Chapter 1

Introduction
Human immunodeficiency virus (HIV) is a lentivirus (a member of the retrovirus family) that causes acquired immunodeficiency syndrome (AIDS) (Weiss 1993; Douek et al., 2009), a condition in humans in which the immune system fails, finally leading to life-threatening opportunistic infections was discovered in 1983 (Barré-Sinoussi, 1983; Broder 1984; Gallo, 1984). Infection of the target cells by HIV-1 is dependent on the surface representation of cluster determinant 4 (CD4), which serves as a specific virus receptor (Klatzmann et al, 1984; Maddon et al, 1986). The high-affinity interactions between CD4 and the surface (SU) envelope glycoprotein of HIV-1 (gp120) are critical to initiate infection and viral entry (Klatzmann et al, 1984; McDougal et al, 1986). The HIV infection leads to progressive loss of CD4-expressing cells, including CD4+ T lymphocytes (Barre-Sinoussi et al, 1983) that are the regulators of immune responses, and the host becomes susceptible to opportunistic infections, ultimately develops AIDS (Gupta 1993; Vento 1993; Fauci, 1988). In addition to peripheral blood lymphocytes, HIV-1 targets CD4+ cells in the thymus, which is the primary site for CD4+ T-cell selection (Bonyhadi et al, 1993; Stanley et al, 1993), and the lymph nodes, which are major HIV reservoirs in asymptomatic infected individuals (Pantaleo et al, 1993). The HIV infection causes cytopathic effects, including syncytium formation and cell death (Lifson et al, 1986). HIV persistently infects CD4+ cells both in vivo and in vitro (Hoxie et al, 1985; Psallidopoulos et al, 1989; Schnittman et al, 1989), which result in immunological dysregulation and loss of immune competence in infected individuals (Fauci, 1993).
HIV has been found in saliva, tears, nervous system tissue and spinal fluid, blood, semen (including pre-seminal fluid, which is the liquid that comes out before ejaculation), vaginal fluid, and breast milk. However, only blood, semen, vaginal secretions, and breast milk generally transmit infection.

The initial infection with HIV occurs when the body fluids from an infected person is transferred to an uninfected one. The first stage of infection/ the primary/ or acute infection, is characterized by rapid viral replication (Mellors 1995) and in the absence of any detectable adaptive immune response, the viral titers reaches the levels >100 million copies of HIV-1 RNA/ml. This initial cycle of viral replication is important for the development of pathogenic processes i.e., the infection of range of tissue reservoirs and the destruction of HIV-1 specific CD4+ T lymphocytes. The high levels of HIV-1 viremia persist for a short phase and decline by several orders of magnitude before reaching a viral setpoint. The host then generates an HIV-1-specific humoral and cellular immune responses to control initial viral replication (Mellors, 1995).

Several factors influence the viral replication during acute infection. These include the fitness of the infecting virus, host genetic factors and host immune responses. The reports of pantaleo (1993) show that there is clonal expansion of CD8+ T cell (HIV-1-specific) responses, which is indirectly proportional to the initial decline of viremia (Koup et al., 1995). These CD8+ T cells eliminate HIV-1-infected cells directly by MHC class I-restricted cytolysis or indirectly by producing cytokines, chemokines or other soluble factors and restrict the generation of new viral progeny (Yang 1999). Host genetic factors play an
important role in both susceptibility and resistance to HIV-1 infection. One of these is the well characterized CCR5delta32, a major coreceptor for the HIV-1 entry (Samson 1996). Homozygotes for this 32 base pair deletion (CCR5delta32) do not express the receptor at cell surface and are not susceptible for infection with R5 tropic viruses (Samson 1996). HLA class I alleles have also been reported to be associated with both, lower viral set points and slower disease progression, including HLA-B27 and -B57 (Brien 2001, Kaslow 1996). Individuals expressing HLA-B57 exhibited a better control of viral replication following an acute infection (Altfeld 2003). Therefore host immune system and genetic factors influence the clinical manifestations of acute HIV-1 infection and the speed of disease progression.

**Treatment**

There is no cure for AIDS at this time. However, a variety of treatments are available that can help keep symptoms at bay and improve the quality of life for those who have already developed symptoms. Antiretroviral therapy suppresses the replication of the HIV virus in the body. A combination of several antiretroviral drugs, called highly active antiretroviral therapy (HAART), has been very effective in reducing the number of HIV particles in the bloodstream. This is measured by the viral load (how much virus is found in the blood). Preventing the virus from replicating can improve T-cell counts and help the immune system recover from the HIV infection. HAART is not a cure for HIV, but it has been very effective for the past 12 years. People on HAART with
suppressed levels of HIV can still transmit the virus to others through sex or by sharing needles. There is good evidence that if the levels of HIV remain suppressed and the CD4 count remains high (above 200 cells/mm$^3$), life can be significantly prolonged and improved. Treatment with HAART has complications. HAART is a collection of different medications, each with its own side effects.

**HIV Structure**

Obtained from Johnston and Fauci, NEJM, 2007

**Fig 1.1**

HIV is different in structure from other retroviruses. It is around 120 nm in diameter (around 60 times smaller than a red blood cell) and roughly spherical.

HIV-1 is composed of two copies of single-stranded RNA enclosed by a conical capsid comprising the viral protein p24, typical of lentiviruses (Figure 1). The RNA component is 9749 nucleotides long (Ratner et al., 1985). This is in turn surrounded by a plasma membrane of host-cell origin. The single-strand RNA is
tightly bound to the nucleocapsid proteins, p6, p7 and enzymes that are indispensable for the development of the virion, such as reverse transcriptase and integrase. The nucleocapsid (p7 and p6) associates with the genomic RNA (one molecule per hexamer) and protects the RNA from digestion by nucleases. A matrix composed of an association of the viral protein p17 surrounds the capsid, ensuring the integrity of the virion particle. Also enclosed within the virion particle are Vif, Vpr, Nef, p7 and viral Protease. The envelope is formed when the capsid buds from the host cell, taking some of the host-cell membrane with it. The envelope includes the glycoproteins gp120 and gp41.

As a result of its role in virus-cell attachment, the structure of the virus envelope spike, consisting of gp120 and gp41, is of particular importance. It is hoped that determining the envelope spike's structure would contribute to scientific understanding of the virus and its replication cycle, and help in the creation of a cure. The first model of its structure was compiled in 2006 using cryo-electron microscopy and suggested that three copies of gp120-gp41 heterodimers are thought to form a trimer as the envelope spike (Zhu et al., 2006). However, published shortly after was evidence for a single-stalk "mushroom" model, with a head consisting of a trimer gp120s and a gp41 stem, which appears as a compact structure with no obvious separation between the three monomers, anchoring it to the envelope (Zanetti et al., 2006). There are various possibilities as to the source of this difference, as it is unlikely that the viruses imaged by the two groups were structurally different (Subramaniam, 2006). More recently, further evidence backing up the heterodimer trimer-based model has been found (Zhu et al., 2008).
HIV has several major genes coding for structural proteins that are found in all retroviruses, and several nonstructural ("accessory") genes that are unique to HIV. The \textit{gag} gene provides the basic physical infrastructure of the virus, and \textit{pol} provides the basic mechanism by which retroviruses reproduce, while the others help HIV to enter the host cell and enhance its reproduction. Though they may be altered by mutation, all of these genes except \textit{tev} exist in all known variants of HIV.

- \textit{gag} (group-specific antigen): codes for the Gag polyprotein, which is processed during maturation to MA (matrix protein, p17); CA (capsid protein, p24); SP1 (spacer peptide 1, p2); NC (nucleocapsid protein, p7); SP2 (spacer peptide 2, p1) and p6.

- \textit{pol}: codes for viral enzymes reverse transcriptase, integrase, and HIV protease.
• \textit{env} (for "envelope"): codes for gp160, the precursor to gp120 and gp41, proteins embedded in the viral envelope which enable the virus to attach to and fuse with target cells.

• Transactivators: \textit{tat, rev, vpr}

• Other regulators: \textit{vif, nef, vpu}

• \textit{tev}: This gene is only present in a few HIV-1 isolates. It is a fusion of parts of the \textit{tat, env, and rev} genes, and codes for a protein with some of the properties of \textit{tat}, but little or none of the properties of \textit{rev}.

\textbf{RNA secondary structure}

Several conserved secondary structure elements have been identified within the HIV RNA genome. These include the trans-activating responsive (TAR) element located within the 5' end of the genome and the HIV Rev response element (RRE) within the \textit{env} gene (Berkhout, 1992; Paillart et al., 2002). RNA secondary structures have been proposed to affect the HIV life cycle by altering the function of HIV protease and reverse transcriptase, although not all elements identified have been assigned a function.

An RNA secondary structure determined by 2' hydroxyl acetylation and primer extension (SHAPE) analysis has shown to contain three stem loops and is located between the HIV protease and reverse transcriptase genes. This \textit{cis} regulatory RNA has been shown to be conserved throughout the HIV family and is thought
to influence the viral life cycle (Wang et al., 2008). The complete structure of an HIV-1 genome, extracted from infectious virions, has been solved to single-nucleotide resolution (Watts et al., 2009).

**HIV Tropism**

HIV Tropism refers to the cell type that the human immunodeficiency virus (HIV) infects and replicates in. HIV tropism of a patient's virus is measured by the Trofile assay. HIV can infect a variety of cells such as CD4+ helper T-cells and macrophages that express the CD4 molecule on their surface. HIV-1 entry to macrophages and T helper cells is mediated not only through interaction of the virion envelope glycoproteins (gp120) with the CD4 molecule on the target cells but also with its chemokine coreceptors.

Macrophage (M-tropic) strains of HIV-1, or non-syncitia-inducing strains (NSI) use the beta-chemokine receptor CCR5 for entry and are thus able to replicate in macrophages and CD4+ T-cells (Barré-Sinoussi et al., 1983). The normal ligands for this receptor, RANTES, macrophage inflammatory protein (MIP)-1-beta and MIP-1-alpha, are able to suppress HIV-1 infection in vitro. This CCR5 coreceptor is used by almost all primary HIV-1 isolates regardless of viral genetic subtype.

T-tropic isolates, or syncitia-inducing (SI) strains replicate in primary CD4+ T-cells as well as in macrophages and use the alpha-chemokine receptor, CXCR4, for entry (Barré-Sinoussi et al., 1983). The alpha-chemokine, SDF-1, a ligand for
CXCR4, suppresses replication of T-tropic HIV-1 isolates. It does this by down regulating the expression of CXCR4 on the surface of these cells.

Viruses that use only the CCR5 receptor are termed R5, those that only use CXCR4 are termed X4, and those that use both, X4R5. However, the use of coreceptor alone does not explain viral tropism, as not all R5 viruses are able to use CCR5 on macrophages for a productive infection (Barré-Sinoussi et al., 1983). HIV can also infect a subtype of dendritic cells (Gallo, 1983), MDC-1, which probably constitute a major reservoir that maintains infection when T helper cell numbers have declined to extremely low levels.

**Life Cycle**

**Entry**

HIV can only replicate (make new copies of itself) inside human cells. The process typically begins when a virus particle bumps into a cell that carries on its surface a special protein called CD4. The envelope spikes gp-120 on the surface of the virus particle interacts with its high affinity receptor CD4 on the target cell via a conserved site on it. Then the V3-loop of gp-120 binds to the core receptor CD26 on the target cell. Proteolytic cleavage of the V3 loop by CD24 induces a conformational change in gp120 & gp41 exposing the fusogenic domain in gp-41. This fusogenic domain then mediated fusion between the viral envelop and target cell plasma membranes. The contents of the HIV particle are then released into the cell, leaving the envelope behind.
**Reverse Transcription and Integration**

Following the nucleocapsid into the host cell cytoplasm, the core proteins are removed, releasing single stranded RNA and RT which copies the single stranded RNA forming RNA-DNA hybrid. After the original RNA template is partially degraded by ribonucleaseH, the synthesis of the second DNA strand proceeds. The double stranded DNA is then translocated to the nucleus and integrated into the host chromosomal DNA by the viral integrase enzyme which is packaged together with the RT in the virion. Once integrated the viral DNA is permanently associated with the host cell DNA. In this form, the virus can remain latent for years together without expressing its genes. Once integrated, the HIV DNA is known as provirus.

**Transcription and Translation**

HIV provirus may lie dormant within a cell for a long time. But when the cell becomes activated, it treats HIV genes in much the same way as human genes. First it converts them into messenger RNA (using human enzymes). Then the messenger RNA is transported outside the nucleus, and is used as a blueprint for producing new HIV proteins and enzymes.

**Assembly, Budding and Maturation**

Among the strands of messenger RNA produced by the cell are complete copies of HIV genetic material. These gather together with newly made HIV proteins and enzymes to form new viral particles, which then begin to assemble within the
host cell, begin to assemble within the host cell, the host cell plasma membrane is modified by insertion of gp-41 & associated gp-120. The viral RNA and core proteins then assemble beneath the modified membrane, acquiring the modified membrane as its envelope during budding. The enzyme protease plays a vital role at this stage of the HIV life cycle by chopping up long strands of protein into smaller pieces, which are used to construct mature viral cores.

The newly matured HIV particles are ready to infect another cell and begin the replication process all over again. In this way the virus quickly spreads through the human body. And once a person is infected, they can pass HIV on to others in their bodily fluids.

**Structure of envelope**

The HIV envelope spike is formed as a complex between gp120 and gp41 (Wyatt and Sodroski, 1998). The gp120 unit mediates attachment of the virus to the target cell, whereas gp41 is required for the fusion of virus and target cell membranes. During HIV infection, the viral envelope spike is first synthesized as a single polypeptide precursor (Wyatt and Sodroski, 1998). In the Golgi complex, the protein subsequently oligomerizes and undergoes extensive glycosylation. The glycosylation process, which is required for proper folding and conformational stability of the envelope glycoprotein (Fenouillet et al., 1994), mainly involves the attachment of N-linked highmannose-type oligosaccharides to the protein backbone. As the glycoprotein is transported through the Golgi, accessible glycan moieties are trimmed and modified by various cellular enzymes (Wyatt and
Sodroski, 1998). These modifications generate complex-type oligosaccharides; glycans that are relatively inaccessible to modifying enzymes remain as high-mannose type glycans (Leonard et al., 1990). The resulting glycoprotein, which has a molecular mass of ~160 kDa, is cleaved in the trans-Golgi network by furin or equivalent endoproteases into gp120 and gp41 (Wyatt and Sodroski, 1998). The gp120-gp41 complexes, which remain associated through weak non-covalent interactions, are initially expressed at the surface of infected cells. During the HIV budding process, the gp120-gp41 complexes are then incorporated into the virus envelope and displayed on its surface as viral spikes (Wyatt and Sodroski, 1998).

**Organization of gp120 on the Viral Surface**

Experimental evidence suggests that the functional unit of the envelope spike is a heterodimeric trimer complex of gp120 and gp41. Electron tomography study revealed structures on the surface of negatively stained virions of SIV and HIV-1 appear to be tri-lobed envelope glycoproteins (Zhu et al., 2003). The HIV core matrix that interacts with gp41 is organized in a trimeric configuration (Hill et al., 1996), and the crystal structures of HIV-1 gp41 cores resemble the transmembrane proteins of other viruses that have been shown to display trimeric envelope spikes (Chan et al., 1997; Tan et al., 1997; Weissenhorn et al., 1997). There are evidences that other envelope species may be present on the surface of HIV-1. Atomic force microscopy analyses have failed to reveal any uniform trimeric envelope species on the surface of virions (Kuznetsov et al., 2003). It was shown that viruses can be captured onto ELISA plate wells using
antibodies that are unable to neutralize viral particles in solution (Poignard et al., 2003; Herrera et al., 2005). Though trimers may likely represent the functional envelope spike, both functional and nonfunctional forms of the envelope may be present on the virion surface. These nonfunctional envelope entities may be monomers, dimers, or tetramers and could possibly arise as the result of (a) the dissociation of functional gp120-gp41 complexes, which could perhaps cause gp120 to be shed from the viral surface, or (b) inefficient trimerization of the spike in the Golgi (Wyatt and Sodroski, 1998; Parren et al., 1997; Burton and Montefiori, 1997).

**Topology of gp120**

Based on comparative sequence analyses, gp120 is divided into five conserved (C1–C5) and five variable (V1–V5) segments (Modrow et al, 1987; Willey et al., 1986). Before obtaining the gp120 core crystal structures, many topological features of monomeric and oligomeric gp120 were deduced from antibody binding experiments and mutagenic studies (Helseth et al., 1991; Kowalski et al., 1987; Moore et al., 1994(a); Pollard et al., 1992; Moore and Sodroski, 1996; Olshevsky et al., 1990; Moore et al., 1994(b)). The C1 and C5 regions were thought to be the main areas on gp120 for contact with gp41, as these regions are accessible to antibody on monomeric gp120 but not on gp120-gp41 complexes (Helseth et al., 1991; Moore et al., 1994(a); Moore et al., 1994(b)). Major segments of the C2, C3, and C4 regions were thought to form a hydrophobic core buried within the gp120 molecule (Moore et al., 1994(a); Pollard et al., 1992). It was proposed that this gp120 core harbors several discontinuous neutralizing
antibody epitopes that overlap the binding sites for CD4 and the coreceptor (Moore et al., 1994(a); Pollard et al., 1992; Moore and Sodroski, 1996; Olshevsky et al., 1990). In contrast to the conserved regions, the variable regions (mainly V1, V2, and V3) were well exposed on the surface of monomeric gp120 (Moore et al., 1994(a)). Deletion of V1/V2 and V3 generally increases the binding affinity of antibodies to epitopes that overlap the binding sites for CD4 and the coreceptor, which suggests that these variable regions may cover conserved epitopes from antibody recognition (Sullivan et al., 1998; Wyatt et al., 1995; Wyatt et al., 1993; Cao et al., 1997). Deletion of the V4 region has been shown to disrupt gp160 folding (Pollard et al., 1992; Wyatt et al., 1993). V4 also seems to tolerate insertion of foreign antibody epitopes (Ren et al., 2005). Determination of the structures of gp120 molecules from HIV and SIV has supported many of the interpretations made from these earlier observations.

**Molecular Structure of gp120**

Four crystal structures of HIV-1 gp120 and one of SIV gp120 have been reported (Chen et al., 2005; Kwong et al., 2000; Kwong et al., 1998; Huang et al., 2005). All five structures are of the gp120 core; i.e., the structures lack the V1/V2 and V3 variable regions, and the N and C termini are truncated. The HIV-1 gp120 structures were determined in complex with the D1D2 fragment of CD4, whereas the SIV gp120 structure was solved unliganded. Because all crystal structures determined so far are of monomeric gp120, they may not adequately represent the structure of oligomeric gp120 on the virus. Despite these and other caveats the structures do provide insight into the conformational flexibility of monomeric
gp120 as well as the locations of receptor-binding sites and putative antibody epitopes on gp120.

**Crystal Structure of HIV-1 gp120 Core in Complex with CD4**

The structure of the gp120 core from the laboratory-adapted virus HXB2 was the first determined in complex with CD4 (Figure 1.4) (Kwong et al., 1998). HXB2 is highly sensitive to antibody neutralization. The second crystal structure was that of CD4 complexed with the gp120 core of the primary virus YU-2 (Kwong et al., 2000(a)), which exhibits a marked resistance to antibody neutralization. The two gp120 core structures are virtually superimposable (Kwong et al., 2000(a)), which is consistent with earlier predictions that the ability of HIV to resist antibody neutralization may manifest mainly in the context of the gp120 quaternary structure on the viral surface rather than in the monomeric gp120 form (Moore et al., 1995). Based on these CD4-bound structures, gp120 is organized into three general areas (Figure 1.5a): (a) the inner domain, (b) the outer domain, and (c) the bridging sheet (Wyatt et al., 1998).

**Inner domain, outer domain, and bridging sheet**

The inner domain is formed mainly by the C1 and C5 regions and is devoid of glycans (Kwong et al., 1998; Wyatt et al., 1998), which implies that these regions function as the major contact interface with the gp41 transmembrane unit. The outer domain, in contrast, is largely covered by glycans (Kwong et al., 1998; Wyatt et al., 1998). Modeling of the gp120 oligomer suggests that these glycans cover large sections of the outer surface of the spike to lower its overall
immunogenicity (Wyatt et al., 1998; Kwong et al., 2000(b)). Comparison of the liganded gp120 structure of HIV-1 (Figure 1.5a) and the unliganded gp120 structure of SIV (Figure 1.5b) shows that the respective outer domains are highly similar (Chen et al., 2005). The

![Diagram of HIV gp120 glycoprotein with five constant (C1–C5) and five variable (V1–V5) regions](image)

**Inner domain** - C1 – C5 regions – no glycans - binds to gp41  
**Outer domain** - V1 – V5 regions – glycosylated – binds to CD4  
**Bridging sheet** – 4 β strands – binds to CCR5

Adapted from Wyatt et al., Nature, 1998  
**Fig 1.3**

conformation of the inner domain in the unliganded structure deviates significantly from the conformation in the liganded structures. This observation suggests that the inner domain may have significant conformational flexibility in the absence of CD4. Comparison of the inner domain substructures in the unliganded and liganded core structures suggests that, upon CD4 binding, these substructures are repositioned somewhat independently of each other, rather than a shift of the inner domain as a single unit (Chen et al., 2005). The large structural
rearrangements associated with the repositioning of inner domain substructures correlates with the large negative entropy and enthalpy changes measured by isothermal titration microcalorimetry (Myszka et al., 2000). It is noteworthy that the majority of gp120 conformational shifts resulting from CD4 binding are in the portion of gp120 that interacts with gp41. Thus, these conformational changes may be necessary to lock the coreceptor-binding site into a fixed conformation and also trigger gp41 into initiating the fusion process (Wyatt and Sodroski, 1998). The conformational changes that occur within the inner domain also affect the formation of the bridging sheet, which links the inner and outer domains. In the CD4-liganded gp120 conformation, the bridging sheet is folded into a compact antiparallel, four-stranded β-sheet (β2-β3 and β20-β21) (Kwong et al., 1998; Wyatt et al., 1998). However, in the unliganded structure, the β-strands that constitute the bridging sheet lie in separate pairs at a distance of approximately 20 Å (Chen et al., 2005); the two β-strands (β2-β3) that constitute the V1/V2 stem are located in the vicinity of the inner domain, whereas the other two strands (β20-β21) are situated near the outer domain in approximately the same location as they are on the liganded structure. Conformational changes that occur within the inner domain upon CD4 binding would result in a 40 Å shift of the V1/V2 stem to form the bridging sheet (Chen et al., 2005). However, molecular modeling of the liganded and unliganded gp120 structures suggests that the unliganded structure may be one of many conformations that gp120 may adopt in the absence of CD4 (Pan et al., 2005). In fact, the models suggest that the β2-β3 strands may oscillate from the conformation observed in the unliganded structure
to a conformation resembling the CD4-bound structure via a series of intermediate conformers (Pan et al., 2005).

Adapted from Kwong et al., Nature 393:648–59

**Fig 1.4:** Crystal structure of HIV-1 gp120 complexed to CD4 and an antibody antigen-binding fragment (Fab). The gp120 core of HXB2 (Protein Data Bank ID 1G9M) (gray), CD4 (orange), the antibody heavy chain (H) (blue) and light chain (L) (green) are shown.

Epitope-mapping studies of antibodies to the CD4-binding site (CD4bs) also suggest that gp120 can adopt a conformation resembling the CD4-bound form relative to the conformation of the unliganded structure.
The CD4-binding site

The binding site for CD4 on the liganded gp120 structure is formed by the interface between the inner domain, bridging sheet, and outer domain (Kwong et al., 1998; Wyatt et al., 1998). At the center of this interface lies a hydrophobic cavity that has been dubbed the Phe43 cavity (Figure 1.5a) (Kwong et al., 1998). However, most of the CD4 contact residues are located on the outer domain of the liganded HIV-1 gp120 structures and form a contiguous binding region (Figure 1.5c). On the unliganded SIV gp120 structure no such region is discernible (Figure 1.5d) (Chen et al., 2005), assuming that equivalent residues in SIV and HIV-1 contact CD4. On the unliganded SIV gp120 core, many of the residues that are presumed to contact CD4 upon complexation with gp120 are located near or within a long cavity that is formed primarily by portions of the inner and outer domains and the β20-β21 segment of the bridging sheet. The location of these conserved residues minimizes their immediate recognition by antibodies, while preserving the ability to contact CD4. In this regard, the curved structure of the D1D2 fragment of CD4 is particularly noteworthy (Figure 1.4); it permits CD4 to curl over the outer domain, so residues located near or within the cavity formed by the inner and outer domains can be reached. Given that the CD4bs is not coherently present on the unliganded structure, it indeed seems likely that gp120 transiently forms conformations that are reflective of the liganded structure; upon interaction with CD4, the gp120 structure is locked in the bound conformation.
Fig 1.5: Comparison of the crystal structures of HIV-1 and SIV gp120 core. (a) Structure of the CD4-liganded HIV-1 gp120 core (HXB2), viewed from the perspective of CD4. The gp120 inner domain (blue), outer domain (yellow), and bridging sheet (orange) are shown. The locations of various gp120 regions are also denoted. (b) Structure of the unliganded SIV gp120 core (Protein Data Bank ID 2BF1), viewed from the perspective of CD4 as in (a). (c) HIV core gp120 in same orientation as (a), depicting CD4 contact residues (orange) and residues that influence coreceptor binding (green). (d) SIV gp120 core in the same orientation as in (b), colored according to the scheme for HIV gp120 in (c).
The coreceptor-binding site

The region that is important for the interaction with the β-chemokine receptor CCR5 has been mapped to residues in the bridging sheet and near the V3 stem (Rizzuto et al., 1998; Rizzuto and sodroski, 2000). These residues lie close together on the liganded HIV-1 gp120 structure, but the equivalent residues on the unliganded SIV gp120 structure are separated into two areas (Figure 1.5c,d) (Chen et al., 2005). These differences are consistent with the notion that CD4 binding is required to lock these areas into a contiguous binding site. The fact that the coreceptor site is not presented until after CD4 binding suggests that the site may be susceptible to antibody recognition. Several studies have shown that HIV strains that do not require CD4 for entry are highly sensitive to antibody neutralization (Hoffman et al., 1999; Edwards et al., 2001; Kolchinsky et al., 2001; Decker et al., 2005). Due to neutralizing antibody–driven selection pressure in vivo, the prevalence of such viruses during infection is likely low. However, in the absence of circulating neutralizing antibodies, e.g., in the central nervous system, such viruses may occur more frequently (Decker et al., 2005; Martin et al., 2001). We note here that the unliganded SIV gp120 structure is derived from strain SIVmac32H. This simian virus is able to infect CD4-negative target cells in vitro with medium efficiency (Reeves et al., 1999). The gp120 from this strain may thus harbor certain structural features that are not observed in the gp120 of HIV-1 or SIV strains that require CD4 for entry.
Antibody epitopes on HIV-1 gp120

The b12 epitope: This is the most broadly cross neutralizing epitope present on the CD4 binding site. The possible aminoacids that interfere with gp120-CD4 binding are Asp 457, Pro 470.

The 2G12 epitope: This present on the silent face of gp120. The aminoacids that aid in the neutralization are Asn 295, Thr 297, Ser 334, Asn 386, Asn 392, Asn 397. The mechanism of virus neutralization is unknown.

The 447–52D epitope: This is another neutralization restricted epitope on the V3 loop region which is often masked by V1 and V2 regions in various viral clades. The aminoacids responsible for the activity are Glu 370, Tyr 384.

There are other CD4 – induced epitopes against which two antibodies 17b and 48d are produced that will interfere with coreceptor binding. The aminoacids involved here are Asn 88, Lys 117, Lys 121, Lys 207, Ser 256, Thr 257, Asn 262, DV3, Glu 370, Glu 381, Phe 382, Arg 419, Ile 420, Lys 421, Gln 422, Ile 423, Trp 427, Tyr 435, Pro 438, Met 475.

Mechanism of fusion

Infection of target cells by HIV is a complex, multi-stage process involving attachment to host cells and CD4 binding, coreceptor binding, and membrane fusion. The initial interaction between HIV and a target cell may be facilitated by nonspecific interactions between positively charged domains on the gp120 protein and negatively charged proteoglycans on the cellular membrane (Mondor et al., 1998; Moulard et al., 2000) or by specific interactions with cell surface lectin
binding proteins such as DCSIGN. Such attachment factors can enhance the efficiency of virus infection (Geijtenbeek et al., 2000). HIV interaction with CD4 occurs at a structurally conserved, recessed surface on gp120 that is formed by epitopes that are discontinuous in the primary protein sequence (Kwong et al., 1998). Upon binding to CD4, gp120 undergoes a conformational shift that has several important consequences. First, two sets of β-sheets that are spatially separated in unbound gp120 are brought together by CD4 binding into a four-stranded β-sheet minidomain called the bridging sheet (Chen et al., 2005; Kwong et al., 1998). Second, CD4 binding results in movement and exposure of the V1/V2 and V3 loop structures. Third, binding of CD4 changes the orientation of gp120 such that the bridging sheet and the V3 loop are directed towards the host cell membrane, where they can subsequently interact with coreceptor (Huang et al., 2005; Trkola et al., 1996; Wu et al., 1996). Thus, CD4 binding is a prerequisite to the formation and exposure of the coreceptor binding site of gp120. In humans, the primary coreceptors for HIV-1 are the chemokine receptors CCR5 and CXCR4, members of the seven-transmembrane G protein-coupled receptor family (Alkhatib et al., 1996; Choe et al., 1996; Deng et al., 1996; Doranz et al., 1996; Dragic et al., 1996; Feng et al., 1996; Oberlin et al., 1996; Zhang et al., 1998). These proteins are integral membrane proteins with seven transmembrane helices, an extracellular N-terminus and three extracellular loops (ECLs) that form a small pocket. The N-terminus of chemokine receptors contains sulfated tyrosine residues and elements within and around ECL2 that are critical for gp120 binding (Atchison et al., 1996; Edinger et al., 1997; Rucker et al., 1996). These spatially
separated domains of CCR5 interact with distinct regions of gp120: the N-terminus with the bridging sheet and the base of the V3 loop, and ECL2 with the tip of the V3 loop (Basmaciogullari et al., 2002; Cormier and Dragic, 2002; Cormier et al., 2001; Hoffman et al., 1999; Huang et al., 2007). The relative dependence on the N-terminus compared with the ECL2 loop appears to vary for different R5 viruses. The N-terminal interaction has now been investigated using a crystal structure between gp120 and an unusual sulfated antibody, 412d, that mimics the N-terminus of CCR5 (Huang et al., 2007). These studies suggest that prior to coreceptor binding, the V3 loop of gp120 is flexible and located close to the target cell membrane. Engagement of the N-terminus by gp120 requires the formation of a conserved sulfotyrosine binding pocket and converts the V3 stem from a flexible structure into a rigid β-hairpin. In contrast, the interaction between the tip of the V3 loop and ECL2 is less well defined, but contact between these regions is particularly important for HIV entry (Lee et al., 1999; Platt et al., 2001; Samson et al., 1997; Wu et al., 1997). These data are consistent with a crystal structure of gp120 in which the V3 loop is found to extend nearly 30Å from its base towards the cellular membrane, where it could presumably make contact with the Extracellular loops (ECLs) of the coreceptor (Huang et al., 2005). Binding of gp120 to CXCR4 appears to occur in a similar fashion (Basmaciogullari et al., 2002; Chabot et al., 1999; Doranz et al., 1999; Lin et al., 2003a), although the V3 loop of viruses that utilize CXCR4 tend to be more positively charged, particularly at positions 11, 24, and 25 of V3 (De Jong et al., 1992; Fouchier et al., 1992; Milich et al., 1993; Xiao et al., 1998). Binding of
gp120 to coreceptor triggers further conformational changes in the envelope trimer that result in the exposure of the hydrophobic fusion peptide of gp41 and its insertion into the host cell plasma membrane. After the insertion of the fusion peptide, the heptad repeat regions HR1 and HR2 of gp41 undergo a highly energetically favorable rearrangement in which they fold back on each other. In a functional trimer spike, this forms a six-helix bundle structure where the three HR1 domains form a central coiled-coil and the three HR2 domains wrap around in an anti-parallel direction around the central coil (Chan et al., 1997; Weissenhorn et al., 1997). This structural rearrangement brings the transmembrane region of gp41, which is embedded in the viral membrane, into close proximity to the fusion peptide, which is inserted into the host cell membrane. This juxtaposition results in the formation of the fusion pore, allowing the viral capsid to enter the cell. The entry process of HIV has traditionally been thought to occur at the plasmamembrane of the cell, but recent evidence suggests that endocytosis of viral particles may be required for complete fusion (Miyauchi et al., 2009).

**Entry inhibitors**

Antiretroviral (ARV) drugs are broadly classified by the phase of the retrovirus life-cycle that the drug inhibits. They include nucleoside and nucleotide reverse transcriptase inhibitors (NRTI), Non-nucleoside reverse transcriptase inhibitors (NNRTI), protease inhibitors, integrase inhibitors, entry inhibitors and maturation
inhibitors. In turn entry inhibitors are classified into 3 types depending upon their mode of action namely,

a. Inhibitors blocking the gp120–CD4 interaction

b. Inhibitors blocking the gp120–coreceptor interaction

c. Inhibitors blocking gp41-mediated membrane fusion

Adapted from Moore and Doms, PNAS, 2003

**Fig 1.6**

*a. Inhibitors blocking the gp120–CD4 interaction*

In contrast to cell associated CD4, which is required for HIV entry, soluble CD4 (sCD4) was found to inhibit HIV entry at high doses *in vitro*. But clinical administration of this protein did not reduce viral loads in HIV-infected patients, and detailed analysis revealed that the levels of sCD4 achieved in patients were not sufficient to inhibit primary HIV isolates (Daar et al., 1990). Still, class of sCD4 derivatives and CD4 mimics, including the PRO-542 CD-IgG2 tetrameric fusion protein and the NBD-556 and NBD-557 compounds (Allaway et al., 1995;
Arthos et al., 2002; Martin et al., 2003; Schon et al., 2006; Trkola et al., 1995) (Fig. 3) were developed. These compounds appear to work by inducing a short-lived activated state of gp120 that spontaneously and irreversibly converts into a nonfunctional conformation. In contrast, the activated intermediate of gp120 generated by cell-surface CD4 is far more stable (Haim et al., 2009). Other small-molecule inhibitors of the CD4-gp120 binding interaction are the BMS-378806 and BMS-488043 compounds. These agents also target the conserved CD4-binding site of gp120, but their precise mechanism of action is unclear. Some studies suggest that these compounds compete with sCD4 for binding to gp120 (Hoet et al., 2006; Lin et al., 2003b), while others indicate that they do not block sCD4 binding (Schon et al., 2006) and may exert their antiviral effects by preventing conformational changes in gp120 upon CD4 engagement (Si et al., 2004). The clinical utility of BMS-806 is limited by a low genetic barrier to resistance, as 1–2 amino acid changes of gp120 result in 40–500-fold resistance to drug (Lin et al., 2003b). Amino acids that confer resistance to BMS-806, including Trp 112, Thr 257, Ser 375, Phe 382, and Met 426, line the ‘phenylalanine 43 cavity’ on gp120 that is involved in stabilization of the CD4-bound conformation of gp120 (Madani et al., 2004). Although BMS-806 was discontinued in phase II clinical development, it has shown success in an animal model as a potential topical microbicide (Veazey et al., 2005). The humanized antibody ibalizumab (TNX-355) binds to the D2 domain of CD4 and blocks CD4-induced conformational changes in gp120 (Moore et al., 1992). This agent has been demonstrated to reduce viral loads and increase CD4 T cell counts in
combination with optimized background therapy (Kuritzkes et al., 2004), but is not orally bioavailable.

**b. Inhibitors blocking the gp120–coreceptor interaction**

The discovery of CCR5 and CXCR4 as the critical coreceptors for HIV entry (Alkhatib et al., 1996; Choe et al., 1996; Deng et al., 1996; Doranz et al., 1996; Dragic et al., 1996; Feng et al., 1996; Oberlin et al., 1996; Zhang et al., 1998) was rapidly followed by the identification of a subset of individuals that were homozygous for an inactivating deletion for CCR5,Δ32-ccr5, which conferred high-level resistance to HIV-1 infection without significant immunological consequences (Dean et al., 1996; Liu et al., 1996; Samson et al., 1996). Coupled with the observation that patients heterozygous for Δ32-ccr5 had delayed rates of disease progression (Dean et al., 1996; Huang et al., 1996; Michael et al., 1997; Rappaport et al., 1997; Samson et al., 1996), these findings indicated that pharmacological blockade of the gp120–CCR5 interaction could be an effective strategy for inhibiting HIV infection. A number of naturally occurring ligands for the CCR5 receptor block HIV infection, including CCL3 (MIP-1α), CCL4 (MIP-1β), and CCL5 (RANTES) (Cocchi et al., 1995). These chemokines exert their antiviral effects by blocking Env binding to CCR5 and by inducing the internalization of CCR5 from the cell surface (Alkhatib et al., 1997), but have potentially undesirable agonist activity on CCR5. Several RANTES derivatives, including AOP-RANTES, NNY-RANTES, and PSC-RANTES, have been developed in an effort to maintain anti-HIV activity while reducing or
eliminating the agonistic effects on CCR5 (Mosier et al., 1999; Simmons et al., 1997). PSC-RANTES is in development as a potential microbicide for HIV (Cerini et al., 2008; Lederman et al., 2004). These agents compete with gp120 for binding to coreceptor and therefore are competitive antagonists of HIV infection. Another strategy for inhibiting gp120–coreceptor interactions is binding of small compounds to a hydrophobic pocket in the transmembrane helices of CCR5 and are believed to exert their antiviral effects by altering the conformation of the extracellular loops that HIV interacts with during coreceptor binding. These agents do not bind to the same binding site as gp120, making them allosteric rather than competitive inhibitors. Many small-molecule antagonists have demonstrated efficacy against HIV replication in vitro, and three of these agents have been tested extensively in humans. Aplaviroc (GW873140) was tested in phase IIb studies before reports of idiosyncratic hepatotoxicity halted its development in 2005 (Nichols et al., 2008). Vicriviroc (SCH-D, SCH-417690) is currently in phase III clinical trials, and is a second-generation compound based on ancriviroc (SCH-C), which was discontinued after being associated with an elongated QT cardiac interval in clinical trials (Strizki et al., 2005; Tagat et al., 2004). Maraviroc (UK-427857) was approved in 2007 by the FDA for the treatment of HIV-infected patients with viral replication and HIV strains resistant to multiple antiretroviral agents. Maraviroc is the result of medicinal chemistry optimization of the compound UK-107543 and is effective against CCR5-using (R5-tropic) HIV strains in the low nanomolar range (Dorr et al., 2005). It is administered orally twice daily, with dosages that vary depending on the presence
of strong CYP3A inducers or inhibitors in the antiviral regimen. Antibodies that block the CCR5 receptor and prevent HIV infection have also been developed. PRO-140 is a humanized mouse anti-CCR5 antibody that prevents gp120 from engaging CCR5 but does not block CCR5 ligand activity. It has demonstrated potent efficacy against CCR5-tropic HIV strains both in vitro and in HIV infected adults (Jacobson et al., 2008; Trkola et al., 2001). It is currently in phase II clinical trials. Unlike CCR5, CXCR4 is essential for multiple physiological processes. In mice, knockout of the CXCR4 gene or its ligand, CXCL12 (SDF-1), is embryonic lethal due to defects in vascularization, hematopoiesis, cardiogenesis, and abnormal cerebellar development (Tachibana et al., 1998; Zou et al., 1998). In humans, heterozygous truncating mutations in the cytoplasmic tail of CXCR4 have been associated with WHIM syndrome, an immunodeficiency syndrome characterized by warts, hypogammaglobulinemia, infection, and myelokathexis (Hernandez et al., 2003). Several polypeptide mimics of the natural ligand for CXCR4, CXCL-12, have been developed. These compounds, including T-22, T-134, T-140, and ALX40-4C, act by binding specifically to the CXCR4 receptor and preventing gp120 binding (Arakaki et al., 1999; Doranz et al., 2001; Tamamura et al., 1994, 1998). ALX40-4C was tested in humans prior to the identification of CXCR4 as a coreceptor for HIV, and despite being well-tolerated did not have a significant effect on reduction of HIV viral loads. However, the majority of patients in this study were later found to have CCR5-tropic strains of HIV (Doranz et al., 2001). Small-molecule antagonists and partial agonists of CXCR4 are under development. The bicyclam analog AMD3100
demonstrates potent activity against CXCR4-using (X4-tropic) strains of HIV in vitro, but possesses cardiac abnormalities and no significant viral load reduction was observed (Hendrix et al., 2004). Interestingly, patients treated with AMD3100 were found to have increased mobilization of CD34+ stem cells into the peripheral circulation (Liles et al., 2005), and the drug was approved by the FDA as a hematopoietic stem cell mobilizer for transplantation under the trade names Plerixafor and Mozobil. The compound AMD070 is a third-generation small molecule CXCR4 antagonist that is orally bioavailable and inhibits X4-tropic strains of HIV with similar potency to AMD3100 (Stone et al., 2007), but development has been halted due to liver histological changes in preclinical toxicity studies. Although no CXCR4 antagonists are in active clinical trials for the treatment of HIV-1 infection, several are in development and have demonstrated potent inhibitory effects on X4-tropic strains (Iwasaki et al., 2009; Murakami et al., 2009).

**c. Inhibitors blocking gp41-mediated membrane fusion**

Pharmacological agents that disrupt gp41-mediated membrane fusion, collectively called fusion inhibitors, were the first entry inhibitors to be approved for the treatment of HIV infection. Synthetic peptides corresponding to the HR1 and HR2 domains of gp41 were found to have potent antiviral effects (Jiang et al., 1993; Wild et al., 1992). These agents were initially analyzed during epitope-mapping experiments designed to identify targets for vaccine development (Matthews et al., 2004). However, biochemical and crystallization studies subsequently
revealed their true mechanism of action: prevention of the formation of the six-helix bundle by competing for binding to the HR1 and HR2 domains on gp41 (Chan et al., 1997; Weissenhorn et al., 1997; Wild et al., 1994). The fusion inhibitor enfuvirtide (T-20) was approved by the FDA for treatment-experienced, HIV-infected patients in 2003. Enfuvirtide is a linear, 36 amino acid synthetic peptide with a sequence identical to part of the HR2 region of gp41 (Wild et al., 1993), and competes for binding to HR1. It has demonstrated potency against HIV in clinical trials (Kilby et al., 2002; Lalezari et al., 2003), although there is considerable variability in the enfuvirtide sensitivity of primary virus strains (Derdeyn et al., 2000; Melby et al., 2006; Reeves et al., 2002). A number of other next-generation peptidic fusion inhibitors are under investigation, several of which have improved pharmacodynamics and efficacy compared with enfuvirtide (Dwyer et al., 2007; Lalezari et al., 2005b). Additionally, certain agents are active against some enfuvirtide-resistant strains of HIV (Dwyer et al., 2007; He et al., 2008), and fusion inhibitors that bind to different functional domains of gp41 can have synergistic effects (Pan et al., 2009). However, since peptidic fusion inhibitors are not orally bioavailable and must be administered via injection, the development of small-molecule inhibitors of gp41-mediated fusion remains a goal in drug development. An alternative approach that has shown considerable potential is the generation of d-peptide pocket-specific inhibitors of entry (PIEs) that bind to gp41 and which, unlike natural l-peptides, are not digested by proteases and have the potential for oral bioavailability (Welch et al., 2007).
Vertical transmission of HIV

The present work has its origins in the protection offered by the maternal placental environment during pregnancy. The vertical (mother-to-infant) transmission of HIV-1 is the major cause of AIDS in children and accounts for more than 10-39% (Domachowske, 1996; Zachar et al., 1999, Ahmad, 2005). The viral load, which is related to clinical and immunological status in mother, is the main contributing factor for HIV-1 vertical transmission (Bongertz, 2001) along with numerous maternal parameters like low CD4+ve lymphocyte counts, disease progression rate and immune competence during various stages of pregnancy (Ahmad, 2005; Elizabeth et al., 2003). Low frequency of HIV transmission is due to the presence of various innate immune factors expressed in women during pregnancy. Also, the viral load in seropositive women reported to decrease during pregnancy (Cameron, 2009). Placenta is not only the major nutrient supplier but also a barrier to various pathogens that attack the developing foetus (Burton and Watson, 1997). So protective environment around this is highly important in control of viremia by the maternal immune system. Such innate immune factors include 55KDa placental factor (Tiensiwakul et al., 2004), LIF (Patterson et al., 2001), IFN γ (Druckmann et al., 2005), Killer specific secretory protein (Hayano, 2002), Indoleamine-2, 3-deoxygenase (Sedlmyr et al., 2002), collectins (Ohtani et al., 2001) and hormonal factors (Hunt et al., 2000) to improve protection of the foetus from harmful pathogens.
Epap-1 and its genesis

Earlier studies identified and characterized a 90 KDa Epap-1, which is isolated from the first trimester placental tissue (Kondapi et. al., 2002; Rani et. al., 2006). Epap-1 was also found to be present in blood and urine of pregnant women during first trimester. Epap-1 inhibits HIV-1 entry and is found to have affinity to HIV-1 gp120. Molecular analysis of Epap-1 and gp120 interaction showed that it possesses recognition sites that can interact with V3 and C5 regions of gp120 (Rani et. al., 2006). Further, Epap-1 expressed in bacterial as well as baculoviral system was shown to be biologically active as the native form. Anti-HIV-1 activities of native and recombinant forms are conserved and specific to V3 and C5 regions. In addition, Epap-1 inhibits the replication of both X4 and R5 viral strains.

Rationale

Epap-1 was shown to block the recognition of V3 and C5 regions of gp120, identification of the region of Epap-1 interacting with V3 loop region would help in understanding the molecular action of Epap-1 in blocking gp120 fusion. Some proteins like Cyanovirin (a protein of 30KDa) suffered in preclinical and clinical studies. So the larger molecular weight of Epap-1(90 KDa) may have various limitations in stability, immunogenicity and production. Thus a peptide may be more stable and can evade unwanted immune reactions if any. Hence the identification of gp120 targeted HIV-1 neutralizing peptides derived from Epap-1 will form new lead in the development of entry inhibitors.
Objectives

- Analysis of anti-HIV potential of proteolysed extract of Epap-1
- Modeling of Epap-1 and gp120 interactions to identify potential gp120 binding regions in Epap-1
- Analysis of cytotoxicity and anti-HIV activity of Epap-1 derived peptides
- Characterization of molecular action of potent peptides
- Design of small peptides with enhanced affinity and activity