Abstract
**Abstract**

HIV-1, like many other viruses, is surrounded by a lipid membrane from which protrudes a virally encoded type I membrane protein, envelope (Env). The Env is initially synthesized as a single polypeptide precursor which gets cleaved by cellular proteases into a surface subunit, gp120 that remains non-covalently attached to a transmembrane domain subunit, gp41. The primary receptor for HIV-1 is the CD4 antigen, to which the gp120 subunit of Env binds. This causes structural alterations in gp120, enabling it to bind subsequently to a second receptor, termed a co receptor and this binding gives the final trigger that leads to membrane fusion.

The HIV-1 Env protein is the target of virus-neutralizing antibodies (NAbs), but it does not normally elicit a strong NAb response in infected individuals. Most of these antibodies elicited against the HIV-1 Env during natural infection or after vaccination are incapable of neutralizing HIV-1 infectivity in vitro. While several such antibodies effectively neutralize viruses that are adapted to replicate in immortalized T-cell lines (TCLA strains).

HIV has ability to evade the immune system, which has been associated with the rapid variability of the HIV env sequence and the masking of epitopes by glycosylation. Three mechanisms can contribute to the escape from NAbs: point mutations, changes in glycosylation patterns, and insertions and deletions in the HIV
Env. HIV-1 Env is a highly glycosylated protein. Removal of these glycosylation sites increases neutralization sensitivity of virus. There are reports on evolution of Env and pattern of glycosylation, but most of the results are based on subtype B viruses obtained from developed countries. Moreover most of the published data is with the TCLA strains which may not hold true for primary isolates. 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 (NLGs). In addition, the transmembrane subunit gp41 of HIV-1 Env consists of the cytoplasmic tail (CT) which is required for various functions of the virus. Truncation in the CT, affects the biological functions of the Env glycoprotein. The CT of Env is known to alter the biochemical and immunologic properties of the Env ectodomain, and also increases neutralization sensitivity of virus.

HIV-1 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 role of the cytoplasmic tail are limited. Hence, this work was initiated to get insight with the following objectives.

1. To determine genetic diversity of Indian subtype C gp120 from env of closely related African subtype C, well-studied subtype B and other major clades of HIV-1.
2. To determine genetic diversity of Indian subtype C gp41 from env of closely related African subtype C, well-studied subtype B and other major clades of HIV-1.

3. To determine the role of the crucial N-linked glycosylation sites on HIV-1 Env protein function and immunogenicity.

4. To determine the role of the cytoplasmic domain on HIV-1 Env protein function and immunogenicity.

In brief, this work was aimed at studying a unique, unexplored niche (in HIV-1 Indian subtype C) in Env and our results would supplement other studies for better understanding the phylogenetic, functional and immunological relevance of HIV-1 Env.

**gp120 sequences from HIV-1 subtype C early seroconverters in India.**

The study participants included six recently seroconverted individuals from a cohort of persons with high-risk behavior during their follow-up visits. PBMCs were separated from non-coagulated blood and DNA was extracted, followed by nested PCR for gp120 amplification. The amplified gp120 was cloned and sequenced. HMA analysis confirmed HIV-1 genotype C-3 infection in all the six study participants.

The phylogenetic analysis revealed that all Indian sequences segregated away within subtype C forming a tight cluster. This may be due to the founder effect.
indicating that the viruses have descended from a single strain introduced at the beginning of the epidemic in India as reported earlier. Similarity Plot analysis confirmed the phylogenetic results, as higher percent similarity was observed between the Indian subtype C sequences when compared with non-Indian subtype C potential vaccine candidate strain sequences and sequences representing other clades.

The constant domains of gp120 were relatively conserved in all the sequences analyzed. Among the variable loops, V1/V2, V4 and V5 were found to be more variable compared to V3 region. The V3 domain was found to be well conserved containing the GPGQ motif and absence of potential N-linked glycosylation site at the 5’ end of V3 loop, which is a characteristic feature of subtype C viruses. The cysteine residues in the gp120 region of all the sequences were well conserved.

In silico prediction for CTL and antibody epitopes was performed for gp120 sequences. Total of 23 CTL epitopes were predicted for gp120 region of env sequences out of which, 7 predicted CTL epitopes are well conserved with those reported in the HIV immunology database from Loa Alamos. Out of rest 16 epitopes, 12 predicted CTL epitopes are unique to Indian subtype C sequences and 4 CTL epitopes are conserved across the non-Indian HIV-1 subtype C sequences. Total of 22 antibody epitopes were predicted for gp120 region of env sequences out of which, 9 predicted epitopes are conserved with those reported in the HIV database. Out of the remaining 13 epitopes, 5 epitopes are unique for Indian subtype C sequences, and 8 epitopes are well conserved across non-Indian subtype C sequences.
Molecular Analysis of gp41 Sequences of HIV-1 Subtype C from India.

The study population was comprised of eight HIV-1 infected individuals. Route of infection for seven subjects was through sexual contact and vertical transmission from mother to child had occurred for one individual. Twenty sequences obtained from clones representing eight HIV-1 study participants were analyzed using subtype C specific primers designed for gp41 region after sequencing.

The gp41 sequences including sequences of the other subtypes were spanned thoroughly using similarity program and the percent similarity indicated that, major variabilities were located in 286bp- 506bp region. The phylogenetic tree for this region segregated the subtype C sequences from gp41 sequences of the other major subtypes suggesting that the region is highly polymorphic. Phylogenetic tree constructed for gp41 sequences, indicated that subtype C gp41 sequences clustered away from reference sequences of the other clades in M group forming a monophyletic lineage within subtype C. These results were in agreement with our earlier results of phylogenetic analysis for gp120 sequences. The tree further indicated that Indian subtype C sequences formed subclusters within Indian subtype C supporting the high genetic polymorphism in sequences representing each subject.

The amino acid residues of cytoplasmic tail and transmembrane were all almost conserved in our gp41 sequences. We could identify some potential replacements in some of the amino acid residues of gp41 region. HIV-1 gp41 is a multifunctional protein. It plays multiple roles in viral pathogenesis and replication.
The linear organization of gp41 includes N-and C-terminal heptad repeats, transmembrane and the cytoplasmic domain. The sequence variation in gp41, could affect the structure and function of the various domains and recognition by antibodies raised against gp41. The functional domains of gp41 were found to have substitutions in our sequences.

In silico prediction of CTL and antibody epitopes was performed for the gp41 sequences. Total of 27 CTL epitopes were predicted for gp41 sequences out of which, 7 predicted CTL epitopes are well conserved with those reported in the HIV database. Out of the remaining 20 epitopes, 5 predicted CTL epitopes are unique to Indian subtype C sequences and 15 CTL epitopes are conserved across the non-Indian HIV-1 subtype C sequences. Comparison of the predicted antibody epitopes with reported functional epitopes in the database revealed that very few epitopes were perfectly conserved. Hence, the predicted epitopes were analyzed for sequence conservation across the non-Indian subtype C sequences and total of 11 epitopes are well conserved across non-Indian subtype C sequences but they have not been listed in the HIV database.

**Modification in HIV-1 env from Primary Isolate of Indian Subtype C: Effect on Expression and Immunogenicity in Murine model.**

For this study, the HIV-1 TM protein was modified either by removal of the NLGs or by truncation of the CT, in addition to removal of conserved NLG site from gp120- V3 region from Indian subtype C HIV-1 Env. The effect of these modifications
was studied by expressing the native and mutant HIV-1 Env using the replication
defective human adenovirus serotype 5 system. The RAd5-Env constructs were used
for inoculation into mice and development of antibodies and subsequently their
neutralization ability was assessed.

The cytoplasmic tail of HIV Env is known to play important role in modulating
surface expression of the Env protein. The deletion in the CT to generate the
membrane anchored gp150 protein retained on the cell surface. Similarly, the mutation
in the CT to generate the secretory gp140 has shown better surface expression of the
Env protein in terms of the mean fluorescent intensity values as compared to native
gp160 by FACs. The CD4+MolT 4 infection and syncytia formation confirms that, both
gp150 and gp140 produces the oligomer similar to that of native gp160.

Mutation in the three consensus NLGs situat ed on the extracellular domain of
gp41 were assessed to determine the significance of each of these sites in relation to the
structure and function of the viral Env glycoprotein. Site-specific mutants of each of
the asparagine residues did not eliminate the ability of formation of syncytia in CD4+
MolT4 cells. In addition to that, significant effects neither on Env protein synthesis
could be observed as evidenced through western blot data, nor on the Env protein
conformation, processing, multimerization could be observed as evidenced through
FACS data.
The immune response following a single intramuscular administration of RAd5 HIV-1 Env constructs was evaluated separately for each construct with C57Bl/6 mouse pooled sera. In all the cases generation of mice IgG antibodies were determined along with their HIV-1 neutralization ability. All mice receiving the RAd5-Env produced an HIV-1 Env-specific antibody response, at week 4, 6 and 8 and the highest level was determined at week 6 after immunization.

HIV Env-specific neutralizing antibodies were also seen at week 6 for all these constructs. The neutralizing response against the homologous strain was dose dependant and higher RAd5-HIV-1 Env dose (1 X 10^8 PFU) was more potent for induction of neutralizing antibodies. However, no better immunogenicity could be observed in any of the mutant constructs as compared to the native construct. However, the construct with deletion in the cytoplasmic tail showed better expression as compared to the other constructs.

Overall, modification in HIV-1 Env has some effect on expression of the protein, but has no effect on biological function of the modified protein. The data also demonstrated that using an in vivo expression system with HIV-1 Env derived from a primary isolate can induce antibodies against HIV-1 and that such antibodies are capable of primary virus neutralization.