3. Review of Literature

3.1 HIV-1 discovery

In February 1982 at a Cold Spring Harbour Conference on AIDS, Robert C Gallo put forward the idea that Acquired Immunodeficiency Syndrome (AIDS) was likely to be caused be a HumanT-Lymphotropic Virus (56). In 1983, there were two independent reports of the isolation of a retrovirus from individuals with signs and symptoms of AIDS by Robert Charles Gallo from the National Cancer Institute, Bethesda, Maryland, USA and Luc Montagnier, Pasteur Institute, Paris, France (57, 58). The isolation followed presentation of patients with illnesses like *Pneumocystis carinii* and Kaposi’s sarcoma in 1981 (59-61). The affected individuals included drug abusers and homosexuals. In 1982, there were reports of haemophilia associated AIDS in the United States (62). Laboratory tests on the affected individuals revealed a perceptible defect in the cell mediated immune mechanisms (62). In 1984, AIDS was also reported among children and infants born to infected mothers (63, 64).

The virus was initially referred to as HumanT-Lymphotropic Virus Type –III (HTLV-III) by R.C. Gallo and Lymphadenopathy Associated Virus (LAV) by L. Montagnier, the two groups that had isolated the virus (57, 65, 66). The International Committee on the Taxonomy of Virus (ICTV) named the virus as the Human Immunodeficiency Virus (HIV) in 1986 (67). HIV is now recognized to consist of two species HIV-1 and HIV-1 (68). There are significant differences both in the nucleotide and amino acid sequence levels which are reflected in the differences in the biological properties among the two viruses (4).
3.2 Origin of HIV

Reports of the incidence of AIDS among humans in 1982 were followed by identification of simian AIDS in primate centres among Asian rhesus macaques (69, 70). The simian immunodeficiency virus (SIV) was later isolated from the primates which was concomitant with demonstration of cross reactivity of the sera from the infected primates to HIV-1 antigens by Kanki et al (71, 72). It has been subsequently shown that the HIV1 and 2 were transmitted to humans through cross species transmission of non-human primate lentiviruses (73-75). While HIV-1 entered humans from chimpanzee (*Pan troglodytes*) through the introduction of the SIVcpz, HIV-2 was transmitted from sooty mangabey (*Cercocebus atys*) (73-75). The chimpanzee has been the source of at least 3 separate inductions of the virus into human population whereas HIV-2 is the result of at least 4 transfers (73, 76).

Analyses using molecular clock to calculate ancestral dates suggests that last common ancestor of HIV-1(group M) was in 1930 (77). The earliest known case of AIDS was believed to be from a sailor in Manchester reported by Corbitt et al (78). Based on their findings the authors dated the first case of HIV infection to 1959. However these findings were later retracted by the authors (79). The earliest confirmed report of HIV-1 infection based on serological and molecular evidence from a plasma sample that originated in Africa and dated to 1959 (80). The rapid spread of HIV transmission to an epidemic scenario in the last decade has been hypothesized to three factors ‘i) social changes accelerated sexual transmission ii) health care changes accelerated parenteral transmission and iii) serial passaging adapted HIV for persistent infection and sexual transmission’ (81).
Theories have been put forward to explain the possible routes of zoonotic transmission of HIV to humans. Wolfe et al have reported the natural transmission of the simian foamy virus (SFV), a retrovirus, among natives of central Africa. This transmission was associated with direct contact with blood and body fluids of wild non-human primates during activities like hunting (82). The authors suggest that their findings could indicate these routes also provide opportunity for cross species transmission of other retroviruses. A theory known as the ‘OPV/AIDS theory’ has also been suggested implicating primate tissue used in preparation of oral polio vaccines in the transmission of HIV to humans. This theory has however been refuted (83, 84). A new theory proposed by Moore et al suggests that the interaction between colonial practices like labour camps and non sterile vaccination campaigns with traditional bushmeat hunting in French Equatorial Africa maybe the basis for the origin of the disease (85).

3.3 Biology of HIV-1

3.3.1 Classification

The Human Immunodeficiency virus type 1 belongs to the family *retroviridae* (68). Retroviruses are so termed because of the presence of the enzyme reverse transcriptase (RT) (86). The RT transcribes the viral genomic RNA into DNA (86). The DNA motif is integrated in a proviral state by the integrase enzyme of the virus (41). Retroviruses are classified into two subfamilies and 7 genera as follows (68):
Family: Retroviridae

1) Subfamily: Orthoretrovirinae which has been further classified into six genera
   Alpharetrovirus, Betaretrovirus, Gammaretrovirus, Deltaretrovirus, Lentivirus and
   Epsilonretrovirus.

2) Subfamily: Spumaretrovirinae which has one genus Spumaretrovirus

The Human Immunodeficiency virus type 1 is the type species of the genus Lentivirus (68).
The Human Immunodeficiency virus type 1 and 2 belong to the primate lentivirus group of
the genus Lentivirus.

3.3.2. Overall structure

A mature HIV-1 consists of the following structure

1) Envelope: The envelope is derived for the host cell during budding from the host
cell(41). The outer layer of HIV-1 is characterized by mushroom shaped glycoprotein
which has a head of molecular weight 120 KD and is termed as gp120(87). The
transmembrane stalk is known as gp 40 and attaches the gp120 molecule to the
membrane of the virus (88).

2) Core: The central part of the HIV-1 is an electron dense conical component. The core
contains RNA, enzymes and the tRNA lys 1,2 primer (41). The tRNA lys 1, 2 serves
as the template for RNA replication. The enzymes contained in the core are reverse
transcriptase, protease and integrase (41). The RNA which is the genome material of
HIV-1 is represented as two single stranded copies of RNA to which are bound ribonucleoproteins (41).

3) Matrix: The electronlucent region between the core and the envelope is known as the matrix (41). In a mature viral particle it lines the inner surface of the virion membrane (89).

### 3.3.3 Viral genes and products

The HIV-1 RNA genome which is represented by two single stranded copies is ~ 9kb in size and genomic organization is shown in Fig.1 (89). The proteins encoded by the HIV-1 genome can be classified as structural, regulatory and accessory proteins (41, 89). The genes encoding for the structural proteins are *gag*, *pol* and *env*. Besides this the genome also encodes for two regulatory proteins (Rev and Tat) and four accessory proteins (Vif, Vpr, Vpu and Nef) (41, 89).

The HIV-1 genome is also flanked by a 630 bp sequence known as long terminal repeats (LTR) at both the 5’ and 3’ region of the genome (41, 90). The functions of LTR are highly varied. They are needed for synthesis of viral DNA from RNA, integration of viral DNA, regulation of viral RNA synthesis from proviral DNA and HIV-1 viral RNA packaging during formation of new viral particles (41, 90, 91). The LTR can be divided into three regions: the U3 region, the R region and the U5 region (41, 90).
Fig. 1: HIV-1 Genome and its structural components

This Figure shows the genomic organization of HIV-1 (top) and a mature HIV-1 particle (bottom). This Fig. was modified from Frankel et al (89).
3.3.3.1 Structural Proteins

The structural proteins are synthesized as a large precursor protein which is further broken into smaller functional components by the HIV-1 proteases.

3.3.3.1.1 Gag

The precursor protein of gag denoted as Pr \text{55}^{\text{gag}} is a 55 kDa protein. This is subsequently broken down into the functional proteins by the protease enzyme into a 17 kDa matrix (p17), 24 kDa capsid (p24), 7 kDa nucleocapsid and smaller proteins denoted as p1, p2 and p6 (41, 92).

The matrix (MA) is derived from the N-terminal component of the gag polyprotein and is composed of 132 residues(89). The MA is involved in transferring the gag and gag-pol precursor proteins for viral assembly in infection of cells. The MA also transports the envelope glycoprotein for incorporation in the viral particle (89, 93, 94). The core of HIV-1 is formed by approximately 2000 molecules of capsid (CA). The nucleocapsid (NA) is associated with the genomic RNA and play an important role in delivering the RNA into the viral particle during assembly (89).

3.3.3.1.2 Pol

The \textit{pol} encodes for the reverse transcriptase (RT), protease (Pr) and integrase (In). The \textit{pol} is transcribed by an unspliced mRNA which yields a large gag-pol (Pr \text{160}^{\text{gag-pol}}) fusion protein (90, 95). The gag and pol polyprotein are encoded by the same mRNA with a frameshift
mutation leading to the production of Pr $^{55}\text{gag}$ and Pr $^{160}\text{gag-pol}$ (90, 95, 96). The Pr $^{160}\text{gag-pol}$ is cleaved to yield RT (p66/p51), Pr (p10) and IN (p32).

### 3.3.3.1.2.1 Reverse Transcriptase

The p66 homodimer is cleaved to yield a heterodimer consisting of the p66 and the p51 subunit. The functional HIV-1 RT thus consists of the p66 and the p51 subunit (97, 98). The p51 subunit consists of 440 amino acids while the p66 consists of an additional 120 amino acids to contain a total of 560 amino acids (97, 98). The p66 contains the functional component of the HIV-1 RT with the additional 120 amino acids constituting the RNase H activity of the RT (38, 99). The p51 contains no catalytic activity and functions more as a scaffold for the p66 subunit (38, 99).

The ‘anatomical’ shape of the p66 subunit has been likened to that of a ‘right hand’ and is composed of the ‘fingers’ (amino acid residues 1-85, 118-155), ‘palm’ (86-117, 156-237) and ‘thumb’ (238-318) (37). The ‘palm’ sub domain is the location of the catalytic aspartate residues (D110, D185, D186) (37). In addition, it also has the ‘connection’ (319-426) and the RNase H (37). The subdomains of p66 are show in Fig. 2. The p51 has a compact structure which is referred to as the ‘closed’ conformation while in comparison the p66 has an ‘open’ conformation (37, 100).
Fig. 2: The polymerase domains of the catalytic p66 subunit of the HIV-1 RT

The ‘anatomical’ shape of the p66 subunit has been likened to that of a ‘right hand’ and is composed of the ‘fingers’ (amino acid residues 1-85, 118-155), ‘palm’ (86-117, 156-237) thumb’ (238-318) and the ‘connection’ (319-426) domains. The catalytic residues are present in the ‘palm’ sub domain. This Fig. has been modified from Kohlstaedt et al., 1992 (37).
The template binds along a groove which is present between the active site of the polymerization component and the RNase H active site of the RT (37). In the complex of the dsDNA template-primer and HIV-RT the DNA duplex has an A-like structure near the polymerization active site and a B-like structure near the RNase H active site thus giving the DNA a bend appearance while present in the binding groove (101). During binding of the template to the RT there is movement of the ‘fingers’ bringing the ‘finger tips’ towards the palm region. The ‘thumb’ along with ‘finger’ subdomains serve as clamps holding the template in position (101, 102). When the incoming dNTP binds to the growing strand there is an apparent conformational change in the RT leading to chain extension (103). It has been postulated that after addition of one dNTP, the RT moves to the original ‘open’ conformation allowing for binding of the next dNTP (102).

### 3.3.3.1.2.2 Protease

The HIV-1 protease produced by cleavage of the Pr 160\textsuperscript{pap-pol} contains 99 amino acid residues (104). HIV-1 protease is a homo-dimer with each monomer consisting of 99 amino acids (40). The HIV-1 protease is in the shape of an oblate ellipsoid with dimensions of 55 by 35 by 25 Å\textsuperscript{3} (40). It is an aspartyl protease whose catalytic activity can be inhibited by pepstatin, \textit{in vitro} (105, 106). The amino acids Asp-Thr-Gly at positions 25 - 27 from each monomer constitutes the active site at the dimer interface with substitution of Asp at position 25 leading to complete inhibition of the protease activity (40, 107, 108).
In the absence of protease only immature virions are produced that are incapable of completing the replicative cycle (107, 109). The HIV-1 protease is responsible for post-translational modification of gag and gag-pol polyprotein (41). The hydrophobic substrate cleft recognizes and cleaves 9 different peptide sequences to produce the matrix, capsid, nucleocapsid and p6 proteins from the gag polyprotein and protease, RT and integrase proteins from the gag-pol polyprotein (41). The initial cleavage of gag-pol precursor polyprotein is an autocatalytic process brought about by dimerization of the gag-pol polyprotein (110). The autoprocessing is negatively influenced by p6 and positively influenced by the nucleocapsid domain (110).

The HIV-1 protease structure can be divided into the following substructures ‘fulcrum’ (amino acid positions 10-23), ‘flap elbow’ (amino acid positions 35-42), ‘flap’ (amino acid positions 43-58), ‘flap tip’ (amino acid positions 49-52) and ‘cantilever’ (amino acid positions 59-75) based on the ‘cantilever-fulcrum’ model of flap opening 61. The substructures of the HIV-1 protease are shown in Fig. 3a. The active site of the protease is covered by two β hairpin structures or ‘flaps’ which serves as clamps (42, 43). The protease is said to be in an ‘open’, ‘semi-open’ or ‘closed’ conformation depending on the positions of these flaps (42). When the substrate is bound to the active site of the protease molecule the enzyme is in a closed conformation (42). The opening of the protease is associated with a concert of movements leading to download motion of the cantilever, fulcrum and flap elbows and upward movement of the flap. This leads to the upward movement of the active site which could be to facilitate ligand binding (111). In the unbound state the HIV-1 protease is in equilibrium in the ‘semi-open’ state (111). The positions of the flaps in the ‘closed’ and ‘semi-open’ are shown in Fig 3b.
Fig. 3a: The HIV-1 protease and its substructures


Fig. 3b: The positions of HIV-1 protease flaps

i) Closed

ii) Semi-open

Change in ‘handedness’ of the flap tips during transition from a closed (i) to a semi-open (ii) conformation. The figure was modified from Hornak et al (2007)
3.3.3.1.2.3 Integrase

HIV-1 integrase is plays an important role in the integration of HIV-1 viral DNA into the host chromosome (41). The HIV-1 integrase consists of a catalytic domain which is between the N-terminal and C terminal. This division is based on the susceptibility of the integrase to proteolytic enzymes and functional studies (112-114). The Aspartic acid at position 64 and 116 along with glutamic acid at position 152 in the catalytic domain constitute important residues needed for catalytic activity of HIV-1 integrase (112).

The first step in integration of the viral DNA is known as 3’-end processing and involves exposing the 3’- hydroxyl group that is to be joined to the target DNA(115). The first step involves removal of 2 nucleotides from the 3’ end and this is followed by the DNA strand transfer. This involves incorporation of the viral DNA into the host DNA (115). Substitution in the catalytic domain of integrase leads to production of HIV-1 viral particle that are unable to infect T-lymphoid cell lines in vitro (116, 117).

3.3.3.1.3 Env

The env gene encodes for the gp120 and gp41 subunits which are present on the outer membrane of HIV-1. These proteins are translated from a singly spliced mRNA which is initially translated to an 88 kDa protein(89). This protein is further glycoslated in the endoplasmic reticulum of the infected host cell to produce a 160 kDa (gp160) protein(89). This larger precursor protein (gp 160) is subsequently cleaved by cellular enzymes to produce the functional gp 120 and gp 41 subunits(41, 89). The gp 120 and gp 41 subunits are then transferred to the cellular membrane to be incorporated into the budding HIV-1 particle(41).
The gp 120 is anchored to the viral membrane through the transmembrane (TM) subunit gp41 (117, 118).

There are 5 conserved (C1 to C5) and variable (V1 to V5) domains in the gp120 subdomain (41). The gp 120 domain is involved in binding of the HIV-1 to CD4 receptors. This binding to CD4 receptor is mediated by the V3 region of gp120 (119, 120). The gp41 subunit undergoes conformational changes following binding of gp120 with the CD4 receptor resulting in fusion of the virus with the host cell (121).

### 3.3.3.2 Regulatory Proteins

#### 3.3.3.2.1 Tat (Transcriptional Transactivator)

The HIV-1 Tat plays an important role in transcription of the HIV-1 LTR (122). The Tat is unusual as it binds to a RNA rather than a DNA molecule (89). A recent study has shown that HIV disease progression can be influenced by the subtype (123). Studies have also shown that there is a variation in levels of Tat transactivation among the different subtypes (124, 125).

The tat mRNA is a multiply spliced mRNA and consists of two coding (1 and 2) and one noncoding exon (90). Mutational analysis studies have shown that Tat protein can be functionally organized into different domains(126). The functional domains of Tat include: N-terminal domain (amino acid positions 1-20), Cys-rich domain (amino acid positions 21-40), Lys X Leu Gly Ile X Tyr motif (amino acid positions 41-48), basic domain (49-57) and
the auxiliary domain (amino acid positions 58-67). Studies have shown that alteration of even one of the domains can affect the proper functioning of the Tat protein (126). The Cys-rich domain has been suggested to be important for protein dimerization whereas the basic domain contains the arginine-rich RNA binding motif (ARM) and acts as a nuclear localization signal (NLS) (90). The auxiliary domain is believed to contribute to Tat activity by structural stabilization or by direct functional contribution (126).

3.3.3.2.2 Rev (Regulator of viral protein expression)

The Rev is a 116 amino acid protein that helps in transportation of unspliced RNA from the host cell nucleus to the cytoplasm (41, 90). Like the Tat, the Rev also has an ARM which serves as NLS. This region of the Rev binds to RNA. In addition, Rev also has a region known to function as nuclear export signal (NES) (127). The NES region is a leucine rich region and mutation in the region inhibits movement of the transcripts from the nucleolus and cytoplasm and vice versa (127).

The region of mRNA that is recognized by the Rev is known as the Rev Responsive Element (RRE) (128, 129). The NES region of the Rev transports these transcripts mediating interactions with import proteins, Imp β and export protein, Crm 1 (90). These export and import proteins interact with nucleoporins found in the nuclear pores (90).
3.3.3.3 Accessory Proteins

3.3.3.3.1 Nef (Negative factor)

Nef is a 27 kDa protein composed of 206 amino acids (89, 90). Nef has been shown to down regulate CD4 and major histocompatibility complex 1 (MHC 1) expression on the surface of the infected cell resulting in inhibition of CTL mediated lyses and enhancement of virion infectivity (90). The down regulation of CD4 helps in incorporation of env proteins on the budding viral particle(89). Nef is also known to enhance the infectivity of the HIV-1 during budding by a CD4 independent mechanism (90). Nef defective virions are associated with a slower disease progression (130).

3.3.3.3.2 Vpr (Viral protein R)

This is a 14 kDa protein and is composed of 96 amino acids (89, 90). The Vpr helps in transportation of the uncoated nucleoprotein into the nucleus (89). Vpr helps in transportation into the nucleus even in non-dividing cells like macrophages (131). This is mediated by NLS present in Vpr which transports viral nucleoprotein even in the absence of a mitotic nuclear envelope breakdown(89). The Vpr is also associated with arrest of cell cycle at the G2 phase(132). Vpr thus indirectly enhances viral infectivity since HIV-1 LTR is transcriptionally more active in the G2 phase (90, 133).

3.3.3.3.3 Vif (Virus infectivity factor)

This is a 23 kDa protein and is composed of 192 amino acids (89, 90). The Vif protein helps in replication of HIV-1 in non permissive cells or cells which prevent the replication of HIV-
1. It is possible that Vif neutralizes an inhibitory component present in non-permissive cells (89). Vif may also play a role in viral assembly and/or maturation as it is observed that Vif mutant viruses have an altered core structure (89).

3.3.3.3.4 Vpu (Viral protein U)

The Vpu is composed of 81 amino acids and is unique to HIV-1(90). Vpu increase HIV-1 infectivity by two possible ways. Vpu helps in degradation of CD4 complexed with the gp120 and gp41 subunits in the endoplasmic reticulum (89). Vpu is also associated with promoting and facilitating the budding of virions from the infected cell (90).

3.3.3.3.4 Tev

Another novel hybrid protein designated Tev has been shown in certain strains of HIV-1. It is a hybrid of Tat, Env and Rev and contains 72, 38 and 91 amino acids from Tat, Env and Rev, respectively (90). The biological function of the protein is not known.

3.3 Replication of HIV-1

The replication of HIV-1 consists of a number of steps which can be conveniently divided into two phases (134). The early phase which consists of viral attachment, fusion, reverse transcription and integration into the host chromosome and the late phase which involves transcription, nuclear export, assembly, budding and finally maturation of the viral particle (134). The entire life cycle is completed in approximately 24 hours (41).
3.4.1 Viral attachment

This step primarily involves interaction between the gp120 present on the viral envelope and CD4 molecule present on the surface of the T cells (135). Besides CD4, viral attachment is also mediated by coreceptors CCR5 and CXCR4 (136, 137). Although coreceptors like CCR2b and CCR3 can be also used, it appears that CCR5 and CXCR4 are the most important (138, 139). Both CCR5 and CXCR4 are transmembrane proteins and function as chemokine receptors (136, 137). While CCR5 binds to macrophage inflammatory protein (MIP) 1α, 1β and regulated upon activation normal T cell expressed and secreted (RANTES), CXCR4 binds to stromal cell-derived factor-1 (SDF-1) (136, 137).

3.4.2 Fusion and entry

The binding of gp120 to CD4 results in the exposure of sites in the N-terminal of gp41 (41). The process of membrane fusion is dependant greatly on the structural integrity of the gp41 segment(121). The conformational change that accompany binding of gp120 to CD4 leads to attachment of CD4 to coreceptors. The cascade of events resulting in gp120 attachment to CD4 and coreceptors causes the N-terminal of gp41 to be pushed and subsequently inserted into the host’s cell membrane(140). This helical portion of the N-terminal region known as HR1, subsequently folds to reveal the HR2 region. The HR2 region contains the C-terminal region of gp41 (141). The resulting bundle consisting of three HR1 and HR2 regions brings the two cells involved in fusion close to each other (141, 142). The structural transitions in gp41 and the free energy changes associated with it are believed to result in lipid mixing and membrane fusion(143).

The steps in entry of the HIV-1 are shown in Fig.4.
Fig. 4: Model of HIV-1 viral entry

HIV enters host cells through the interaction of HIV envelope glycoproteins with the host cell CD4 receptor and a coreceptor, either CCR5 or CXCR4. The HIV envelope glycoprotein unit consists of 3 gp120–gp41 heterodimers.

Once bound to both CD4 and the coreceptor, the HIV envelope glycoprotein undergoes a second conformational change that allows insertion of the hydrophobic fusion domain of gp41 into the membrane of the host cell.

Each of the 3 gp120 molecules folds upon itself forming a 6-helix bundle, which brings the viral membrane and host cell membrane in close proximity.

As a result of several such interactions between the HIV envelope and host cell CD4 receptor and CCR5 or CXCR4 coreceptors, the viral membrane and host cell membrane fuse, and contents of the virion enter the host cell.

JAMA, Aug 16 2006, Vol 296 and Pg No 821, Copyright © (2006), American Medical Association. All Rights reserved.
The process of fusion is followed by the uncoating of the virus resulting in the formation known as reverse-transciptase complexes (RTCs) and preintegration complexes (PICs) (134). The PICs differ from RTCs since they are capable of integration as reverse transcription is complete in these complexes (134).

3.4.3 Reverse Transcription

Reverse transcription does not occur in the intact virion and is initiated only after the viral genome has been released from the core (p24) (144). Initialization of reverse transcriptase requires tRNA\textsubscript{Lys} binding to the primer binding site (PBS) located on the downstream of the U5 region in the 5’ end of the LTR located in the HIV-1 genome (90, 145). The RT then acts on this complex and transcribes till the R (strong stop) region located upstream of the U5 region in the 5’ end of the LTR (90). RNase H acts on the region upstream of the R region and the newly synthesized DNA then reanneals to the 3’R(90). DNA synthesis is then initiated again using this as the starting site resulting in the synthesis of a negative strand DNA. RNaseH then destroys the original RNA except the portion at the 3’ which hybridizes to the polypurine tract present in the negative- strand DNA. This portion of the RNA serves as a template for the synthesis of positive-strand DNA (90).
3.4.4 Integration

For integration of HIV-1 DNA with the host chromosome to occur the viral DNA has to be transported into the nucleus. The transportation of the viral DNA into the nucleus is mediated by the accessory protein, Vpr (146). The first step in integration of the viral DNA is known as 3’-end processing and involves exposing the 3’- hydroxyl group that is to be joined to the target DNA (113). The first step involves removal of 2 nucleotides from the 3’ end and this is followed by the DNA strand transfer. This involves integration of the viral DNA into the host DNA (113). The HIV-1 viral DNA incorporated into the host chromosome is known as provirus (41).

3.4.5 Viral protein synthesis

Viral protein synthesis is initiated by binding of the Tat protein to the transactivation response (TAR) RNA (90). The TAR RNA has been shown to have a stem-loop-bulge structure with Tat binding to the bulge portion (90). Transcription requires the concerted interaction of Tat,NF-kB, Sp1 and RNA pol II (41). HIV-1 synthesizes three types of mRNA which are categorized into genomic RNA (Gag and Gag-pol precursor), singly spliced mRNA (env, Vpu, Vif and Vpr) and multiply spliced mRNA (tat, rev and nef). The mRNAs are then transported to the cytoplasm by Rev protein where translation occurs (127). The proteins are transferred to the nucleus and the plasma membrane depending upon the function of the protein (41).
3.4.6 Viral assembly and release

The viral protein synthesis occurs in the cytoplasm though the sites of the synthesis are different. The Env protein is synthesized in the endoplasmic reticulum and then transported to the plasma membrane where they form the gp120 and gp41 components (41). The Gag and Gag-Pol precursor proteins are synthesized in the free cytoplasmic ribosomes and are subsequently transferred to the plasma membrane (41). Viral particles bud from the plasma membrane after the association of the Env proteins with the Gag and Gag-Pol precursor proteins. The formation of these immature particles before budding requires p17 (90). The viral particles which bud from the plasma membrane are non infectious and become infectious after cleavage of the Gag-Pol precursor protein by protease (41, 90). The viruses have been shown to bud through cholesterol rich lipid rafts on the membrane (147).

3.5 Genetic Diversity of HIV-1

Since, the commencement of the AIDS epidemic more than two decades ago, HIV-1 has evolved, differing from one geographical location to another (148). This variability at the genomic level can be attributed to high mismatch error rate of the HIV reverse transcriptase (RT) enzyme coupled with the absence of proofreading capacity, diploid genome, the rapid turnover of the virus in vivo, viral fitness, immune response and recombination events that are taking place during replication (149, 150). It has been reported that HIV-1 RT has an average error rate of 1/1700 per detectable nucleotide incorporation (149). Further, certain regions of the genome are mutational hotspots with an error rate as high as 1 per 70 nucleotides (149). As a result when the nucleotide sequence of HIV-1 is analyzed, a great deal of heterogeneity
is observed. The most notable manifestation of HIV-1 genetic diversity is the phylogenetic clustering of viral isolates geographic region wise referred to as clades or subtypes based on the 20-50 per cent differences in env nucleotide sequences. The inherent diversity of HIV-1 is compounded by three separate induction of the virus from chimpanzees to humans (1, 151). Due to this reasons, there are now three groups of HIV-1: M (Major/Main), N (Non-M, Non-O/New) and O (Outlier). The env proteins of group M and O can show a variation of as much as 30-50 per cent. The N subtype appears to be phylogenetically equidistant from M and O (148). The M group is the most prevalent among of the three groups. It now has nine subtypes (A-D, F-H, J, K), all of which have originated from Central Africa. The amino acid distances in the env gene between the subtypes in the major group have reached 25-35%, while in the gag gene it is about 15% (1, 152). Within subtypes A and F, there are separate subclusters that are related closely to each other than to other subtypes. They are designated A1, A2 and F1, F2, respectively (1). Within subtypes B, C and G there are geographically localized sub-clusters that share a common ancestry as suggested in phylogenetic analysis: subtype B from Thailand (153), subtype C from India (154) and Ethiopia(155), and subtype G from Portugal (156).

Majority of the HIV-1 strains analyzed till date are subtyped. However, there are a few HIV-1 strains with genomes having regions represented from different subtypes (recombinants) seen in geographic areas where more than one type is circulating. This recombinant status is reinforced by the finding that irrespective of the regions of their genome analyzed they fail the criteria of a single designated subtype. These ‘hybrid strains’ are the products of recombinant events taking place in the virus. This is due to the “template switching” ability of the reverse transcriptase enzyme (157). Two types of recombinant forms have been identified – circulating recombinant forms (CRF) and the unique recombinant forms (URF)
If the recombinant is identified in at least three epidemiologically unlinked individuals characterized by full-length genome sequencing, they are designated CRFs. Three near full-length genomic sequences are preferred, but two complete genomes in conjunction with partial sequences of a third strain are sufficient (for CRF, the partial sequences must also confirm the CRFs mosaic structure) (158). There are currently 43 recognized CRFs (159). The formerly designated subtypes E and I are now reclassified as CRFs (159). In addition to CRF, several HIV-1 strains with unique mosaic structures have been reported in epidemiologically linked persons. These forms known as unique recombinant forms (URF) have not show any evidence of epidemic spread and are thought to arise due to secondary recombination of a CRF (1). Currently there are 30 of them (160).

Though the sequence diversity within HIV-1 group O is nearly as great as observed within group M, clades are not clearly differentiated phylogenetically. Hence, subtypes within the HIV-1 O group are not yet defined. Since not many group N strains have been sequenced, no subtype has been determined till date. Because of the high degree of divergence, the homologous recombination between group M and O viruses were not expected. Contrary to this, there are recent reports of the intergroup recombinants reported from Cameroon (158, 161).

All groups of HIV-1 are found in Africa. While group M is prevalent all over the continent, groups N and O are geographically restricted to Central Africa (162-165) (116-119). Subtypes A and D are prevalent in East Africa, subtype A in West Africa, and subtype C in south Africa. In West and Central Africa, the most prevalent genetic form is however a recombinant virus CRF02_AG (1). Subtype B is the most prevalent form in western and central Europe, the Americas and Australia. Subtype C is most prevalent in the Indian subcontinent and
recombinants CRF01_AE and subtype B in South East Asia (1). Globally subtype C is the most predominant subtype causing 50 per cent of infections(166). Even in areas that were traditionally non-C in nature, it is becoming more predominant. Subtype A was estimated to be the second leading cause of the pandemic (12%) followed by subtype B strains (10%) (166). Among the recombinant forms CRF01_AE and CRF02_AG were responsible for 5% of infections each (166).

There are a number of methods, to understand the genetic heterogeneity of HIV-1 subtypes, the reference method being sequencing and phylogenetic analysis. This method not only determines the subtype but also examines the relationship between a set of sequences (1). However, there are other methods that are economically feasible and more accessible in the developing world.

Heteroduplex mobility assay (HMA) makes use of the difference in electrophoretic mobility of a heteroduplex formed between the amplified PCR product of a sample and a reference strain to identify genotypes (167). HMA can employ different target sequences with the gag and the env region (gp120) being the preferred region. The env gp41 based HMA is considered to be a useful tool to monitor subtypes in countries with divergent strains of HIV-1. It circumvents problems arising due to variation in sequences as it targets a relatively conserved region (168).

Hyendrickx et al have used a combination of gag/ env primers to detect recombinant strains of HIV-1 group M isolates (169). Another well described and simple laboratory method for the differentiation of HIV-1 subtypes is the V3 serological sub-typing (170-172). This
method is based on the binding of antibody to peptides from the V3 loop of the envelope from different subtypes.

Genotyping and serotyping have shown a good correlation in areas where a single subtype circulates (173). Serotyping with the V3 peptide depends on an immune response to an antigen coded by a very small antigenic domain and thus a single amino acid substitution can affect serotyping. By contrast, genotypic methods are based on the analysis of a much larger domain and are hence more specific. Though simple, it is however not very practical in areas where multiple subtypes co-circulate. Further, it is not reliable in differentiating subtype C and A (174).

### 3.6 Routes of Infection

The three documented routes of transmission of HIV-1 can be classified into sexual, vertical and percutaneous routes. Irrespective of the route of infection, transmission depends on factors like HIV-1 concentration in the relevant body fluid and susceptibility of the naive host besides virus specific determinants (175, 176).

In the beginning of the HIV-1 pandemic because of the population affected homosexual spread of transmission was highlighted(177). There were subsequent reports of heterosexual mode of transmission (178, 179). In the heterosexual transmission route, the female partner of an infected male has a greater chance of getting infected than the male partner of an infected female (180). The heterosexual route of transmission accounts for over 50% of infections spread in Asia, Sub-Saharan Africa and Latin America (including the Caribbean) (176).
HIV-1 transmission also occurs from an infected mother to a child. The risk of maternal to foetal transmission is 13-40% and can be reduced by treatment of the mother and infant before and after delivery, respectively (181-183). Postpartum transmission occurs by breast feeding and is associated with a transmission rate of 16% (184). HIV-1 transmission has also been associated with transfusion of infected blood and blood products and reuse of contaminated needles (185).

Epidemiologically the HIV-1 pandemic is seen to have two distinct patterns: ‘concentrated’ and ‘generalized’. The ‘generalized’ pattern is seen in sub-saharan African countries where the epidemic is self sustaining in the population. In the ‘concentrated’ pattern which is seen in the rest of the world the epidemic is seen only in specific risk groups like sex workers (176, 186).

3.7 Immunopathogenesis of HIV-1 infection

The median time taken for HIV-1 infection to for progression to AIDS following infection is approximately 8 to 10 years and includes three stages a) primary infection b) clinical latency and c) AIDS-defining illness (187, 188). The time period between infection with HIV and appearance of symptoms like fever, rash, pharyngitis, headache and lymphadenopathy is known as primary infection. The appearance of these symptoms coincides with immune response to HIV and usually occurs 4-6 weeks after infection (188, 189). This period is marked by high viral load and decline in CD4 counts (187). The viral load stabilizes within 6 to 12 months and this is known as the ‘viral set point’ (187). The viral set point marks the
beginning of clinical latency and is characterized by the absence of any particular signs and symptoms (187). This is followed by the terminal stage of HIV infection is characterized by increasing viral load and decline in CD4 counts less than 200 cells/µl (187, 189). The AIDS defining illnesses are mentioned in section 3.8 of review of literature.

Based on studies on disease progression conducted on cohorts of HIV-1 infected individuals they are further grouped into rapid progressors, long-term non progressors and long-term survivors (188). Individuals in whom disease progression is rapid are known as rapid progressors while long-term non progressors do not show disease progression even after 8 to 10 years (188). Long term survivors are individuals in whom both clinical and laboratory parameters remain stable even though they have progressed to AIDS with a time frame similar to typical progressors (188).

Production and release of new HIV-1 particles occur following infection of CD4 T-cells (135, 190). HIV-1 infection is thus characterized by an impaired immune response as a consequence of the gradual decrease in lifespan and number of T-cells (191, 192). The target cells of HIV-1 also include macrophages, monocytes and dendritic cells (188). Macrophages and monocytes are however resistant to the cytopathic effects of HIV-1 infection and can function as reservoirs for HIV in an infected individual. The HIV-1 present in these cells also escapes detection by the immune system (187). The Langerhans cells present on the mucosal surface serve as ‘trojan horses’ helping in presenting HIV-1 to susceptible cells without showing the cytopathic effects of HIV-1 infection(187). The Langerhans cells are disseminated primarily to the lymph nodes where they present HIV to the susceptible CD4 T cells (187). Infection of CD4 of T cells can lead to the production of viral particles and
subsequent lysis of the cell or lead to a state of latency (189, 193). In the latent state the virus escapes detection by the immune system. Activation of the T-cells harbouring latent viruses on exposure to antigens leads to production of viral particles (189). The $t_{1/2}$ of a virus producing T cell is $2.0 \pm 0.9$ days (194).

Besides CD4, viral attachment is also mediated by coreceptors CCR5 and CXCR4 (136, 137). Although coreceptors like CCR2b and CCR3 can be also be used, it appears that CCR5 and CXCR4 are the most important (138, 139, 195). Both CCR5 and CXCR4 are transmembrane proteins and function as chemokine receptors (136, 137). While CCR5 binds to macrophage inflammatory protein (MIP) 1$\alpha$, 1$\beta$ and regulated upon activation normal T cell expressed and secreted (RANTES), CXCR4 binds to stromal cell-derived factor-1 (SDF-1) (136, 137). CCR5 co-receptors are used in the early stages of HIV-1 infection but there is a switch to CXCR4 usage in the later stage (50). This occurs in about 50% of individuals infected with subtype B strains (51). This switch in co-receptor usage is not common among clade C strains (52, 53). Previous data on co-receptor usage in HIV-1 clade C strains from India has shown the absence of this co-receptor switch (53). In this study from India, the coreceptor usage was determined using GHOST cell assay on a total of 40 HIV infected individuals. Subtype C strains from 39 HIV-1 infected individuals were observed to use CCR5 and one HIV-1 infected were observed to use CCR5 and one HIV-2 strain was found to be dual tropic. A recent report from Africa has however shown an emergence of X4 utilizing clade C strains to the extent of 30% in infected individuals (54). HIV-1 strains which use CCR5 coreceptors are known as nonsyncitium inducing (NSI) whereas CXCR4 are known as syncitium inducing strains (SI) (50, 51). Syncitium is characterized by cell clustering followed by fusion. The role of CCR5 in HIV-1 infections can be observed in individuals
who have a 32bp deletion mutation in the CCR5 gene (196, 197). Both in vitro and in vivo studies have observed inhibition of HIV-1 entry into cells with this mutation (198, 199).

3.8 CDC classification of HIV disease

In 1993, the centres for disease control (CDC) and prevention issued a revised classification of HIV disease emphasizing the clinical importance of the CD4+ T-lymphocytes count in the categorization of HIV related clinical condition (200). It stated that antiretroviral treatment be considered in all individuals with a CD4 count < 500 cells/µl. It recommended the institution of prophylaxis for Pneumocystis carinii pneumonia in all those who have a CD4 count < 200 cell/µl. HIV infected individuals are classified based on clinical condition into 3 categories. The clinical conditions and the corresponding clinical stages are given below.

Clinical Category A

Category A consists of one or more of the conditions shown below in individuals’ ≥13 years of age with documented HIV.

- Asymptomatic HIV infection
- Persistent generalized lymphadenopathy
- Acute (primary) HIV infection with accompanying illness or history of acute HIV infection
**Clinical Category B**

This consists of symptomatic conditions in an HIV infected adolescent or adult not included among the conditions that fall in the clinical category C

- Bacillary angiomatosis
- Candidiasis, oropharyngeal (thrush)
- Candidiasis, vulvovaginal, persistent frequent or poorly responsive to therapy
- Cervical dysplasia, moderate or severe cervical carcinoma *in situ*
- Constitutional symptoms, such as fever $\geq 38.5$ °C or diarrhoea lasting $>1$ month
- Oral hairy leukoplakia
- Herpes zoster involving at least two distinct episodes or more than one dermatome
- Idiopathic thrombocytopenic purpura
- Listeriosis
- Pelvic inflammatory disease, particularly if complicated by tubo-ovarian abscess
- Peripheral neuropathy

**Clinical Category C**

This includes all the conditions described as AIDS surveillance case definition. For classification purposes, once a category C condition has occurred, the person will remain in the same category

- Candidiasis of bronchi, trachea or oesophagus
- Cervical cancer, invasive
- Coccidiodomycosis, disseminated, extra-pulmonary
• Cryptococcosis, extrapulmonary
• Cryptosporidiosis, chronic intestine (> 1 month duration)
• Cytomegalovirus disease other than liver, spleen, nodes)
• HIV encephalopathy (dementia)
• Herpes simplex: Chronic ulcer (>1 months’s duration); or bronchitis, pneumonia or oesophagitis
• Histoplasmosis, disseminated or extra pulmonary
• Isosporiasis, chronic intestinal (>1 month duration)
• Kaposi’s sarcoma
• Lymphoma, Burkitt’s, immunoblastic, primary of brain
• Mycobacterium avium complex disseminated or extra pulmonary
• Mycobacterium tuberculosis (pulmonary or extra pulmonary)
• Pneumocystis carinii pneumonia
• Progressive multifocal leukoencephalopathy
• Toxoplasmosis of the brain
• Recurrent salmonella septicaemia
• HIV wasting syndrome – profound involuntary weight loss > 10% of baseline weight plus either chronic diarrhoea (at least two loose stools per day for \( \geq 30 \) days\(^0\) or chronic weight loss and documented fever (for \( \geq 30 \) days intermittent or consistent) in the absence of concurrent illness or condition other than HIV infection that could extend the findings (e.g. cancer, tuberculosis, cryptosporidiosis or other specific enteritis)
• Recurrent pneumonia –(more than one episode in 1 year period) acute (new evidence not present earlier) pneumonia diagnosed by both
i) Culture (or other organism specific diagnostic method) obtained from a clinically related specimen of a pathogen that typically causes pneumonia (either than *Pneumocystis carinii* or *Mycobacterium tuberculosis*) and

ii) Radiological evidence of pneumonia, cases that do not have laboratory confirmation of a causative organism for one of the episodes of pneumonia will be considered to be presumptively diagnosed.

The nine groups that are based on the combination of clinical condition and CD4 counts are given below as per the CDC classification (200)

<table>
<thead>
<tr>
<th>CD4+ T Cell categories</th>
<th>Clinical Categorization</th>
<th>CD4 T cells/µl</th>
<th>CD4%</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥500 cells/µl</td>
<td><strong>A</strong> (Asymptomatic primary HIV or PGL*)</td>
<td><strong>A1</strong></td>
<td>≥29</td>
</tr>
<tr>
<td>200-499 cells/µl</td>
<td><strong>B</strong> (Symptomatic not A or C)</td>
<td><strong>B1</strong></td>
<td>14-28</td>
</tr>
<tr>
<td>&lt;200 cells/µl#</td>
<td><strong>C</strong> (AIDS indicator conditions)</td>
<td><strong>C1</strong></td>
<td>&lt;14</td>
</tr>
</tbody>
</table>

*PGL – persistent generalized lymphadenopathy  # - AIDS indicator T-cells

The equivalent CD4% for each category of CD4 count was also given and is given below

<table>
<thead>
<tr>
<th>Category</th>
<th>CD4 T cells/µl</th>
<th>CD4%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>≥500 cells/µl</td>
<td>≥29</td>
</tr>
<tr>
<td>2</td>
<td>200-499 cells/µl</td>
<td>14-28</td>
</tr>
<tr>
<td>3</td>
<td>&lt;200 cells/µl#</td>
<td>&lt;14</td>
</tr>
</tbody>
</table>
3.9 Laboratory Methods

3.9.1 Tests for diagnosis of HIV-1

Tests for diagnosis of HIV infection individuals were not introduced until 1987 when the United States Public Health Service (UPHS) issued guidelines for counselling and testing of individuals most likely to be infected or those belonging to the high risk group (201). Before these guidelines were introduced, HIV testing was employed from 1985 to protect blood supply. The delay in introducing the tests for patient care was due to a lack of consensus on the implications of a positive result on transmission of HIV(201).

Laboratory tests for diagnosis of HIV infection can be classified into direct and indirect. Direct methods are those used for detection of nucleic acid and antigens. Serological tests form the mainstay of HIV diagnosis. The physiological and sociological implication associated with a positive result makes it necessary to develop tests with the highest specificity and sensitivity.

3.9.1.1 Serodiagnosis of HIV -1

The time between HIV infection and the development of an antibody response known as the ‘serological window period’ is approximately 25 days (189). There are 4 generations test used for detection of HIV-1 infection. The first generation tests make use of whole viral lysate as antigen but had a problem with specificity due to reactivity of antibodies against HLA antigens (189). This problem was circumvented by the use of synthetic peptides or antigens in the second generation kits. Both the first and second generation kits could detect only IgG antibody (189). The third generation kits were able to detect IgM antibodies in
addition to IgG antibodies which resulted in reduction of the window period compared to the 
first and second generation kits (189). The various studies on third generation tests have 
shown their abilities to detect antibodies 3 days after p24 detection, 8.5 days after RNA 
detection and the earliest was estimated to be 8 days after infection (202, 203). The fourth 
generation tests which can detect HIV-1 infected individuals earlier than third generation 
assays can detect p24 antigens of HIV-1 in addition to IgG and IgM antibodies (204-206).

For diagnosis of HIV in emergency scenarios like needle stick injury (where the HIV status 
of the source has to be determined) and in parturient women (to administer ART to prevent 
mother to child transmission) rapid testing devices are used (189). These devices can also be 
used in resource limited settings where they can supplement ELISA(206). The rapid devices 
can be classified as agglutination assays, flow through devices, lateral flow through devices, 
dot immunoassay and capillary flow devices (207). In a study done to compare 14 such 
devices, the sensitivity and specificity ranged from 95.3% to 100% and 73% to 100%, 
respectively (207).

Western blot tests for HIV-1 are confirmatory test used for diagnosis of HIV infection. These 
are called confirmatory tests are these are not done on samples that give indeterminate results 
in screening assays and not on samples that yield negative results (189). Line immunoblot 
assay make use of recombinant proteins and give less indeterminate results than a western 
blot (206).
3.9.1.2 Direct Detection of HIV

3.9.1.2.1 HIV-1 RNA detection

HIV-1 RNA can be detected as early as one week before detection of p24 and if a theoretical detection of 1 RNA/ml is used then it can be detected 11 days earlier (203). Tests for quantitation of HIV-1 RNA have been shown to have a sensitivity of 100% and specificity of 97% and can be detected one week prior to onset of symptoms (202, 208).

3.9.1.2.2 HIV-1 Proviral detection

HIV-1 proviral detection involves detection of HIV proviral DNA from peripheral blood mononuclear cells (PBMC) (206). This can be employed for detection of neonatal HIV-1 infection(189).

3.9.1.2.3 HIV-1 p24 antigen detection

HIV-1 p24 can be detected in the early stages of infection before seroconversion and in the terminal AIDS stages. Following serconversion immune complexes of p24- antibody are created and hence p24 cannot be detected (189). Assays detecting p24 have been shown to have sensitivity of 79% and specificity of 99% (208). However, p24 antigen assays using heat denatured plasma render p24 assays as useful as HIV-1 RNA detection by PCR (209).
3.9.1.2.4 HIV-1 isolation

Though rarely used for diagnosis of HIV infection, HIV-1 can be cultured from PBMC of infected individuals. The overall sensitivity is 95% in patients with CD4 counts below 500 cell/µl (189).

3.9.2 Tests for monitoring HIV-1 infections

Tests for monitoring HIV-1 infections refer to tests that are employed to determine the time needed to initiate treatment and also for monitoring response to treatment.

3.9.2.1 CD4 estimation

The CD4 counts are good prognostic markers of development of AIDS and are also an important marker for treatment initiation (210, 211). Although flow cytometry is currently the standard technique for estimation of T cell subsets there are a number of other techniques that can be used to give comparable results (212, 213).

It has been recommended that CD4 counts should be determined every 3 to 6 months and a significant change between two tests is defined as approximately 30% change of absolute count (211). While on treatment adequate viral load suppression is defined as an increase in CD4 cell count that average 100-150 cells/µl with an accelerated response in the first three months (211).
3.9.2.2 HIV-1 plasma viral load

HIV-1 viral load is used to monitor response to treatment. An infected individual on ART should have undetectable viral load when measured using any of the three FDA approved kits for viral load measurement. This should be achieved 16-24 weeks after initiation of treatment.

The three FDA approved kits with their detectable limits are (211):

1) Amplicor HIV-1 Monitor Test, version 1.5, manufactured by Roche diagnostic. The lowest level of detection being 50 RNA copies/ml. This is based on reverse transcriptase polymerase chain reaction (PCR).

2) NucliSens HIV-1 QT manufactured by bioMerieux. The lowest level of detection being 80 RNA copies/ml. This is based on nucleic acid sequence based amplification (NASBA).

3) Versant HIV-1 RNA 3.0, manufactured by Bayer. The lowest level of detection being 75 RNA copies/ml. This is based on the principle of signal amplification (bDNA).

According to the recommendation of the 2008 guidelines for the use of antiretroviral agents in HIV-1 infected adults and adolescent by the Department of Health and Human Services, USA (DHHS), plasma viral load should be measured immediately before treatment and at 2–8 weeks after treatment initiation or treatment changes because of suboptimal viral suppression. There should be a decrease of at least a 1.0 log10 copies/ml after 2-8 weeks on a regimen (211).
3.9.2.3 HIV-1 drug resistance testing

The DHHS panel recommends the following in the 2008 guidelines for the use of antiretroviral agents in HIV-1 infected adults and adolescent. The recommendations given below are as per reference cited (211).

- HIV drug resistance testing is recommended for persons with HIV infection when they enter into care regardless of whether therapy will be initiated immediately (AIII). If therapy is deferred, repeat testing at the time of antiretroviral therapy initiation should be considered.
- A genotypic assay is generally preferred for antiretroviral-naïve persons.
- HIV drug resistance testing should be performed to assist in selecting active drugs when changing antiretroviral regimens in cases of virologic failure.
- Drug resistance testing should also be performed when managing suboptimal viral load reduction.
- Drug resistance testing in the setting of virologic failure should be performed while the patient is taking his/her antiretroviral drugs, or immediately (i.e., within 4 weeks) after discontinuing therapy.
- Genotypic resistance testing is recommended for all pregnant women prior to initiation of therapy and for those entering pregnancy with detectable HIV RNA levels while on therapy.
- Drug resistance testing is not advised for persons with HIV RNA <1,000 copies/mL, because amplification of the virus is unreliable.

Plasma is the material of choice for drug resistance assays as studies have shown that it helps in early detection of drug resistance mutations in patients on ART (214, 215). Resistance testing can be done by the following techniques
3.9.2.3.1 Phenotypic Assays

These assays measure the ability of the virus to grow in different concentration of drugs. The pol gene which encodes the protease and reverse transcriptase are inserted into lab strains using recombination methods (24, 216). These recombinant strains are grown in different concentrations of the drug and their replication ability is compared to that of a reference strain grown in similar concentration of the same drug. The replication is measured by use of p24 testing or by the use of reporter genes (211). The fold resistance is then calculated for the patient’s strain (24, 211). The fold resistance is the ratio of the IC50 of the test and reference strain. The value of fold resistance levels needed to classify a strain as resistant are not the same for all drugs. Results can take 2-3 weeks and are useful to determine the resistance patterns of strains with unusual mutations and also to determine the drugs to be avoided in heavily treated patients failing a regimen (24, 211).

There are a number of variables which have to be accounted for like inoculums of virus tested, cell lines used for carrying out the assay and the means of assessing viral replication (24). Phentoypic assays costs more than genotypic assays and are hence less commonly employed in clinical practice.

3.9.2.3.2 Genotypic Assays

Genotypic assays do not involve culturing the virus and involves sequencing of the strain directly from the plasma samples to determine the presence of drug resistance conferring mutations. The protocol of these assays can be separated into components requiring machines and software (technical) and those requiring interpretation of the individual carrying out the assay (subjective) (217).
The first few steps of the genotypic assays consist of technical components which are sample preparation and extraction of RNA, preparation for sequencing followed by sequencing. This is followed by the subjective components which include evaluation of the quality of sequences, editing and construction of a consensus sequence using the appropriate software. The last step is the submission of the consensus sequence for analyses which constitutes the technical component (217).

Genotypic assays are less cumbersome than phenotypic assays, require relatively less time, indicate drugs that should be avoided and require knowledge of mutations that confer resistance and cross-resistance to drugs (24, 211).

Besides CD4 estimation, viral load monitoring and drug resistance assays the DHHS in 2007 also recommended HLA-B*5701 screening and coreceptor assays.

### 3.9.2.5 HLA-B*5701 screening

HLA-B*5701 has been observed in individuals showing hypersensitivity to abacavir which is in the initial six weeks of abacavir treatment (218). Abacavir hypersensitivity is the most important cause for discontinuation of abacavir because of a multiorgan failure observed in 5% to 8% of individuals in clinical trials (211). Based on these findings the DHHS panel recommends the following in the 2008 guidelines for the use of antiretroviral agents in HIV-1 infected adults and adolescent with regard to HLA-B*5701 screening (211).

- The Panel recommends screening for HLA-B*5701 before starting patients on an abacavir-containing regimen, to reduce the risk of hypersensitivity reaction.
• HLA-B*5701-positive patients should not be prescribed abacavir.
• The positive status should be recorded as an abacavir allergy in the patient’s medical record.
• When HLA-B*5701 screening is not readily available, it remains reasonable to initiate abacavir with appropriate clinical counselling and monitoring for any signs of hypersensitivity reaction.

3.9.2.6 Coreceptor assays

Maraviroc, a CCR5 antagonist, was the first member of this group to be approved by the US Food and Drug Administration (FDA) in August 2007 (219). The DHHS panel in the 2008 guidelines for the use of antiretroviral agents in HIV-1 infected adults and adolescent made the following two recommendations with regard to coreceptor assays (211).

• Coreceptor tropism assay should be performed whenever the use of a CCR5 inhibitor is being considered.
• Coreceptor tropism testing might also be considered for patients who exhibit virologic failure on a CCR5 inhibitor.

3.10 Treatment and Prevention of HIV-1 infections

3.10.1.1 Treatment of HIV-1 infections

HIV replicates in the cells of the immune system particularly CD4+ T cells of the infected individual leading to a state of progressive immunodeficiency (190, 191). The HIV-1 genome becomes integrated in CD4+ T cells of infected individuals where they as exist in a proviral state (193). These memory CD4+ T cells in which HIV-1 are present persist even in patients
receiving treatment for HIV-1 infection. Studies have shown that even treated individuals who have maintained undetectable levels of viral load for more than 7 years; the complete eradication of this latent HIV-1 reservoir is not possible (220, 221). Following discontinuation of treatment there is a rebound of HIV-1 viral population from these reservoirs (222, 223). Treatment of HIV-1 infected individuals is useful in reducing but not eliminating HIV-1 replication. Once an infected individual is started on treatment it is important that the individual maintains lifelong adherence to the regimen.

A HIV-1 infected individual is put on combination of drugs which target various proteins of HIV-1 which prevent the completion of the HIV-1 life cycle and hence the production of mature viral particles in the infected individual are inhibited (11-14, 16). This combination of drugs is called antiretroviral therapy (ART) has been found to be effective in controlling HIV infections than administration of only one drug (15, 16). The antiretroviral drugs used in the ART regimen can be broadly classified into the following groups:

1) **Nucleoside and nucleotide reverse transcriptase inhibitors (NRTI)**

As their name suggests members of this group inhibit HIV-1 reverse transcriptase. Since the initial demonstration of inhibition of HIV-1 infectivity by 2',3'-dideoxynucleosides this group now consists of seven nucleoside and one nucleotide reverse transcriptase inhibitor (11). (http://www.fda.gov).

The steps involved in the polymerization reaction catalysed by HIV-1 RT can be divided into the following steps (224, 225).

a) Binding of the template and the primer complex to the RT
b) Binding of the nucleotides along with the divalent cations to the template-primer complex to initiate chain extension.

c) Chain extension by formation of a phosphodiester bond between the 3’-OH present at the primer terminus and the α-phosphate present in the incoming nucleotide

Drugs in this group lack the 3’-OH bond thereby prevent the formation of the phosphodiester bond resulting in premature chain termination. Thus NRTIs inhibit RT activity by competitive inhibition. The nucleosides must be triphosphorylated while nucleotides must be diphosphorylated before they can become functional (24). Nucleoside inhibitors that have been approved by FDA include zidovudine (AZT), didanosine (ddI), lamivudine (3TC), stavudine (d4T), abacavir (ABC) and emtricitabine (FTC). The only nucleotide inhibitor used in treatment of HIV infection is Tenofovir.

Besides this, there are other NRTIs which are in various stages of development e.g., Apricitabine which is active against HIV strain resistant to AZT and 3TC (226, 227), alovudine which is a nucleoside analogue related to AZT and d4T, having activity against nucleoside resistant strains of HIV-1 (228).

2) Non-nucleoside reverse transcriptase inhibitors (NNRTI)

This group of drugs also inhibit the activity of reverse transcriptase like the NRTIs (229). The NNRTIs are however capable of directly binding to the hydrophobic pocket which is close to the active site of the enzyme (230, 231). The hydrophobic pocket of the RT to which the NNRTIs bind is located between the β6-β10-β9 and β12-β13-β14 of the p66 subunit (232). This pocket does not exist in the p51 subunit (37). The pocket is about 10Å from the aspartic
acid triad at positions 110, 185 and 186 (232). Comparison of unliganded and NNRTI bound RT have revealed that the palm sub domain of p66 of RT undergoes changes following the binding of NNRTI to the RT (233, 234). The pocket for the binding of NNRTIs is formed by binding of the inhibitor to the RT and is characterized by the movement of the side chains of Tyr at positions 181 and 188 from a ‘down to up’ position (230, 234).

Inhibition of RT by NNRTIs like nevirapine occurs by two possible mechanisms. The drug has been postulated to either prevent the translocation of the template-primer complex following incorporation of a new nucleotide by restricting the movement of the thumb (‘molecular arthritis’) or by altering the orientation of the carboxylate side chains involved in the catalysis by altering the conformation of β6-β10-β9 (37).

The NNRTIs can be grouped into two classes: the first generation which was discovered by random screening of compounds like nevirapine and delavirdine and the second generation which had a more scientific approach in its development employing techniques of molecular modelling. e.g. Efavirenz (235). The first generation NNRTIs are associated with a rapid development of resistance while the second generation are potent inhibitors of RT from drug resistant strains (235). Currently the NNRTIs employed for the treatment of HIV infections are delavirdine, efavirenz, etravirine and nevirapine.

3) Protease inhibitors (PIs)

The protease enzyme is essential for HIV-1 infectivity and mutation in the protease enzyme prevents the virions from infecting MT4 cell lines which are sensitive to infection by wild
type HIV-1 virions (107). Inhibition of the activity of HIV-1 protease enzymes by natural and synthetic compounds results in the development immature viral particle which cannot infect susceptible cells (105, 236). The crystal structure of HIV-1 protease enzyme was reported from many laboratories and this data was used in designing protease inhibitors (40, 237, 238). The protease inhibitors are examples of drugs developed using a structure assisted drug design and which employs techniques such as protein crystallography, NMR and computational biochemistry (44, 239).

Saquinavir (SQV), developed by Hoffmann-La Roche, was the PI to be approved by the FDA and like the subsequent PIs are competitive inhibitors of HIV-1 protease (240). However, low bio availability on oral administration of SQV leads to the development of rapid resistance (241). This has led to the development of newer PIs like atazanavir, tiprnavir and darunavir which are smaller in size and have higher inhibitory activity against the protease from drug resistant variants (242-245). Concentration of PIs can vary among individuals following PI monotherapy as reflected by inhibitory quotient (IQ) of these drugs (IQ is the ratio of the trough drug concentration divided by the inhibitory drug concentration) (246). This has led to the concept of ‘boosted PIs’ wherein the PIs is given along with a low combination of a ritonavir (a P450 enzyme inhibitor) (247). This was first employed in the combination of lopinavir and ritonavir in which ritonavir inhibits lopinavir metabolism (248).

The PIs are not administered in the first line regimen in India because of it higher costs compared to NRTIs and NNRTIs. Currently the PIs employed for the treatment of HIV infections are atazanavir, darunavir, fosamprenavir, indinavir, lopinavir, nelfinavir, saquinavir and tipranavir.
4) Fusion and Entry inhibitors

Enfuvirtide (T20) is the first member of the HIV fusion inhibitors that binds to sites on gp41 and gp120 (249). Maraviroc was the entry inhibitor to be approved by the US Food and Drug Administration (FDA) in August 2007 (219). With the identification of roles played by CD4 receptor and coreceptors like CCR5 and CXCR4 in HIV-1 viral entry numerous attempts have been made to inhibit this step of the viral cycle (135-137, 250). Maraviroc, a CCR5 antagonist, prevents the interaction of CCR5 with gp 120 (251). Maraviroc is known as a CCR5 antagonist as it not only inhibits the binding of HIV-1 to CCR5 but also can inhibit the signals induced by various chemokines that bind to CCR5 (252). Maraviroc exclusively targets CCR5 utilizing strains hence it necessary to determine the tropism of the HIV-1 strains in an infected individual before administration. CCR5 tropic viruses are present early in infection and have shown to decrease with exposure to antiretrovirals (50, 253). Hence, they may have a greater potential in regimens employed in treatment naïve individuals although studies have shown their usefulness among treatment experienced individuals’ also (251).

5) Integrase inhibitors

The FDA in October 2007 approved the use of Raltegravir which is the first member of new novel class of antiretroviral drugs known as integrase inhibitors (254). This class of drugs interferes with a step essential for integration of the viral DNA into the host chromosome (116). Raltegravir has demonstrated antiviral activity in treatment experienced individuals and in treatment naïve individuals as well (255).
6) Maturation inhibitors

Members of this class of drugs are still in phase II clinical trials and the results obtained have been encouraging (256). The only member of this group Bevirimat, inhibits HIV-1 replication by interfering with the cleavage of the viral gag protein by HIV-1 protease(257).

3.10.1.2 Treatment Guidelines

Complete elimination of HIV-1 form an infected individual is not possible as the virus remains latent in cells. For an ART regimen to be successful in effective viral suppression it should contain three drugs which belong to multiple drug classes. Hence, the primary goals of ART initiation are(22, 211):

1) To reduce HIV- related morbidity and prolong survival (clinical goal)
2) Improve quality of life
3) Restore and preserve immunological function (immunological goal)
4) Maximally and durably suppress viral load (virological goal)
5) Prevent vertical HIV transmission

According to guidelines, ART should be initiated in patients with a history of AIDS defining illness or with CD4 counts less than 200 cells/µl (22, 211). According to the recommendation of the NACO (2007) guidelines, 3TC should be included in all first line regimens along with a NRTI (AZT or d4T) and one NNRTI (NVP or EFV).
3.10.2 Prevention of HIV-1 infections

The development of HIV-1 vaccines has thrown up numerous challenges. The vaccine would have to elicit both a mucosal and systemic immune response in addition to both humoral and cellular response to prevent transmission (258). The vaccine would also have to take into account the genetic variability of the virus. One major hindrance is the lack of suitable animal models to study vaccine response (258).

The recent failure of the Merck HIV-1 vaccine has been a setback to efforts to develop HIV-1 vaccines. The Merck vaccine made use of sequences from the gag, env and the nef region(259). The vaccine in addition to failure to elicit a strong immune response enhanced HIV infections in individuals who had previous immunity to the adenovirus vector used in the vaccine(259).

Vaccines developed by Vical and GenVec in collaboration with Vaccine Research Centre (VRC), NIH, USA are now presently in clinical trials for immunogenicity and efficacy. While the vaccine manufactured by Vical contains gag, pol, nef (clade B), env (A, B, C), GenVec’s vaccine contain gag, pol (B), env (A, B, C) (259). Tat based DNA constructs and recombinant protein are also being considered for its potential use in HIV-1 vaccines (260). Although, Tat based vaccines would not be able to prevent an individual acquiring infection it could block viral replication and disease onset (260).
3.11 Drug Resistance Monitoring

One of the steps taken to tackle HIV infection is the increased availability of antiretroviral drugs thereby increasing the coverage of Antiretroviral Therapy (ART). ART is known to reduce morbidity and mortality associated with HIV infection but this benefit is hampered by the emergence of drug resistance. Drug resistance can not only contribute to reduction in the efficacy of ART but if unchecked can also lead to spread of drug resistance strains among the population (21). The frequency of 1 or more major resistance mutation in recently infected individuals increased from 3.8% to 10.2% while on treatment in a 5 year period as reported from the West (21). Numerous polymorphisms and mutations have been reported which are known to contribute to drug resistance in clade B strains (29). Although HIV-1 subtypes other than B are responsible for most new HIV infections worldwide, the understanding of drug resistance of HIV-1 is largely limited to those areas of the world where subtype B viruses predominate and antiretroviral therapy has been accessible. Data from parts of the world with non B subtypes is slowly emerging.

HIV-1 has a high variability at the genomic level due to a high mismatch error rate of the HIV reverse transcriptase (RT) enzyme coupled with the absence of proofreading capacity, diploid genome, the rapid turnover of the virus \textit{in vivo} and recombination events that are taking place during replication (149, 150). It has been reported that HIV-1 RT has an average error rate of 1/1700 per detectable nucleotide incorporation (149). Approximately $10^4$ to $10^5$ point mutations occur every day in a HIV-1 strain in a treatment naive individual (261). Hence, if the regimen does not suppress viral replication completely mutations conferring resistance can develop in these strains.
Drug resistance mutations can be characterized by studying the effect of site directed mutations on susceptibility to drug or by characterizing the mutations observed in strains isolated from individuals showing treatment failure (24). Studying strains isolated from individuals failing a regimen provides additional information such as background polymorphisms, significance of the mutations in vivo and provide a larger spectrum of mutations for analysis (24). Mutations are typically described by a letter indicating the consensus B wild-type amino acid followed by the amino acid residue number and then a letter indicating the mutation (eg, M184V). However, if a mixture of more than one amino acid exists at a position, the amino acids are written after the position, separated by a slash (eg, K103K/N denotes that the sequence has a mixture of the wild-type residue lysine (K) and the mutant residue asparagine (N) at position 103). Mutations that are seen in the HIV-1 reverse transcriptase (RT) and protease region are classified as primary and secondary mutations. Primary mutations are mutations which reduce drug susceptibility directly whereas secondary mutations reduce drug susceptibility or improve replication fitness of the virus in conjunction with primary mutation(24). Though in vitro a single mutation in the protease can confer resistance to PIs, multiple mutations are required in vivo to confer resistance. This requirement for multiple mutations to confer resistance is known as ‘genetic barrier’ to drug resistance (262, 263).

3.11.1 Mutations in the Protease

HIV-1 protease is a homo-dimer with each monomer consisting of 99 amino acids(40). The amino acids Asp-Thr-Gly at positions 25 - 27 from each monomer constitutes the active site at the dimer interface (40, 107, 108). The HIV-1 protease structure can be divided into the following substructures ‘fulcrum’ (amino acid positions 10-23), ‘flap elbow’ (amino acid
positions 35-42), ‘flap’ (amino acid positions 43-58), ‘flap tip’ (amino acid positions 49-52) and ‘cantilever’ (amino acid positions 59-75) based on the ‘cantilever-fulcrum’ model of flap opening (43). The active site of the protease is covered by two β hairpin structures or ‘flaps’ which serves as clamps (42, 43). Mutations can be observed in conserved regions of the protease as well as in the substrate cleft and the flap.

3.11.1.1 Substrate Cleft mutations

Mutations observed in the substrate cleft that are known to confer resistance to PIs are D30N, G48V, I50V, V82A/T/F/S and I84V. D30N is observed only in patients on a nelfinavir containing regimen and causes a 5 to 20 fold reduction in nelfinavir susceptibility. D30N is often followed by the development of N88D and this combination further reduces nelfinavir susceptibility by about 50-fold (264, 265). The G48V mutation occurs in patients receiving saquinavir and causes 10-fold resistance to saquinavir. It is also associated with about threefold resistance to indinavir, ritonavir, and nelfinavir (264). G48V usually occurs with mutations at positions 54 and 82(266). The V82A mutation along with G48V also occurs in isolates from patients receiving prolonged therapy with saquinavir (264). When present alone mutations at codon 82 confer reduced susceptibility in vitro to indinavir, ritonavir, and lopinavir but when present with other PI mutations, V82A/T/F/S contribute phenotypic and clinical resistance to each of the PIs (262-264, 267).
3.11.1.2 Protease Flap Mutations

The flap tips are the site of many drug resistant mutations especially positions 46, 47, 53, and 54 making important contributions to drug resistance. Mutations at position 54 contribute resistance to each of the approved PIs (262, 263, 267).

3.11.1.3 Protease Mutations at other conserved regions

The L90M mutation contributes or directly confer in vitro and in vivo resistance to each of the seven approved PIs (262, 263). Mutations at position 88 (N88D and N88S) commonly occur in patients receiving nelfinavir and occasionally in patients receiving indinavir. Mutation at this position is usually associated with the D30N mutation which induce high-level nelfinavir resistance (268).

3.11.2 Mutations in the reverse transcriptase

The ‘anatomical’ shape of the functional domain of the RT has been likened to that of a ‘right hand’ and is composed of the ‘fingers’ ( amino acid residues 1-85, 118-155), ‘palm’ (86-117, 156-237) and ‘thumb’ (238-318)(37). The ‘palm’ sub domain is the location of the catalytic aspartate residues (D110, D185, D186)(37). In addition, it also has the ‘connection’ (319-426) and the RNase H (37). The positions of the mutations that confer resistance to NRTIs and NNRTIs are different.
3.11.2.1 NRTI resistance mutations

NRTIs are competitive inhibitors of RT activity inducing premature chain termination. There are two possible mechanisms by which mutations can confer resistance to a strain. One possible mechanism is by allowing the RT to discriminate between the NRTI and the natural dNTPs (102, 269). The other mechanism also known as pyrophosphorolysis, nucleotide excision and primer unblocking allows the RT to remove the incorporated NRTI and continue with chain elongation (270, 271).

Thymidine analogue mutations (TAMS) are primer unblocking mutations which occur at various combinations at positions 41, 67, 70, 210, 215 and 219 (272, 273). These mutations occur primarily in patients on a regimen containing AZT or d4T alone or in combination with other NRTIs (274, 275). The Q151M in association with mutations at positions 62, 75, 77 and 116 confers high level resistance to AZT, ddI and d4T and low level resistance to 3TC (24).

The M184V mutation arises in individuals receiving either 3TC monotherapy or on a 3TC containing regimen which fails to suppress viral replication completely (276, 277). The M184V mutation confers resistance by the discrimination mechanism (278). The M184V mutation has a M184I intermediate which arises since the RT has a bias for G to A transition than for A to G transition [ATG (methionine), ATA (isoleucine), GTG (valine)] (279). The M184V mutation is more advantageous to possess for the strain than M184I and hence there is a transition (280). The M184V mutation is also associated with hypersusceptibility to AZT as it interferes with primer unblocking (281).
Another common mutation in the RT is the T69D mutation which confers resistance to ddC (282). However, *in vitro* studies have shown that other substitutions induced at position 69 can cause resistance to AZT, ddI, ddC and d4T (283). Position 69 in the RT is also associated with an insertional mutation (usually serine) which emerge in viruses that have several TAMS confers resistance to tenofovir and other nucleoside analogues (24, 284). K65R mutation which confers resistance by the mechanism of discrimination confers resistance to ddI and d4T (24).

**3.11.2.2 NNRTI resistance mutations**

The NNRTIs bind directly to a hydrophobic pocket which is about 10Å from the aspartic acid triad at positions 110, 185 and 186 (230-232). The hydrophobic pocket of the RT to which the NNRTIs bind is located between the β6-β10-β9 and β12-β13-β14 of the p66 subunit (232). This pocket does not exist in the p51 subunit (37). The mutations that are commonly known to confer resistance to NNRTIs are L100I, K101E, K103N, V106A, V179D, Y181C/I, Y188L, and G190A (24).

The most common NNRTI resistance conferring mutation in clinical isolates is K103N which confers at least 20-fold resistance to each of the NNRTI (24). Analyses of the RT structure of wild type and K103N mutant reveal a hydrogen bond network that is absent in the wild type. The unliganded mutant strain had hydrogen bonds between Asn at position 103 and Tyr at position 188. Hydrogen bonds were also present between water molecules in the hydrophobic pocket interfering with the ability of the inhibitor to bind to the enzyme (285). High level resistance to nevirapine is associated with V106A, Y181C/I, Y188L (24). L100I is
associated with intermediate resistance to efavirenz and delavirdine while V179D causes low level resistance to each of the NNRTIs (24)

3.12 3D Modelling

Information on 3D structure of proteins like HIV-1 protease has helped in designing and development of new drugs for treatment (44, 286). Designing of drugs using involves protein crystallography, nuclear magnetic resonance (NMR) and computational biochemistry (239, 286). The structures of solved proteins are deposited in the Protein Data Bank (PDB).

Crystallization involves the ordered formation of large stable crystals sufficient enough to diffract X-ray to enable their structures to be studied (287). These structures are then studied using nuclear magnetic resonance. The nuclear magnetic resonance makes use of the magnetic properties of certain nuclei to generate NMR spectra. The spectra are generated as the nuclear spins from the ground to the excited state. The frequency of nuclear transition differed in frequency for different nuclei and also showed differences based on the chemical group. This is thus useful for determining protein structure as resonance pattern would be dependent on the chemical group (287).

Homology modelling involves construction of protein models from its amino acid sequence. The known amino acid sequence known as the query sequence is then aligned with a template sequence whose structure is known. This information on the structure can be obtained from
the Protein Data Bank and can be used to generate the structure of the query sequence. Error in homology modelling can arise due to error in query-template sequence alignment and use of a poor template (288, 289). Hence, though structures developed by homology modelling may not be useful for drug design and protein-protein interaction they can be useful for formulating hypotheses which can be tested. Software which can be used for generating of structures using homology modelling are Swiss model which is an automated server (http://swissmodel.expasy.org) and MODELLER which is a downloadable software (http://salilab.org/modeller/).

3.13 Drug resistance databases and algorithms.

For analyses of sequences to determine drug resistance using genotypic assays an expert system is required (24). The expert system consists of a computerized knowledge base and an inference engine (24). The features of an expert system for HIV-1 genotypic interpretation consist of the following components i) data input ii) knowledge base iii) algorithm iv) data input. The algorithm component determines the level of resistance for each drug. The algorithm component can have drug-based rules, mutation-based rules or machine learning. An algorithm based on drug based rules confers resistance to a particular drug if a combination of mutations or mutation exists in the sequence. A mutation based rule derives a resistance pattern by combining interpretation of individual mutations. A machine learning system can be based on pattern matching, neural network, support vector machines, decision trees etc. Prospective studies have shown that availability of drug resistance reports for infected individuals have allowed more effective regimens to be prescribed as compared to a control group whose physicians had no access to drug resistance reports (35, 36).
In Stanford drug resistance data base (SHDB) is based on the mutation based rule. The score for each drug is totalled and based on the score, resistance classified as: susceptible, potential low-level resistance, low-level resistance, intermediate resistance and high level resistance. The scores for each mutation are assigned based on published studies associating mutations with antiretroviral drugs (290). The geno2pheno (G2P) estimates phenotypic drug resistance data based on the submitted sequence (291). G2P uses two machine learning approaches viz., decision trees and support vector machines (SVM) (291). The output thus contains interpretation based on these two systems. This database also proposes a score for the probability of the submitted sequence to originate from a resistant subpopulation rather than a susceptible population (291). The DR_SEQAN system is an offline system and reports the fold increase of the IC50 based on reports from published literature while the genotypic resistance profile is done based on a series of rules (292). The SHDB thus generated only one output while G2P and DS have 3 and 2 outputs, respectively. The susceptibility patterns were obtained for nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs) which included lamivudine (3TC), abacavir (ABC), zidovudine (AZT), stavudine (d4T), didanosine (ddl), tenofovir (TDF), non-nucleoside reverse transcriptase inhibitors (NNRTIs) which included delavirdine (DLV), efavirenz (EFV), nevaripine (NVP) and protease inhibitors (PIs) which included atazanavir (ATV), amprenavir (APV), indinavir (IDV), lopinavir(LPV), nelfanivir (NFV) and saquinavir (SQV). While submitting to G2P the cut off levels for the phenotypic assay was set at the default level. The SHDB algorithm defines five levels of drug resistance: susceptible, potential low level resistance, low level resistance, intermediate resistance and high level resistance. The G2P algorithm defines three levels - susceptible, intermediate and resistant. The DS algorithm defines the levels as susceptible, partial resistance, significant resistance and high level resistance. In addition, it denoted hypersusceptibility as
agonistic; e.g., the mutation M184V, which induces resistance to 3TC, is responsible for hypersusceptibility towards thymidine analogues like AZT and d4T (16).

3.13 HIV-1 in India

The first report of HIV from India was in 1987 from Tamil Nadu. As per the latest NACO surveillance figures India has 2 to 3.1 million HIV (2006) infected individuals (6). This totals to an estimate of a prevalence of 0.36% at the general population (6). India has seen a decline in the number of HIV infections in the country (17, 18). Among the states in India, Andhra Pradesh has the highest prevalence at 21.2%, followed by Maharashtra (20%), Karnataka (11.2%) and Tamil Nadu (10%) (6). Compared to 2004 there has been a gradual increase in the percent distribution of HIV in the >15 and ≥50 years age group in the year 2006. The percent distribution of HIV in the > 15 year age group showed an increase from 3.0 to 3.8% while the ≥50 year age group showed an increase from 5.8% to 7.5%. The 15 to 49 years age group however showed a decrease in percent distribution of the population to 88.7% in 2006 compared to 91.2% in 2004 (6).

The data from three centres in India have shown a higher prevalence of subtype C in comparison to the other subtypes. In north India, 78.4% of the strains were subtype C (293). A subtype study carried out in western India demonstrated 96% of the tested samples to be subtype C with majority of them being C3 (294). A study undertaken among female sex workers in Kolkata showed 95 per cent of strains to be subtype C (295). All these studies demonstrated that over 68% of the strains showed maximum homology to the C3-Indian
reference strain (NIH, HMA panel) (293-295). A study from south India also showed very similar findings to that observed in the north, west and eastern parts of India (296). Ninety five percent of the 83 strains analyzed showed homology with subtype C. Among the subtype C strains that were further characterized, 90.38 per cent had maximum homology with subtype C3. Subtype A was detected in 3.7 per cent of the individuals tested (296). The other subtypes reported in small proportion are subtypes A, B and a report of unique recombinant variant AE (296-298).

In one sentinel study from India, wherein phenotypic analyses was carried out on treatment-naïve strains, primary resistance to protease inhibitors was 2.5% and reverse transcriptase inhibitors was 6.7% (299). Previous reports from ART naïve individuals in India have shown the Pr to be more polymorphic when compared to the RT and a very low frequency of major mutations (30, 31, 300, 301). Sequences of protease and RT genes from individuals failing treatment from India have reported M184V as the most commonly observed mutation in the RT of HIV-1 strains in patients failing ART (33, 34). There is no evidence for any protease mutations in these reports.