Chapter-1

Introduction
Chapter-1

Introduction

1.1 History of HIV and AIDS

1.2 Classification of Retroviruses

1.3 Structure of HIV

1.4 Life cycle of HIV

1.5 AIDS

1.6 Clinical Staging

1.7 AIDS as Zoonosis

1.8 Isolation of HIV

1.9 Epidemiology of HIV

1.10 Why to study Env?

1.10.1 HIV-1 Env binding to CD4 and coreceptor

1.10.2 HIV-1 Env and Tropism

1.10.3 HIV-1 Env and Neutralization

1.10.4 HIV-1 Env and Vaccine development
1.1 History of HIV and AIDS

In 1981, Acquired Immunodeficiency Syndrome (AIDS) was first described, as a distinct clinical disease among young men having sex with men (MSM), who presented either with Pneumocystis carinii pneumonia (now known as named Pneumocystis jiroveci * Frenkel 1999, in honor of the Czech parasitologist) and/or a rare type of cancer of skin, Kaposi’s sarcoma, seen only in immunocompromised persons. They showed unique pattern of underlying immunosuppression without any identifiable cause (Gottlieb, Schroff et al. 1981). Similar cases were subsequently reported in hemophilia patients who had received blood products (Davis, Horsburgh et al. 1983) and intravenous drug users (IDUs) who shared needles and syringes (Hardy, Allen et al. 1985). It was also postulated that a variant of Human T lymphotropic retrovirus (HTLV) might be the etiological agent of AIDS, with similar cell tropism and transmission route (Broder and Gallo 1984). After the recognition of the syndrome, within 2 years the causative agent, Human immunodeficiency virus (HIV) was isolated from the patients of AIDS. The historical data is listed in the table below.

<table>
<thead>
<tr>
<th>Year</th>
<th>Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>Silent period: HIV transmitted before AIDS.</td>
</tr>
<tr>
<td>1981</td>
<td>First few cases of AIDS detected.</td>
</tr>
<tr>
<td>1981</td>
<td>Epidemic of Pneumocystis carinii infection in L.A.</td>
</tr>
<tr>
<td>1982</td>
<td>Epidemic of Kaposi’s sarcoma in New York</td>
</tr>
<tr>
<td>1982</td>
<td>Definition of AIDS by C.D.C., Atlanta</td>
</tr>
<tr>
<td>Year</td>
<td>Event</td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
</tr>
<tr>
<td>1983</td>
<td>Isolation of HIV by Dr. Luc Montagnier in France</td>
</tr>
<tr>
<td>1984</td>
<td>Virus found by Gallo in USA</td>
</tr>
<tr>
<td>1985</td>
<td>ELISA blood test developed</td>
</tr>
<tr>
<td>1986</td>
<td>The International Committee of Nomenclature named it as the Human immunodeficiency Virus</td>
</tr>
<tr>
<td>1986</td>
<td>WHO special programme on AIDS formed</td>
</tr>
<tr>
<td>1987</td>
<td>AZT completes clinical trials</td>
</tr>
<tr>
<td>1991</td>
<td>HIV-2 virus found in AIDS patients in West Africa</td>
</tr>
<tr>
<td>1992</td>
<td>ddI approved</td>
</tr>
<tr>
<td>1995</td>
<td>ddC approved</td>
</tr>
<tr>
<td>1995</td>
<td>3TC approved</td>
</tr>
<tr>
<td>1996</td>
<td>First protease inhibitor approved (Saquinivir) Triple therapy (drug cocktails) lowers plasma HIV levels</td>
</tr>
</tbody>
</table>

### 1.2 Classification of Retroviruses

According to the International Committee on taxonomy of viruses, retroviruses are divided into oncoviruses, lentiviruses, and spumaviruses.

1. The oncoviruses contain the oncogenic retroviruses and are divided into type B, type C, and type D viruses on the basis of their morphology and genome structure.

2. The lentiviruses contain viruses (e.g., human immunodeficiency virus [HIV]) associated with slowly progressive, usually fatal conditions.
These in turn can be divided into three broad groups of which only one contains a human virus, human T cell leukemia virus type (HTLV) which is characterized by the proliferation of mainly CD4+ T-lymphocytes and the development of adult T cell leukemia (ATL).

1. Feline, murine, and avian leukemogenic retroviruses replicate efficiently in their hosts, produce a broad range of diseases, and interact with cellular proto-oncogenes to produce leukemia.

2. Human leukemogenic retroviruses are exemplified by the human viruses HTLV-I and HTLV-2 and the related bovine and simian viruses; they often establish a latent infection in their host cells, have a narrow disease spectrum, and carry a transactivating gene, tax, associated with transformation.

3. Simian type-D retroviruses and mouse mammary tumor virus make up the final group.

The detailed classification with examples is summarized in the table below.

<table>
<thead>
<tr>
<th>Classification of Retroviruses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genus</strong></td>
</tr>
<tr>
<td>1. Avian sarcoma and leukosis viral group</td>
</tr>
<tr>
<td>2. Mammalian B-type viral group</td>
</tr>
</tbody>
</table>
3 Murine leukemia-related viral group
   Moloney murine leukemia virus central, spherical simple core “C particles”
4 Human T-cell leukemia-bovine leukemia viral group
   human T-cell leukemia virus central, spherical complex core
e    Mason-Pfizer monkey virus cylindrical core “D simple particles”
5 D-type viral group
6 Lentiviruses
   human immunodeficiency virus cone-shaped core complex core
7 Spumaviruses
   human foamy virus central, spherical complex core
da Distinctive features seen in transmission electron micrographs.
b Groups 1 through 5 are presently (and, hopefully, temporarily) designated by the awkward descriptive terms listed in the table, awaiting the proposal of more succinct appellations by the International Committee on taxonomy of viruses.

1.3 Structure of HIV

HIV is a virus belonging to the family retroviridae. There are two types of retrovirus: 1) oncogenic or transforming retroviruses, which lead to neoplasms; and 2) cytopathic or lentiviruses, of which HIV is an example.

HIV is a RNA virus. HIV contains two strands of RNA in its innermost core, surrounded by two layers of structural proteins. Together, the RNA and these protein coats make up the nucleocapsid of the virus particle. (A virus particle is also called a virion.). Within the core is also an enzyme called reverse transcriptase. The structural proteins in these two coats are designated "p" for protein, along with a number that indicates the molecular weight. The outermost coat of the virus is a lipid bilayer derived from the host cell's membrane. The virus acquires this membrane as it is released (buds) from the host cell in which it replicated. Embedded in this bilayer are glycoproteins (i.e.,...
proteins with carbohydrate chains added) that are encoded by the *genome* (genetic material) of the virus. These viral glycoproteins are designated by the letters "gp" plus a number indicating the molecular weight. There are two major viral glycoproteins, called gp120 and gp41 (*i.e.*, they have molecular weights of 120,000 and 41,000, respectively) (Fig 1.).

**Fig 1: Structure of HIV-1**

Each virion expresses glycoprotein projections composed of gp120 and gp41. Within the envelope (Env) is the viral core, or nucleocapsid, which includes a layer of protein called p17 and an inner layer of protein, called p24. The HIV genome consists of two copies of single-stranded RNA, which are associated with two molecules of reverse transcriptase and proteins p7 and p9. The gp120 allows the virus to attach to T helper cells, since it binds to CD4 on the surface of the T helper cell. The affinity of gp120 for CD4 is very high. CD4 alone, however, is not sufficient to allow the virus to bind to and enter T helper cells. There are important "co-receptors" on the T cells that are also
necessary. These receptors normally bind chemokines, which are proteins produced by the body’s own cells. HIV, however, uses receptors for chemokines, along with CD4, to help it bind to and invade its target leukocytes. The gp41 in the viral envelope is essential for entry of the virus into host cells, since it allows the viral membrane to fuse with the membrane of the cell.

The genome of the virus is one long strand containing two end regions called long terminal repeats, or LTRs. These LTRs are the binding sites for molecules that normally activate genes in the cells. When these molecules bind to the LTRs, they will also activate the genome of the HIV, and the virus will begin replicating. Once the viral genome is activated, genes are transcribed to form messenger RNA, which is then translated into viral proteins. Three important genes are:

- **gag**: codes for the structural proteins of the nucleocapsid
- **pol**: codes for three important enzymes, reverse transcriptase, protease, and integrase.
- **env**: codes for gp120 and gp41.

In addition, some important regulatory genes are transcribed. The genes of HIV are quite unusual in that they do not code for individual proteins. Rather, they lead to production of large polyproteins, consisting of several individual proteins joined together. In this fused state, the proteins cannot function. They must be cleaved in order to form individual, functional proteins. Enzymes that break these polyproteins into smaller protein fragments are called proteases (Fig 2).
Fig 2: Genetic organization of HIV-1. The three major genes are \textit{gag}, \textit{pol}, and \textit{env}. These genes encode polyproteins that are cleaved to form the nucleocapsid core proteins, enzymes required for replication, and the envelope glycoproteins, respectively.

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Function</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Structural Genes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gag MA</td>
<td>P17</td>
<td>Nuclear transport of viral proteins (core), anchoring to membrane, Interaction with Env</td>
<td>Virion</td>
</tr>
<tr>
<td>Gag CA</td>
<td>P24</td>
<td>Viral core capsid protein</td>
<td>Virion</td>
</tr>
<tr>
<td>Gag NC</td>
<td>P7, p6</td>
<td>Nucleocapsid, binds with RNAs Binds with accessory protein Vpr</td>
<td>Virion</td>
</tr>
<tr>
<td>Protease</td>
<td>P15</td>
<td>Cleavage and maturation of gag-pol precursor</td>
<td>Virion</td>
</tr>
<tr>
<td>--------------</td>
<td>-----</td>
<td>---------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Reverse Transcriptase, RNase H</td>
<td>P66 p51</td>
<td>Reverse Transcription of viral RNA, RNase H activity</td>
<td>Virion</td>
</tr>
<tr>
<td>Integrase DNA</td>
<td>gp120 gp41</td>
<td>DNA provirus integration in Host cellular</td>
<td>Virion</td>
</tr>
<tr>
<td>Env</td>
<td>gp120/gp41</td>
<td>External viral glycoprotein which binds with CD4 and co-receptor of CD4+ T cells</td>
<td>Virion Env, Plasma Membrane (PM)</td>
</tr>
</tbody>
</table>

**Regulatory Genes**

<table>
<thead>
<tr>
<th>Tat</th>
<th>P16/p14</th>
<th>Transactivator protein which influence viral transcription</th>
<th>Host cell Nucleus(HCN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rev</td>
<td></td>
<td>Influence viral RNA transport, stability and functionality</td>
<td>HCN/nucleolus</td>
</tr>
<tr>
<td>Nef</td>
<td>P27/p25</td>
<td>Down-regulation of CD4, MHC-I on CD4+ cells</td>
<td>PM, cytoplasm</td>
</tr>
</tbody>
</table>

**Accessory Genes**

<table>
<thead>
<tr>
<th>Vif</th>
<th>P23</th>
<th>Role in viral maturation and counter Host antiviral proteins Apobec3G</th>
<th>Cytoplasm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vpr</td>
<td>P10</td>
<td>Inhibit host cell division at G2/M stage, Promotes nuclear localization of preintegration complex</td>
<td>Virion, host cell nucleus</td>
</tr>
<tr>
<td>Vpu</td>
<td></td>
<td>Regulate and influence ion channels of host cells, degrade CD4 receptors. It is present only in HIV-1</td>
<td>Integral membrane protein</td>
</tr>
<tr>
<td>Vpx</td>
<td></td>
<td>Only present in HIV-2 and SIV. It is an homologue of Vpr of HIV-1</td>
<td></td>
</tr>
</tbody>
</table>
1.4 Life cycle of HIV

First, as described, the gp120 of the virus contacts CD4 on a T helper cell. Macrophages and microglial cells (which are found in the central nervous system) also have CD4 and can be infected with HIV. When HIV first infects a person, it has a preference for invading macrophages and is called *macrophage-tropic* (or *m-tropic*). As the disease evolves, the virus mutates to preferentially infect T cells (*T-tropic*). Once this shift to a T-tropic strain occurs, the disease usually accelerates. This tropism is possible because HIV needs to interact with cell surface structures in addition to CD4 to infect a host cell. These other structures, as mentioned above, are termed *co-receptors*. Macrophages and T cells have different co-receptors, which allow HIV to discriminate between these two cell types. The co-receptor on macrophages is called CC chemokine receptor 5 (CCR5); the co-receptor on T cells is called CXC chemokine receptor 4 (CXCR4). Using gp41, the virus fuses with the T cell or macrophage and enters it. As the virus is internalized, its lipid and protein coats are shed. The viral single-stranded (ss) RNA is then released free into the cytoplasm of the cell.

Next, reverse transcriptase (RT) makes DNA from the ssRNA. This is the reverse of the normal process in the cell, whereby ss messenger RNA is transcribed from chromosomal DNA: hence the name reverse transcriptase. The two strands of ssRNA from the virus have now been converted into two molecules of double-stranded (ds) viral DNA. Subsequently, the viral enzyme integrase causes the viral dsDNA to be inserted or fused (*i.e.*, integrated) into a chromosome of the host cell. This integrated
viral genome is called a “provirus.” After integration into the host cell genome, either of two things could happen: 1. The cell can remain inactive, and the provirus will remain dormant, 2. Or if the host cell is activated, then the provirus will be activated, too. The provirus will not be activated unless the cell is activated. Whatever processes activate the T cells or macrophages will also cause cellular factors to bind to the LTRs of the provirus, thereby activating the provirus.

When the proviral genome is activated, large messenger RNAs that code for more than one protein are produced. These mRNAs produce large polyproteins that must be cleaved into the individual proteins of the virus by enzymes called proteases. First, a protease provided by the host cell cleaves the polyprotein product of the pol gene to form the viral proteins called reverse transcriptase, integrase, and viral protease. The viral protease then cleaves polyproteins made by the gag and env genes into individual proteins. The env gene codes for a large 160,000 molecular weight protein that is cleaved into gp120 and gp41 by the viral protease. At the same time that messenger RNA is being made from the proviral DNA, viral single-stranded RNA is also being produced to be packaged into new viral particles: The viral proteins and the viral ssRNA assemble together to form the new viral particles. This assembly usually occurs under the host cell membrane. The new particles then "bud" out of the cell. As the virus buds, gp120 and gp41 are inserted into the host cell membrane. The membrane then wraps around the new virion as it is released from the cell to provide its outermost coat (Fig 3).
1.5 AIDS

AIDS is staged or classified based on the number of CD4 cells in the patient's blood (indicated by a number), as well as by the types of opportunistic diseases from which the patient has suffered (indicated by a letter). A normal CD4 T-cell count is about 1100 cells/cu. mm (In Indian population the mean CD4 T-cell count is about 950 cells/cu. mm for male 395-1627 cells/cu. mm, female 511-1817 cells/cu. mm) (Uppal, Verma et al. 2003). In classifying AIDS, a count >500 is category 1, from 200-499 is category 2, and <200 is category 3. Clinically, a patient who is asymptomatic or has an acute infection or lymphoadenopathy only is category A. Category B includes patients with diarrhea, neuropathies, fever, Candida (yeast) infections, pelvic inflammatory disease. Category C, includes patients suffering from "heavy-duty" AIDS defining opportunistic infections such as Toxoplasma, Pneumocystis carinii, Mycobacterium, Mycobacterium avium complex (MAC), cryptococcal infection, disseminated TB. Patients who are least affected would therefore be classified as "A1", whereas those with advanced disease would be "C3". All combinations of numbers and letters are possible.

1.6 Clinical staging

CDC classification of AIDS indicator diseases (1993 REVISION)

The CDC categorization of HIV/AIDS is based on the lowest documented CD4 cell count (Table 1) and on previously diagnosed HIV-related conditions (Tables 2 and 3). For example, if a patient had a condition that once met the criteria for Category B but now is asymptomatic, the patient would remain in Category B. Additionally,
categorization is based on specific conditions, as indicated below. Patients in categories A3, B3, and C1-C3 are considered to have AIDS.

Table 1. CDC Classification System for HIV-Infected Adults and Adolescents

<table>
<thead>
<tr>
<th>D4 Cell Categories</th>
<th>Clinical Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Asymptomatic, Acute HIV, or PGL</td>
</tr>
<tr>
<td>(1) ≥500 cells/µL</td>
<td>A1</td>
</tr>
<tr>
<td>(2) 200-499 cells/µL</td>
<td>A2</td>
</tr>
<tr>
<td>(3) &lt;200 cells/µL</td>
<td>A3</td>
</tr>
</tbody>
</table>

Key to abbreviations: CDC = U.S. Centers for Disease Control and Prevention; PGL = persistent generalized lymphadenopathy.

Table 2. CDC Classification System: Category B Symptomatic Conditions

Category B symptomatic conditions are defined as symptomatic conditions occurring in an HIV-infected adolescent or adult that meet at least 1 of the following criteria:

a) They are attributed to HIV infection or indicate a defect in cell-mediated immunity.

b) They are considered to have a clinical course or management that is complicated by HIV infection.

Examples include, but are not limited to, the following:

1. Bacillary angiomatosis
2. Oropharyngeal candidiasis (thrush)
3. Vulvovaginal candidiasis, persistent or resistant
4. Pelvic inflammatory disease (PID)
5. Cervical dysplasia (moderate or severe)/cervical carcinoma in situ
6. Hairy leukoplakia, oral
7. Idiopathic thrombocytopenic purpura
8. Constitutional symptoms, such as fever (>38.5°C) or diarrhea lasting >1 month
9. Peripheral neuropathy
10. Herpes zoster (shingles), involving ≥2 episodes or ≥1 dermatome

Table 3. CDC Classification System: Category C AIDS-Indicator Conditions

1. Bacterial pneumonia, recurrent (≥2 episodes in 12 months)
2. Candidiasis of the bronchi, trachea, or lungs
3. Candidiasis, esophageal
4. Cervical carcinoma, invasive, confirmed by biopsy
5. Coccidioidomycosis, disseminated or extrapulmonary
6. Cryptococcosis, extrapulmonary
7. Cryptosporidiosis, chronic intestinal (>1-month duration)
8. Cytomegalovirus disease (other than liver, spleen, or nodes)
9. Encephalopathy, HIV-related
10. Herpes simplex: chronic ulcers (>1-month duration), or bronchitis, pneumonitis, or esophagitis
11. Histoplasmosis, disseminated or extrapulmonary
12. Isosporiasis, chronic intestinal (>1-month duration)
13. Kaposi sarcoma
14. Lymphoma, Burkitt, immunoblastic, or primary central nervous system
15. Mycobacterium avium complex (MAC) or M kansasii , disseminated or extrapulmonary
16. Mycobacterium tuberculosis , pulmonary or extrapulmonary
17. Mycobacterium , other species or unidentified species, disseminated or extrapulmonary
18. Pneumocystis jiroveci (formerly carinii ) pneumonia (PCP)
19. Progressive multifocal leukoencephalopathy (PML)
20. Salmonella septicemia, recurrent (nontyphoid)
21. Toxoplasmosis of brain
22. Wasting syndrome due to HIV (involuntary weight loss >10% of baseline body weight) associated with either chronic diarrhea (≥2 loose stools per day ≥1 month) or chronic weakness and documented fever ≥1 month

**WHO Staging System for Developing Countries**

The CD4 lymphocyte count is central to the 1993 CDC classification system for HIV disease and to all of the staging systems proposed in developing countries. Counting subsets of lymphocytes, however, is often not possible in developing countries because the required technology may not be available or may be too expensive for routine use if it is available. The Global Programme on AIDS of the World Health Organization (WHO) has proposed a simplified staging system that is clinically based and flexible enough to be used in different parts of the world. The system is based on four groups of clinical conditions that are considered to have prognostic significance and therefore constitute stages, plus an assessment of physical activity performance expressed as a four-point score. Patients are classified according to the highest stage recorded for either clinical condition or physical activity.
Table 3. World Health Organization Classification System for HIV Infection

(Source: http://hivinsite.ucsf.edu/InSite?page=kb-01-01)

**Clinical Stage 1**

1. Asymptomatic infection
2. Persistent generalized lymphadenopathy
3. Acute retroviral infection

   Performance Stage 1: asymptomatic, normal activity

**Clinical Stage 2**

4. Unintentional weight loss < 10% body weight
5. Minor mucocutaneous manifestations (e.g., dermatitis, prurigo, fungal nail infections, angular cheilitis)
6. Herpes zoster within previous 5 years
7. Recurrent upper respiratory tract infections

   Performance Stage 2: symptoms, but nearly fully ambulatory

**Clinical Stage 3**

8. Unintentional weight loss > 10% body weight
9. Chronic diarrhea > 1 month
10. Prolonged fever > 1 month (constant or intermittent)
11. Oral candidiasis
12. Oral hairy leukoplakia
13. Pulmonary tuberculosis within the previous year

14. Severe bacterial infections

15. Vulvovaginal candidiasis

   Performance Stage 3: in bed more than normal but < 50% of normal daytime
during the previous month

**Clinical Stage 4**

16. HIV wasting syndrome

17. Pneumocystis carinii pneumonia

18. Toxoplasmosis of the brain

19. Cryptosporidiosis with diarrhea > 1 month

20. Isosporiasis with diarrhea > 1 month

21. Cryptococcosis, extrapulmonary

22. Cytomegalovirus disease of an organ other than liver, spleen or lymph node

23. Herpes simplex virus infection, mucocutaneous

24. Progressive multifocal leukoencephalopathy

25. Any disseminated endemic mycosis (e.g., histoplasmosis)

26. Candidiasis of the esophagus, trachea, bronchi, or lung

27. Atypical mycobacteriosis, disseminated

28. Non-typhoid Salmonella septicemia

29. Extrapulmonary tuberculosis
30. Lymphoma

31. Kaposi's sarcoma

32. HIV encephalopathy

The infected CD4 cells in the peripheral blood are the tip of the iceberg. There is a significant reservoir of infected cells residing in the lymph nodes, lymphoid organs (e.g., thymus, spleen), and bone marrow. In a normal, healthy person, the ratio of CD4 T helper cells to CD8 T cytotoxic cells in the blood is about 2:1. This ratio can be reversed in AIDS. The CD4 cells may be depleted by any of several means:

(1) Direct infection of CD4 cells in the blood or lymphoid organs, leading to lysis by budding virus or shutdown of cell functions due to production of large amounts of virus particles or unintegrated viral DNA.

(2) Infected cells in the thymus or bone marrow fuse with one another to form giant, multinucleated cells called syncytia. These fused cells are not functional. Bone marrow and lymphoid organs becomes suppressed and do not produce mature T cells.

(3) Some of the infected CD4 cells will express gp120 on their surfaces and will be eliminated by T cytotoxic cells.

The collapse of the immune system in AIDS reflects the central role of the CD4 cell in the immune response. Not only is there depletion of CD4 cells, but also those that remain may not be functioning properly. As the normal function of the CD4+ T helper cells is lost, B cells begin to behave abnormally. Some of them start to produce immunoglobulins, but these antibodies don’t have any particular specificity. This
production of nonspecific antibodies is called polyclonal B cell activation. Depletion of macrophages also impairs the innate immune response (Fig 4).

**Fig 4: Typical course of HIV infection**

![Graph showing typical course of HIV infection](source.png)  
*Source: www.aids.org Adapted from Fauci et al., 1993a*

HIV infection is often diagnosed by assaying for these anti-HIV antibodies in the blood. When one is first infected with HIV, there is often an early, flu-like illness, which is hard to recognize as an HIV infection because the symptoms are so non-specific (e.g., fever, lymphadenopathy). This is called *acute retroviral syndrome*. During this time, antibodies are produced. As with any immune response, first IgM antibodies appear, then IgG. These antibodies are usually directed against the structural protein p24 or against gp160, the polyprotein that is the precursor to gp41 and gp120. The exact time frame of what happens serologically is as follows: First, p24 antigen appears in the blood. Next come IgM antibodies to p24, followed by IgG antibodies to p24. A little later, an IgG response to gp160 is seen. Current thinking is that it would be most effective to try to eliminate HIV in this acute retroviral syndrome stage or recently
being called as acute seroconverter, when the body is still capable of mounting a good immune response. However, this stage is hard to recognize -- unless the patient is known to be at high risk for HIV infection.

Ultimately, the immune response is ineffective, leading to chronic disease. There is a rapid clearing of virus from the blood after acute infection, but viruses remain within CD4 T cells and macrophages. Moreover, these viruses may mutate so that the initial immune response no longer recognizes them. The gp120 mutates quite rapidly. Some of these mutations may allow viruses to infect more efficiently and/or to evade the protection provided by the early immune response. As the infection proceeds, the immune response become weaker, and it will be less able to deal with any mutant viruses that arise. Eventually, HIV gains an edge over the immune response, and the outcome is almost invariably fatal.

1.7 AIDS AS ZOONOSIS

Evidence indicates that human immunodeficiency viruses (HIV-1 and HIV-2) entered the human population through multiple zoonotic infections from simian immunodeficiency virus (SIV)-infected nonhuman primates. Five lines of evidence have been used to substantiate the zoonotic origins of these viruses: (i) similarities in viral genome organization, (ii) phylogenetic relatedness, (iii) prevalence in the natural host, (iv) geographic coincidence, and (v) plausible routes of transmission (Gao, Bailes et al. 1999).
As such, they have been amply demonstrated to exhibit the remarkable (and perplexing) properties of insidious disease induction, persistence, latency, variation, recombination, and escape from immune and drug pressures. There are two distinct types of human AIDS viruses, HIV-1 and HIV-2, which are distinguished on the basis of their genome organizations and phylogenetic (i.e., evolutionary) relationships with other primate lentiviruses. Both have been further subclassified on the basis of phylogenetic criteria. Current data indicate that HIV-1 comprises three distinct virus groups (termed M, N, and O), with the predominant M group consisting of 9 clades denoted subtypes A through K. Similarly, HIV-2 strains infecting humans have been found to comprise six distinct phylogenetic lineages, subtypes A through F (http://hiv-web.lanl.gov) as shown in Fig 5.

Fig 5: Phylogenetic Association between HIV-1, HIV-2 and SIV
1.8 Isolation of HIV

The virus responsible for AIDS was first reported from Dr. Luc Montagnier’s Laboratory at the Institute Pasteur, Paris in May 1983. The virus was isolated from a lymph node of a person with Lymphadenopathy. The virus was designated as Lymphadenopathy Associated virus (LAV) (Barre-Sinoussi, Chermann et al. 1983). Subsequently in 1984, Gallo and his colleagues demonstrated that the virus was unequivocally the cause of AIDS and related conditions and named it HTLV-III i.e. Human T lymphotropic virus type III (Gallo, Salahuddin et al. 1984). Almost during the same period Levy et al isolated AIDS related virus (ARV) (Levy and Shimabukuro 1985). It was later concluded that all these viral isolates are genetic variants of the same virus. In 1986, The International Committee of Nomenclature named it as the Human immunodeficiency Virus (HIV) (Coffin, Haase et al. 1986; Coffin, Haase et al. 1986).

Molecular genetics data suggest that HIV type 1, the most common specie of HIV that infects humans, has been derived from the Simian Immunodeficiency Virus, SIVcpz, of the Pan Trogloodytes; troglodytes subspecies of chimpanzee, and hence shares high sequence homology with it.

1.9 Epidemiology of HIV

Global summary of the AIDS epidemic December 2007

<table>
<thead>
<tr>
<th>Number of people living with HIV in 2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total 33.2 million [30.6–36.1 million]</td>
</tr>
</tbody>
</table>
Adults 30.8 million [28.2–33.6 million]
Women 15.4 million [13.9–16.6 million]
Children under 15 years 2.5 million [2.2–2.6 million]

People newly infected with HIV in 2007

Total 2.5 million [1.8–4.1 million]
Adults 2.1 million [1.4–3.6 million]
Children under 15 years 420 000 [350 000–540 000]

AIDS deaths in 2007
Total 2.1 million [1.9–2.4 million]
Adults 1.7 million [1.6–2.1 million]
Children under 15 years 330 000 [310 000–380 000]

Based on the phylogenetic analysis of numerous isolates obtained from diverse geographical regions, HIV is divided into types, groups, subtypes, genotypes, circulating recombinant forms (CRFs) and unique recombinant forms (URFs). HIV-1 can be divided into 3 major Groups; M (major), N and O (Outlier). Group M is further subcategorized into many subtypes; A-D, F-H, J and K. Group O is distinctly different and genetically more closely related to simian immunodeficiency virus (SIV) and HIV-2. Group N appears to have arisen from interaction between a group M and a group O virus. Geographical distribution of HIV-1 subtypes is as shown in Fig 6.
During the later part of the epidemic recombination between co-infecting subtypes have given rise to recombinant viruses, some of which have established as CRFs and are transmitted horizontally as well as vertically. At least 16 different CRFs for Group M have been identified to date (http://hiv-web.lanl.gov). The majority of the CRF genomes are highly complex intersubtype recombinants. For instance, CRF01_AE is clearly E in env, U/A/E in the regulatory region and A in gag and pol, while the CRF04_cpx has an extremely complex genome structure including subtypes A, G, H, K and U. To, date CRFs have been identified in nearly every region of the world where two or more subtypes co-circulate and may account for 10 per cent of new HIV-1 infections. In some cases, unique recombinant form (URF) viruses seem to have originated by secondary recombination of a CRF (Lal, Chakrabarti et al. 2005). Mosaics
Involving CRF02_AG have been observed in various African countries, and some of the established CRFs, like CRF11 and CRF13, contain sequences that are derived from CRF01_AE. Recombinations between two CRFs have been described in studies from China and Niger (Lal, Chakrabarti et al. 2005).

In India, the first case of HIV infection was detected in 1986 from Tamil Nadu. Since then, the spread of HIV epidemic in India is a major public health concern. Although the proportion of people living with HIV in India is lower than previously estimated, its epidemic continues to affect large numbers of people. More accurate estimates of HIV indicate that approximately 2.5 million (2 million-3.1 million) people in India were living with HIV in 2006. (http://data.unaids.org/pub/EPISlides/2007/2007_epiupdate_en.pdf).

Genotypically more than 80% of the strains tested from India belonged to subtype C although some B/C and A/C recombinants have also been identified (Lole, Bollinger et al. 1999; Tripathy, Kulkarni et al. 2005). Sequence analysis for most subtype C isolates has revealed that these viruses display certain unique genetic characteristics that, by altering biological activity, may have contributed to the success for the predominance of this subtype. Examples include the presence of three or four NF-B enhancer copies in the long terminal repeat region, Tat and Rev proteins that are prematurely truncated, and a 15-bp insertion of the 5' end of the vpu reading frame, Subtype C also shows a preference for CCR5 coreceptor usage, and the variable V3 loop is relatively conserved in subtype C compared to other subtypes (Shankarappa, Chatterjee et al. 2001).
1.10 Why to study Env?

HIV-1, like many other viruses, is surrounded by a lipid membrane from which protrudes a virally encoded type I membrane protein (Env). The membrane of the virus and that of the cell present a formidable physical and energetic barrier between the viral genome and the cytoplasm of the host cell. To gain entry, all enveloped viruses mediate a membrane fusion reaction such that their lipid bilayers become contiguous with that of a cellular membrane (Hernandez, Hoffman et al. 1996). This process is invariably mediated by a viral fusion protein, such as HIV-1 Env. This homotrimeric protein is initially synthesized as a single polypeptide precursor that is posttranslationally cleaved into a surface subunit, gp120, that mediates receptor binding and that remains noncovalently attached to a transmembrane domain subunit, gp41 (Wyatt and Sodroski 1998). Cleavage liberates the NH2-terminal domain of gp41, a region that constitutes the protein's fusion peptide, a stretch of conserved hydrophobic residues that inserts into the membrane of the host cell during the course of membrane fusion. As such, the cleavage event is a prerequisite for viral infectivity.

The primary receptor for HIV-1 is the CD4 antigen, to which it binds via the gp120 subunit of Env. This causes structural alterations in gp120, enabling it to subsequently bind to a second receptor, termed a coreceptor. Coreceptor binding is thought to be the final trigger that leads to membrane fusion. All HIV-1 strains use one or both of the seven transmembrane domain chemokine receptors, CCR5 and CXCR4, as coreceptors in conjunction with CD4 for virus entry. The differential use of these
receptors, coupled with their patterns of expression, largely dictate the cell types that are susceptible to virus infection in vivo. Individuals who lack CCR5 due to a deletion in the CCR5 open reading frame are highly resistant to virus infection, but are immunologically normal and healthy (Liu, Paxton et al. 1996).

The most widely accepted model describing HIV-1 Env-mediated membrane fusion posits that either CD4 or coreceptor binding results in the formation of a coiled-coil in gp41. This is composed of three NH2-terminal leucine/isoleucine zipper regions, one contributed by each subunit of the Env trimer (Chan, Fass et al. 1997). The NH2-terminal fusion peptide is thereby displaced in the direction of the target membrane into which it can insert (Fig 7). As a result, Env transiently becomes an integral component of two membranes: the viral membrane in which it is anchored, and the cellular membrane that it has gaffed. The exterior surface of the coiled-coil contains grooves into which pack a second, more COOH-terminally oriented heptad repeat region of gp41. In other words, the gp41 subunit folds back on itself, forming an exceptionally stable six-helix bundle, in which the fusion peptide and transmembrane domain of gp41 are now oriented at the same end of the molecule (Weissenhorn, Dessen et al. 1997). Given the stability of this structure, it is likely that the six-helix bundle represents the terminal conformation of a fusogenic Env. Despite considerable differences in primary sequence, many triggered viral fusion proteins share a common core structure involving a six-helix bundle that has the membrane-associated domains at the same end. This indicates that many viral, and perhaps cellular, proteins induce membrane fusion by essentially the same mechanism (Chan and Kim 1998).
Fig 7: HIV-1 Env binding to CD4 and the conformational changes in Env.

A. Model for HIV-1 Env membrane fusion. Binding of CD4 to the gp120 subunit of Env induces exposure of a conserved region in gp120 implicated in coreceptor binding (purple; Rizzuto et al. 1998). In addition, CD4 binding appears to trigger exposure of the triple-stranded coiled-coil, and presumably exposure of the fusion peptide, although coreceptor binding could increase the efficiency and kinetics of this process. It is not known if the more COOH-terminal helical region in each gp120 subunit (red) interact with each other as drawn, but it is known that the extracellular domain of gp120 in general plays an important role in mediating Env oligomerization. Binding to coreceptor could bring Env in closer proximity to the target membrane, enabling the fusion peptide to insert in the bilayer, or it could impact formation of the six-helix bundle, the transition to which leads to membrane fusion. Note that in the six-helix bundle, the N-terminal helices form the core of the helix, with the COOH-terminal helices packing in the grooves on the outside of the structure. It is not known if gp120 remains associated with gp41 throughout the fusion process.

B. Formation of dead spikes. Binding of soluble CD4 to Env can induce shedding of gp120 from gp41, and can even induce formation of the six-helix bundle. A similar process is likely to occur at the cell surface. Such modified Env proteins are not fusogenic, but may serve as immunologic decoys.

Source: Rizzuto et al. 1998

1.10.1 HIV-1 Env binding to CD4 and Coreceptor

In addition to CD4, the HIV requires a coreceptor for entry into target cells. The chemokine receptors CXCR4 and CCR5, members of the G protein-coupled receptor superfamily, have been identified as the principal coreceptors for T cell line-tropic and macrophage-tropic HIV-1 isolates, respectively (Berger, Murphy et al. 1999). Macrophage-tropic HIV-1 viruses primarily use CCR5 (R5) as a coreceptor, whereas T-cell line-tropic viruses use CXCR4 (X4). Dual-tropic viruses (R5X4) use both coreceptors. (Fig 8) A subset of viruses can also use alternative coreceptors, such as CCR2b, CCR8.
Apj, STRL33 (BONZO/CXCR6), GPR1, GPR15 (BOB), CX3CR1 (V28), Chem R23, and RDC-1, for virus entry in transfected cells (Ohagen, Devitt et al. 2003). CCR5 is a receptor for the β-chemokines macrophage inflammatory protein (MIP)-1α and β and RANTES (Deng, Liu et al. 1996). The major coreceptor for syncytium-inducing strains appearing at the late stages of AIDS progression as well as T-cell line-adapted strains is CXCR4, a receptor for the CXC chemokine stroma cell-derived factor (SDF)-1α (Oberlin, Amara et al. 1996). These binding events result in the exposure of the previously hidden fusion peptide of gp41 and its penetration into the target cell membrane, which leads to fusion of the virion with the plasma membrane.

Fig 8: Use of receptors and Coreceptors

1.10.2 HIV-1 Env and Tropism

Although most strains of HIV-1 are able to infect primary lymphocytes, most molecularly cloned viruses are capable of infecting either primary macrophages or T-lymphoid cell lines, but not both cell types (Schuitemaker, Kootstra et al. 1991; Shioda,
Levy et al. 1991) T-cell line-tropic strains are able to induce syncytia in primary lymphocytes and are designated syncytium-inducing strains, whereas most macrophage-tropic isolates display the non-syncytium-inducing phenotype (Fig 9). Additionally, isolates able to infect T-cell lines replicate more rapidly and to higher levels in primary lymphocytes than macrophage-tropic isolates, which replicate more slowly and to lower levels (Schwartz, Felber et al. 1989).

Fig 9: Tropism of HIV-1: A model for the cellular determinants of HIV-1 tropism. Primary macrophages and primary lymphocytes express both CCR5 and CXCR4 (in addition to CD4), while T cell lines express only CXCR4. Macrophage (M)-tropic HIV-1 strains infect macrophages and lymphocytes via CCR5. T cell line (T)-tropic HIV-1 strains infect lymphocytes and T cell lines via CXCR4 but cannot use macrophage CXCR4 for entry. Dual-tropic strains infect all 3 cell types, either by using CCR5 and CXCR4 on macrophages and T cell lines, respectively, or through an ability to use CXCR4 on all target cell types.
The emergence of T-cell line-tropic syncytium-inducing isolates in infected patients correlates with the progression of HIV-1 infection as characterized by declining CD41 lymphocyte counts and the appearance of clinical symptoms of AIDS (Schuitemaker, Koot et al. 1992), whereas macrophage-tropic isolates that display the non-syncytium-inducing phenotype are generally found within the first few months after infection and persist throughout the course of infection (Asjo, Morfeldt-Manson et al. 1986).

The V3 loop plays an important role in regulating HIV-1 infection of other cell types in addition to macrophages, such as T-lymphoid cell lines. However, the V3 loop by itself is insufficient for conferring infection of T-cell lines. Nevertheless, substitution of a T-cell line-tropic V3 loop with a macrophage-tropic V3 loop results in loss of infection of T-cell lines (Cann, Churcher et al. 1992).

Interaction of the V3 loop with other portions of the Env protein is required to establish the proper conformation necessary for infection. A functional interaction between the variable region (V3) and the fourth conserved domain (C4) lying between the fourth and fifth hypervariable regions and is required for infection of T-lymphoid cell lines (Carrillo and Ratner 1996).

1.10.3 HIV-1 Env and Neutralization

The natural history of human immunodeficiency virus (HIV), infection shows that within weeks to months after infection, a detectable immune response associated
with a decline in viremia occurs. This immune response is characterized by the production of HIV-specific antibodies and the expansion of HIV-specific CD4 and CD8 T cells (Moog, Fleury et al. 1997). During the Asymptomatic phase after HIV infection, the level of the antibody responses remains high in the plasma of infected individuals. A neutralizing assay involving cell lines and laboratory-adapted strains allowed neutralizing antibodies (NAbs) to be detected in the plasma early after infection (Zwart, Back et al. 1994). The decreased antibody response associated with the loss of HIV-neutralizing antibodies accompanied clinical symptoms that signify AIDS. The HIV-1 envelope protein (Env) is the target of virus-neutralizing antibodies, but it does not normally elicit a strong neutralizing antibody response in infected individuals. The ability of HIV to evade the immune system has been associated in part with both the rapid variability of the HIV Env protein sequence and the masking of epitopes by glycosylation (Nara, Gerrity et al. 1991).

Most antibodies elicited against the HIV-1 envelope glycoproteins during natural infection or after vaccination are incapable of neutralizing HIV-1 infectivity in vitro (Moore and Ho 1995; Mascola, Snyder et al. 1996). While several such antibodies effectively neutralize viruses that are adapted to replicate in immortalized T-cell lines, three monoclonal antibodies, IgG1b12, 2G12 and 2F5, neutralize a wide range of primary HIV-1 isolates. These antibodies have a high affinity for the native trimer, which is comparable to or sometimes greater than their affinity for the individual gp120 or gp41 subunits. These antibodies may therefore have been raised by an immune
response to virion rather than to viral debris or dissociated subunits (Burton and Barbas 1994; Muster, Guinea et al. 1994).

These three monoclonal antibodies (MAbs) exhibit a higher affinity for oligomeric HIV-1 Env glycoproteins on viruses or cell surfaces than do most antibodies directed against the Env glycoproteins (Parren, Gauduin et al. 1997; Parren, Gauduin et al. 1997). Two of these MAbs bind to the surface glycoprotein, gp120, which is the viral receptor for CD4 and chemokine receptors CCR5 and CXCR4. These MAbs are IgG1b12, which recognizes an epitope overlapping the CD4 receptor site (Burton, Pyati et al. 1994), and 2G12, which recognizes an epitope, based around the C4/V4 region of gp120 and is highly sensitive to the presence of N-linked glycans in this region (Kunert, Ruker et al. 1998; Scanlan, Pantophlet et al. 2002). The MAb, 2F5, binds to an epitope involving a linear motif (ELDKWA) on the membrane proximal region of the transmembrane envelope protein gp41 (Conley, Kessler et al. 1994). Recently, two more MAbs namely, Z13 and 4E10, have been described which recognize a region close to the C terminus of the 2F5 epitope (Zwick, Labrijn et al. 2001).

Studies with HIV-1 subtype C infected individuals by heterosexual transmission have shown that there is preferential transmission of neutralization sensitive virus, containing fewer N-linked glycosylation sites. The exposure of neutralizing epitopes, which are lost in chronic infection because of immune escape, appears to be favored in the newly infected host (Derdeyn, Decker et al. 2004), although this process may not hold for transmission of subtype B HIV-1 (Chohan, Lang et al. 2005).
1.10.4 HIV-1 Env and Vaccine Development

Recombinant protein subunit vaccines based on X4-tropic viral isolates represented the first generation of HIV candidate vaccine strategies. To date, most recombinant HIV-1 glycoproteins tested, as vaccine candidates have been gp120 monomers, which elicit weak antibody responses to homologous viruses and have not been successful at inducing antibodies able to neutralize heterologous primary isolates at significant titers (Poon, Hsu et al. 2005). The antibody responses to plasmid DNA encoding gp120 boosting with a recombinant gp120 protein subunit vaccine elicited cellular and humoral response, but the response was not very effective (Barnett, Rajasekar et al. 1997). Use of live attenuated canarypox virus expressing HIV env and gag induced responses against laboratory isolates (Belshe, Gorse et al. 1998). Phase I studies of volunteers not infected with HIV-1 have shown that immunization with envelope subunit vaccine products elicits antibodies that neutralize only laboratory-adapted HIV-1 strains in vitro. Immune responses were not usually effective in neutralizing primary HIV-1 isolates (Mascola, Snyder et al. 1996; VanCott, Mascola et al. 1999). Similarly, DNA vaccine strategies to date have generally resulted in the induction of low-titer antibody responses, whereas plasmid DNA vaccination and a number of recombinant-vector-based strategies have been shown to induce cellular immune responses against internal proteins.

Considering the facts, a vaccine that elicits broadly specific Nab response along with a robust cellular response which might impart sterilizing immunity, preventing
the virus from binding to and entering the target cells and thereby protect the individual is required. One of the biggest problems in elicitation of the potent Nab response is the virus’ ability to escape from. HIV’s Env protein that is the primary target for Nabs participates in many important stages of the viral replication cycle and the host responses. A number of factors contribute to this genetic diversity such as-

1. Selection for coreceptor usage can generate env diversity,
2. Escape from cellular responses also drives diversity in env, but usually cellular responses are stronger towards other HIV genes,
3. Selection by NAbs results in rapid, continuous evolution of viral escape at the phenotypic level, and hence may be the most significant driving force.

Three mechanisms can contribute to the escape from NAbs: point mutations, changes in glycosylation patterns, and insertions and deletions in the Env. HIV-1 Env is a highly glycosylated protein. As much as half of the molecular weight of Env protein is contributed by the glycosylation, the glycosylation patterns in HIV-1 Env are mostly of N-linked glycosylation. The glycosylation sites are mostly well conserved except those found in variable regions of Env. The glycosylation plays major roles in -1. Proper folding of the protein, giving a stable three-dimensional structure to the protein, 2. Evasion of immune recognition by exploiting the host glycosylation machinery as a “glycan shield” to protect potential protein antigenic epitopes.

There are reports on evolution of Env and pattern of glycosylation, but most of the results are based on western viruses which are subtype B based, moreover most of the published data is with the laboratory adapted strains which may not hold true for
primary isolates. Moreover, the transmembrane subunit gp41 of HIV-1 Env consist of cytoplasmic tail which is almost 150 amino acid long and is required for various functions as -1. Virion incorporation into host cell, 2. Virus infectivity and pathogenicity, 3. And the truncated env gene without cytoplasmic tail is known to induce better immune response.

Env has been one of the components of several vaccine formulations being responsible for inducing one of the major of the immune system. However systematic studies to understand sequence variation in env, and the implications of the glycosylation moieties on immune recognition and viral fitness are limited. Hence, this work was initiated to get insight for answering the following questions.

1. How genetically diverse is Indian subtype C env (gp120 and gp41 as well) from env of closely related African subtype C, well-studied subtype B and other major clades of HIV-1?

2. If some of the crucial N-linked glycosylation sites are modified, then whether newer potential epitopes which are otherwise hidden will get exposed on the surface which might lead to a better immune response as compared to wild type envelope?

3. What happens if the part of the cytoplasmic domain is deleted, and whether it leads to better immunogenicity? In brief, this work was aimed at studying a unique, unexplored niche (in HIV-1 Indian subtype C) in Env characterization and our results will supplement other studies for better understanding phylogenetic, immunological and functional relevance of Env.