Nomaguchi et al 2008]. Vpu also plays role in viral release from host cell membrane [Klimkait et al 1990].

Nef has no role in viral infectivity but plays a role during the biogenesis of viral particles [Laguette et al 2009] and virulence [Piguet et al 1999, Prince et al 2002, Kirchhoff et al 2008]. Nef down regulates the production of major histocompatibility complex type 1 (MHC type 1) in the host cell [Schwartz et al 1996, Collin et al 1998, Cohen et al 1999]. This impairs the function of cytotoxic T lymphocyte cells to clear the infected cells [Collins et al 1998, Tomiyama et al 2002, Yang, et al 2002, Adnan et al 2006]. Nef also down regulates CD4 on host cell surface [Garcia et al 1991, Lama et al 1999] and modulates cellular activation to evade host immune system [Baur et al 1994, Sawai et al 1994, Bodeus et al 1995, Saksela et al 1995, Smith et al 1996].

**Structural genes and proteins**

The Gag gene produces a precursor polyprotein of ~ 500 amino acids long and weighs 55 kilodalton. The Gag precursor (pr55gag) has all the building blocks to form a fully infectious virion, even in the absence of other viral products [Wang et al 1993]. The viral protease enzyme cleaves the gag precursor to yield the structural proteins – matrix, capsid, nucleocapsid and p6 (Figure 4) [Gheysen et al 1989, Hunter 1994, Coffin, Swanstrom and Wills 1997, Freed 1998,]. The cleavage takes place in the nascent virus after budding out from host cell [Gottlinger, Sodroski and Haseltine 1989]. The matrix protein is always at the N-terminal and p6 at the C-terminal of gag precursor [Mervis et al 1988, Henderson et al 1992].

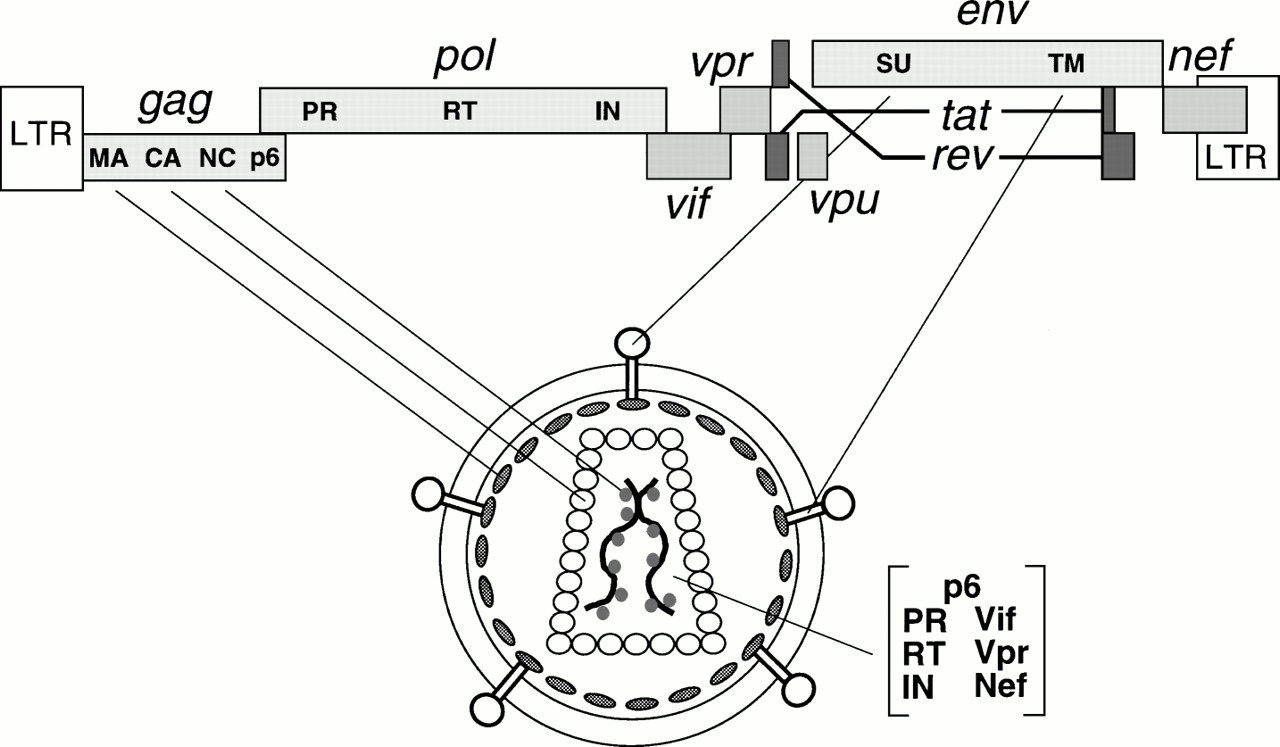


Figure 4: HIV genes and proteins positions in the viral genome and their viral parts. Source: Frankel and Young 1998

In the HIV life cycle, matrix domain of gag plays role in targeting gag precursor to the plasma membrane of the host cell and the viral assembly at the site [Gelderblom et al 1987, Bryant and Ratner 1990, Gottlinger, Sodroski and Haseltine 1989]. The highly basic region in matrix from gag mediates electrostatic association with phospholipids in plasma membrane during assembly [Zhou et al 1994, Hermida-Matsumoto 2000, Ono et al 2000, Ono and Freed 2004]. The capsid is a curved and closed shell consisting 250 hexamers and 12 pentamers [Pornillos et al 2009]. Capsid packs viral proteins and nucleocapsid to pass on to new HIV particles [Gelderblom 1991]. Nucleocapsid plays role in efficient viral assembly by making pr55 –pr55 interprotein contacts [Zhang et al 1998, Dowson and Yu 1998]. P6 protein plays role in detaching and releasing the newly formed HIV particles [Gottlinger et al 1991].

The pol gene is approximately 1000 nucleotides long and as much as 225 nucleotides at 5’ region of the gene overlaps with 3’ region of gag gene REF. Ribosomal frame shift at gag opens reading frame for pol gene to produce a pol precursor. Protease enzyme, packed in nascent virions, cleavages of pol precursor produces highly essential enzymes – protease (PR), reverse transcriptase (RT) and integrase (IN) for viral replication (Figure 4). The 3D structures of these enzymes have been elucidated for anti HIV drug development. The drugs developed bind specifically to interfere the functions the enzymes. The protease enzyme cleaves the gag and pol polyprotein to form the viral structure proteins and functional enzymes respectively [Liang et al 1988, Ross et al 1991]. The reverse transcriptase enzyme produces the cDNA from the viral RNA after infecting host cell [Jacobo-Molina and Arnold 1991, Sarafianos et al 2009]. The RNase H domain in RT degrades the RNA molecule following cDNA production [Davies et al 1991]. The integrase enzyme integrates the proviral cDNA into the host genome [Pruss et al 1994].

Env produces a precursor glycopolyprotein (gp160) that is processed by human convertase enzymes - PC1 and furin to produce glycoprotein 120 (gp120, HIV-1 SU) and glycoprotein 41 (gp41, HIV-1 TM) [Decroly et al 1994]. Gp120 is a non-covalent complex of external protein and gp41 is a trans-membrane protein; both play vital role for initial steps in viral infection [Chan et al 1997]. Three gp120 molecules bound with three gp41 molecules to form envelope spikes [Zhu et al 2006, Lui et al 2008]. They are organized to form trimeric complexes on the surface of HIV and mediate HIV entry into the host cell [Allan et al 1985, Robey et al. 1985, Veronese et al 1985]. The exposed external complex gp120 binds to the CD4 receptor on the host immune cell. This triggers a conformational change that enables binding to a chemokine receptor, usually CCR5 or CXCR4, to facilitate fusion of the viral and target cell membranes.. [Dalgleish eta al 1985, Feng et al 1996, Deng et al 1996, Choe et al 1996, Dragic et al 1996]. The gp120 glycoprotein has been an interest for HIV antibody development, as it is exposed, to neutralize and impair gp120 binding to the host cell receptors [Wyatt et al 1998, Pantophlet and Burton 2006]. Three molecules of trans-membrance glycoprotein gp41 form a six-stranded helical rode structure [Buzon et al 2010]; three from N terminal coiled coil structure and three from C terminal hydrophobic groove [Tan et al 1997]. Gp41 is responsible for viral fusion and fetching the viral contents in to the host cell [Furuta et al 1998, Melikyan et al 2000]. Anti HIV drug development is underway targeting the heptad region of gp41 [Kilby et al 1998, Eckert et al 1999, Lui et al 2007].

**Regulatory Genes:**

Tat is a trans-activating factor for HIV gene expression. After infection of host cell, the HIV proviral genome is integrated in to the host genome, which is regulated by cellular as well as the viral transcription regulatory factors. Tat is the primary transcriptional regulatory factor. An example of Tat action is the control of RNA polymerase II elongation during transcription. In the absence of Tat, polymerase II disengages from the template DNA strand, terminating the transcription prematurely (Kao *et al*., 1987; Kessler & Mathews, 1992; Ratnasabapathy *et al*., 1990; Toohey & Jones, 1989). Sodroski *et al*. (1985) first explained the function of Tat.

Rev is another trans-activating factor for HIV gene expression. It is 19-kD phosphoprotein. Like Tat it is also mainly localized in the nucleus of host cell but cycles rapidly between the nucleus and cytoplasm as it promotes nuclear export of the transcriptional products. Rev binds at the Rev Responsive Element (RRE), which is an RNA element encoded within the env region of the virus.

### HIV replication

The sole purpose of HIV is replication in the host immune cell mainly C4+ expressing T-cells and macrophages. The major events of HIV replication cycle with drug targets are depicted in the Figure 5. The mechanism of the viral entry into the immune cell involves HIV surface glycoprotein gp120 binding to the CD4 receptor of the host cell [Dalgeish et al 1984, Maddon et al 1986, McDougal et al 1986]. This changes conformation in gp120/gp41 complex [Sullivan et al 1998] enabling binding to a chemokine receptor (generally CCR5 or CXCR4) [Sattentau et al 1991, Moore et al 1992, Sattentau and Moore 1993, Berger et al 1999]. The binding triggers gp41 for the membrane fusion [furuta et al 1998]. The gp41 changes its state from native non-fusogenic to pre-hairpin intermediate stage [Lu et al 1995, Chan et al 1997, Finzi et al 2010] and exposes its six helical strands of heptad regions HR1 and HR2 (drug target, labeled a in figure 5) at ectodomain [Furuta et al 1998, Koshiba and Chan 2003] ready to fuse viral membrane with cell membrane. Further, gp41 change confirmation to low energy trimeric hairpin to facilitate the fusion process [Chan et al 1997, Tan et al 1997, He et al 1998, Melikyan et al 2000].

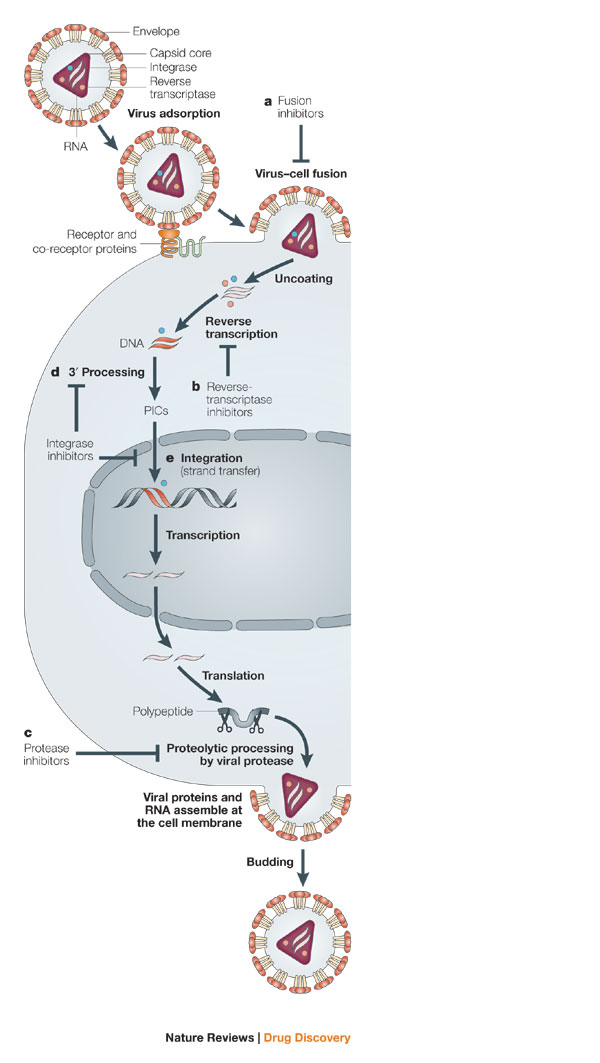
Upon fusion, a series of events occur in the journey of viral genome to the cell nucleus. The capsid is released in to the cytoplasm following fusion. McDonald et al [2002] were able to track the HIV virion pathway to the nucleus using GFP-labeled particles. The study showed that the reverse transcription process starts in the intact capsid and the required deoxynucleotides are imported to reverse transcribes RNA to cDNA by RT enzyme (drug target, labeled b in figure 5) [Cullen et al 2001]. Subsequent uncoating of capsid releases HIV cDNA genome. The capsid and nucleocapsid proteins dissociate from cDNA but the reverse transcription complex remains intact along with viral matrix, integrase and human protein high mobility group I (HMG I (Y)) forming preintegration complex (PIC) [Miller et al 1997]. The PIC protects cDNA from endonuclease degradation [Miller et al 1997]. Vpr protein plays role in ATP dependent [Bukrinsky et al 1992] transportation of PIC on host microtubules towards the nuclear membrane [McDonald et al 2002]. Formation of central DNA flap [Zennou et al 2000], vpr and matrix play critical role in importing the PIC in to the nucleus through nuclear pore [Popov et al 1998a, 1998b].

Figure 5: Major events in HIV replication. The events labeled a – e are the possible target points for disrupting HIV replication cycle with development of antiviral drugs. Source: Pommier et al 2005

Inside the nucleus, the integrase enzyme carries 3’ processing reaction (drug target, labeled d in figure 5) [Hawkins et al 1995] and strand transfer (drug target, event e in Figure 5) of viral cDNA to integrate in to host genome [LaFemina et al 1992]. Host transcriptional co-factor LEDGF/p75 and HIV integrase interact to tether to the host chromosome during the integration process [Emiliani et al 2005].

The integrated HIV provirus hijacks the host cell transcriptional machinery for viral genes to transcribe [Davey et al 2011]. HIV protein Tat promotes the transcription of the viral DNA [Ott et al 2011, Razooky and Weinberger 2011]. The viral transcriptome encodes structural proteins, accessory proteins and viral enzymes necessary for a complete functional HIV.

The viral Rev protein facilitates exporting of the unprocessed viral transcriptome to cytoplasm for translation [Malim et al 1989]. The host translational machinery is exploited for translation of the viral transcriptome.

Upon translation of all viral proteins, initiation of HIV virion assembly occurs at the cell membrane. The viral assembly process is mediated by the gag polyprotein [Freed 1998]. The complete assembled virion particles are released from the plasma membrane by the host ESCRT machinery involving Tsg101 and ALIX regulatory proteins [Fujii et al 2007, Schwedler 2003, Chen and Lamb 2008, Slagsvold et al 2006, Usami et al 2009]. The maturation of the nascent HIV virions begins concomitantly with budding out. The PR enzyme (drug target, labeled c in figure 5) cleaves Gag and Pol polyproteins in the maturation step [Debauck et al 1987, Karacostas et al 1993, Erickson-Viitanen et al 1989, Viegers et al 1998, McQuade et al 1990]. The processed polyproteins produces matrix, capsid, nucleocapsid, p6, protease, reverse transcriptase and integrase proteins [Hill et al 2005] that are rearranged to create mature HIV, ready to new cells [Mariani et al 2003].

Each HIV replication cycle releases new infectious virions in the order of 109 per day [Ho et al 1995]. The number of new HIV infecting new cells determines the replication rate of the virus [Tersmette et al 1989]. With increasing time period of infection, higher turnover rate is observed in both immune cell depletion and viral population expansion [Ho et al 1995]. The measure of low level of CD4+ and higher viral load per ml blood are considered clinical parameters for testing viral infection progress and indicate time point to enroll in anti retroviral therapy or change drug regimen in the therapy. The error prone reverse transcriptase, with mutation rate of order 10-5 per base per cycle [Mansky and Temin 1995], is responsible for generating huge number of HIV variants in the viral population. These HIV variants with very small genetic differences are called HIV quasispecies. Genetic variations in HIV quasispecies add difficulty in anti retroviral drug design and therapy.

### HIV subtype, sub-subtype and Circulating Recombinant Forms

HIV-1 group M is classified into nine subtypes: A, B, C, D, F, G, H, J and K (Figure 6) [Robertson et al 2000]. The subtype classification is based on the phylogenetic and sequence distance analyses forming major clades (Figure 6). The classification of new subtype should also follow the same rule as “roughly equidistant from all previously characterized subtypes in all regions of the genome with a distinct pre-subtype branch similar to those of other subtypes” [Robertson et al 2000]. Subtypes can be further classified to sub-subtype if they form a distinct sister clade [Gao et al 2001] within a clade with the same rule of phylogenetic and distant analyses but not justifiable to call a subtype due to low genetic distance [Robertson et al 2000]. Only subtypes A and F exhibit distinct sister clades. Subtype A has sub-subtypes A1 and A2 (A3 and A4 are mentioned by Taylor et al 2008); Subtype F has sub-subtypes F1 and F2 (Figure 6). The prevalence analysis from sequence data in 2004 showed that “subtype C accounted for 50% of all infections worldwide” [Hemelaar et al 2006]. Prevalence of other subtypes A, B, G and D was found in decreasing order 12%, 10%, 6% and 3% respectively. Subtypes F, H, J and K were found very scarce and on overall, accounts for only 0.94% infections [Hemelaar et al 2006]. Analysis of genetic sequence variation within a subtype and between subtypes showed 15%– 20% and 25% - 35% respectively [Korber et al 2001, Hemelaar et al 2006]. The sequence variation within and between the subtypes displays phenotypes like viral tropism (CCR5 or CXCR4 or dual tropism) [Reeves et al 1999, Taylor et al 2008], differential cellular and antibody response [Barouch 2008], and viral fitness and drug resistance development. The divergence in HIV subtypes is also the major challenge in diagnosis, treatment and antibody based vaccine development [Gaschen et al 2002, Barouch 2008, Lihana et al 2012].

In correlation to very low prevalence and geographical confinement to Western African, the diversity observed in Group N, O and P is very less. It is still unclear if group O can be sub divided into subtypes [Roques et al 2002, Yamaguchi et al 2002]. Group N as well does not show distinct sub clade within itself (Figure 7).



Figure 6: Phylogenetic tree showing HIV-1 group M diversification to subtypes A-D, F-H, J and K, inferred from nucleotide sequence alignment of gag, pol and env genes.

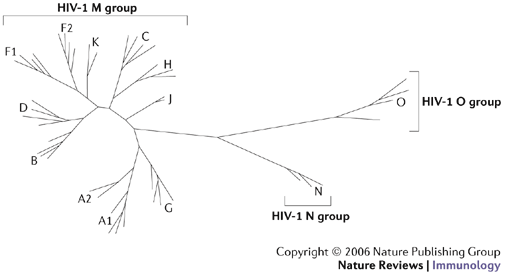


Figure 7: Neighboring joining phylogenetic tree showing HIV-1 group M, N and O. Group M shows distinct nine subtypes A-D, F-H, J, K while no specific subtype is observed in group N and O. Source: Letvin 2006

Recombination between different HIV-1 strains is a common event REF. There are different breakpoints in HIV genome for recombination [McCutchan et al 2002, Archer et al 2008, Kosakovsky Pond et al 2009]. When a single immune cell is dual infected by two different strains of HIV, the recombinant forms of HIV can arise. Recombination can occur between same subtype and different subtype (inter-subtype) (Table 1) [Fang et al 2004, Robertson, Hahn and Sharp 1995]. Different forms of recombinants are circular recombination forms (CRFs), unique recombinant form and inter-subtype recombinant forms (table 1). CRFs account for about 18% of all HIV infections [Hemelaar et al 2006].

|  |  |  |
| --- | --- | --- |
| Classification | Definition | Examples |
| Inter-subtype recombinant forms | Mosaic strains with segments from two or more subtypes alternating across the genome | Common in mixed subtype epidemics; though to result from infection of a person with more than one HIV-1 subtype |
| Circulating recombinant forms | Specific recombinant forms that are spreading in a population; new forms are defined when three people without direct epidemiologic linkage are found to be infected; the assigned name reflects sequence of discovery and subtype composition, with ‘cpx’ indicating forms containing three or more subtypes | Currently, 43 forms are described; CRF01\_AE and CR02\_AG are found principally in Southeast Asia and West Africa, respectively; others have more limited distributions |
| Unique recombinant forms | Inter-subtype recombinant forms recovered from only a single person | Hundreds of forms have been described on the basis of partial or complete genome sequences; their potential for epidemics spread is unknown |

Table 1: Phylogenetic classification of HIV-1. Source: Taylor et al 2008