Summary
Summary

Majority of HIV-1 infections in Southern Africa and India are due to subtype C virus. The rapid increase in the predominance of subtype C might be due to behavioral or biological factors that facilitate transmission and select characteristics of the particular viral strains. HIV viruses isolated from patients in acute primary infection are more likely to be closely related to the transmitted virus. A potent immune response against the transmitted virus could lead to either prevention of HIV infection or modification of disease progression with better prognosis. The Env protein is critical for the entry of the virus into host cells and immune response to Env proteins may have major role in determining the outcome and the course of disease progression.

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

The gp120 exterior Env glycoprotein contains five successive constant (C1 to C5) and variable (V1 to V5) domains. There is marked variation in env gene and this probably is one of the mechanisms used by HIV to escape the host immune response. Hence, characterization of env gene becomes crucial aspect to study molecular evolutionary changes occurring in the sequences over a period of time and also at different stages of HIV infection. Limited data is currently available on gp120 sequences obtained from seroconverter stage virus, particularly for Indian subtype C.

- The gp120 sequences from the six recently infected individuals were analyzed thoroughly. HMA for V3-V5 region of env was performed for subtyping all.
HMA analysis confirmed HIV-1 genotype C-3 infection in all the six study participants.

- Consensus sequence developed from the alignment of all the gp120 sequences from the seroconverter indicated that the constant domains 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 cysteine residues which are required to form the loops in gp120 were well conserved in all six gp120 sequences.

- The C4 domain is positively charged and contains the basic residue lysine present in IKQ segment at amino acid 417-419, which is responsible for the overall positive charge of the CCR5 interactive surface of gp120. The other residues, proline-435 and glycine-438, are also critical for CCR5 binding and alterations in these could affect the binding process. These residues were conserved in all the gp120 sequences in the present study.

- There are 23-24 N-linked glycosylation sites in subtype C. In almost all the sequences, these sites were relatively conserved in all the domains except for those present in variable regions of gp120.

- However, the single N-linked glycosylation site in the V3 domain was conserved in all our sequences.

- The interclade phylogenetic tree constructed showed that all subtype C sequences segregated away from non-subtype C sequences, away from the consensus and ancestral M sequence.
The tree further showed that all Indian gp120 sequences clustered together within subtype C cluster with relative smaller branch lengths.

The similarity plot analysis showed maximum percent similarity (mean 85%; range 80% - 97%) as compared to non-Indian subtype C gp120 sequences (mean 70%; 67% - 95%). Within subtype C sequences, maximum sequence variation could be observed for V1-V2 and V4 regions.

Total of 23 HLA binding peptides 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 database, 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 potential antigenic sites were predicted for gp120 sequences out of which, 9 predicted epitopes are conserved with those reported in the HIV database, out of rest 13 epitopes, 5 epitopes are unique for Indian 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.**

HIV-1 gp41 is a multifunctional protein. It plays multiple roles in viral pathogenesis and replication. The HIV-1 transmembrane Env glycoprotein gp41 consists of fusion domain, two heptad repeats (the N-heptad repeat HR1 and the C-heptad repeat HR2), transmembrane domain (TM) and the long cytoplasmic tail. The NH2 terminal heptad
repeat 1 segment and COOH terminal heptad repeat 2 segment undergoes configuration to form the stable six-helix bundle of gp41 required for viral fusion.

- The gp41 sequences from eight HIV-1 infected individuals were characterized for sequence variability, presence of mutations with effect on function of gp41 and recognition of the epitopes by the immune system.

- The sequences were compared with representative sequences of the other clades, previously published Indian subtype C sequences, non-Indian subtype C sequences. Route of infection for seven subjects was through sexual contact and vertical transmission from mother to child had occurred for one individual.

- HMA for V3-V5 and variable region of gp41 \textit{env} confirmed subtype C infection in all the eight individuals.

- The gp41 sequences including representative sequences of the other subtypes were spanned thoroughly employing neighbor-joining matrix of SimPlot program and the percent similarity indicated that, major variabilities were located in 286bp- 506bp region.

- The phylogenetic tree constructed for 286bp- 506bp region segregated the subtype C sequences from gp41 sequences representing the other major subtypes.

- The Phylogenetic tree for complete gp41 sequences indicated that Indian gp41 sequences form monophyletic lineage within subtype C. The results were in agreement with our earlier results of phylogenetic analysis for gp120 sequences.
Further, the Indian sequences clustered with sequences from China and segregated away from sequences with African origin.

The N-linked glycosylation sites that are required for virus replication (position 104-106, 109-111, 118-120 and 130-132 in consensus) were conserved in all our sequences.

The Cys- (X)₅-Cys motif (position 91-97 in consensus) that plays a role in transmitting a conformational signal from receptor-bound gp120 to gp41 to induce the fusion-activated hairpin was found to be conserved in all the sequences.

The amino acid sequence PRRIR (position 344-348 in consensus) has major role in glycoprotein incorporation into virion was present in all sequences except NARI-IND-18, which had replacement of I347L.

The neutralizing epitope ERDRD (position 232-236 in consensus) defined for subtype B viruses had replacement of R233Q uniquely present in all our sequences.

The diaromatic residues 302Y and 303W of the KYLGSLVQYWGLELK sequence (position 294-308 in consensus) are required for infectivity of the virion were conserved in all our 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.**

We proposed the current work, with the hypothesis that, “the removal of the NLGs and the deletion in the cytoplasmic tail would help the some epitopes which are otherwise hidden to get exposed on the surface which would lead to better expression of HIV-1 Env protein and would lead to better immune response”.

- We describe results of efforts to develop anti-HIV-1 humoral response in murine model.

- In this study, the HIV-1 TM protein was modified either by removal of the NLGs or by truncation of the cytoplasmic tail, in addition to removal of conserved NLG from gp120- V3 region.
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 long 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 as compared to native gp160.

CD4+MolT 4 infection and syncytia formation confirms that, gp150 and gp140 in the present study produces the oligomer as that of native gp160.

Mutation in the three out of four consensus NLGs situated 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 RAd5-Env constructs were used for inoculation into mice and development of antibodies and subsequently their neutralization ability was assessed.

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.

The antibody responses developed in mice were found to neutralize the autologous parental virus. 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, in spite of the fact that the construct with deletion in the cytoplasmic tail expressed better than rest of the other constructs.

Overall, the data 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.