GENETIC CHARACTERISTICS OF THE HYPER-VARIABLE REGION (V3-V5) OF *env* GENE OF HIV-1 DISTINGUISHING SUBTYPE C FROM SUBTYPE B

Supriya Singh¹,², Shashi Khare¹, D S Rawat¹, Saaransh Jain¹, R L Ichhpujani¹, Sudha Prasad², Megha Singhal¹,², L S Chauhan¹ and Arvind Rai¹*

¹National Centre for Disease Control (NCDC), Delhi, India
²Maulana Azad Medical College (MAMC), India

*Corresponding author: Email: drarvindrai21@gmail.com, Phone No.: +91 9810385357

ABSTRACT

**Aim:** Carry out molecular characterization of subtype B and C sequences encompassing the hyper-variable region (V3-V5) of envelope (*env*) gene of HIV-1 to determine the genetic characters that distinguish and may possibly account for differential pathogenicity of the two subtypes. **Experimental Design:** Blood samples were collected from 25 HIV-1 sero-positive patients from India. Hyper-variable region (V3-V5) of *env* gene was amplified and sequenced. 173 sequences belonging to subtypes B and C were retrieved from Los Alamos database. Co-receptor prediction, antigenic loop, co-receptor variability in CCR5 and CXCR4 utilizing subtype B and C sequences, N glycosylation site, signature sequence and phylogenetic analysis was carried out. **Results:** Loss of N glycosylation sites in CXCR4 co-receptor utilizing sequences were observed in 37% of subtype B and 23% of subtype C sequences. Antigenic tip in majority of the subtype C isolates was GPGQTFY. Significant entropy differences were observed in CCR5 and CXCR4 co-receptor utilizing sequences of subtypes B and C. We observed several signature amino acid residues which occurred in high frequencies in Subtype C CCR5/CXCR4 co-receptor utilizing sequences. **Conclusions:** In conclusion, the determined molecular characteristics of subtypes B and C may partially be involved in the high pathogenicity of subtype C.

Keywords: HIV; Co-receptor; V3 loop; N-glycosylation; Envelope; Hyper-variable

INTRODUCTION

As per the UNAIDS Report 2010, since the beginning of the epidemic of HIV in 1981 the deadly virus has infected almost 60 million people worldwide and claimed about 25 million lives due to HIV-related causes (1). India harbours world’s third largest HIV burden after South Africa and Nigeria. Estimates of the department of AIDS control (2009-10) report that 2.27 million people are living with HIV/AIDS in India (2). Human immunodeficiency virus type-1 (HIV-1) exhibits an extensive genetic variability. Among the different subtypes of HIV-1, subtypes B and C are the major subtypes worldwide. HIV-1 subtype C is the predominant subtype in India and is considered to be more pathogenic than subtype B (3, 4, 5, 6). The envelope (*env*) gene (gp120) plays a crucial role in the process of entry of HIV-1 into target cells. The five variable regions (V1-V5) of the *env* encompass the potential pathogenic region of the virus. The V1-V2 region influence replication efficiency, V3 region is important in host cell tropism, V4 and V5 regions are involved in CD4 binding and neutralizing antibody responses (7). The virus attacks human cells by interacting with CD4 receptor which is followed by a conformational change in gp 120. This exposes the co-receptor binding site for CXCR4 or/and CCR5 (major co-receptors for HIV).
Based on differences in growth properties and cytopathic effects on PBMC’s, HIV produces syncytium inducing (SI) and non-syncytium inducing (NSI) viruses in vitro. NSI viruses utilize CCR5 as their major co-receptor, whereas SI viruses utilize the CXCR4 co-receptor. Newly infected patients primarily exhibit CCR5 (R5) co-receptor usage whereas patients in later stages exhibit CXCR4 (X4) co-receptor usage (8). Co-receptor determination can be useful in monitoring disease progression. V3 loop of HIV is a significant region of the hyper-variable region of env as its amino acid sequence can be used for prediction of co-receptor and give a fair idea about the utilized co-receptor while avoiding expensive phenotypic tests. N-glycosylation of the V3 loop shields the virus and prevents it from the neutralizing antibody responses/ host immune responses (9). The antigenic tip of the V3 loop possesses a conserved secondary structure. It is also a target for hosts neutralizing antibodies and is therefore an important determinant of viral pathogenicity (10). Most of the published data on co-receptor utilization, replication efficiency and cytopathic effect relates to HIV-1 subtype B; while limited information is available with regard to subtype C (11, 12, 13, 14). The present study was carried out on Indian samples and sequences retrieved from database to understand the epidemic of HIV-1 subtypes B and C through phylogenetic analysis and molecular characterization of the two subtypes.

MATERIALS & METHODS

Sample collection and study subjects
Ethical approval was obtained from the ethical committee of National Centre for Disease Control (NCDC) and Maulana Azad Medical College (MAMC) & Associated Lok Nayak Hospital. HIV positive cases from different risk groups including heterosexual males, heterosexual females and antenatal risk group were recruited for the study. Informed consent was obtained from all participants.

Peripheral Blood Mononuclear Cells (PBMC) Isolation and DNA Extraction
Whole blood (3-4ml) was drawn in K2EDTA vacutainers and PBMC’s were isolated immediately after collection using Ficoll’s Reagent. Genomic DNA, including proviral DNA of HIV was extracted with the help of QIAamp DNA Blood Mini kit, according to the manufacturer’s protocol.

Polymerase Chain Reaction (PCR)
PCR for β-globin was carried out as described previously by Khoja S et al (15). This was done to check the quality of the genomic DNA. PCR was carried out for partial gag gene of HIV to confirm the serological diagnosis. PCR was carried out as has been described previously by Granqvist et al. PCR for partial env gene (V3-V5) of HIV was carried out (11). Amplifications were carried out in ABI 9700 thermal cycler. The PCR reaction was carried out using Go Taq PCR core system II (Promega) as per the kit protocol.

Automated nucleotide sequencing
PCR products were purified using Promega Wizard SV Gel & PCR Clean up system. Purified PCR products were subjected to automated nucleotide sequencing with forward and reverse primers separately. Sequencing was carried out using Big Dye terminator kit (ABI, USA) using the kit protocol.

Analysis of Sequences
Gene sequences for env obtained in the present study were submitted to GenBank at www.ncbi.nlm.nih.gov and accession numbers were obtained. A BLAST search was carried out in order to confirm identity of the strains. Samples were assigned subtypes based on closest homology found with the subtype references in NCBI. DNA and protein alignments were created using Clustal X program. REGA subtyping tool was used to determine the subtype of our samples (16, 17). Phylogenetic analysis was performed using maximum likelihood with the help of MEGA version 5 and the reliability of the branching orders was determined by bootstrap method (18).

Co-receptor prediction: Co-receptor prediction was performed using the PSSM tool (C-PSSM, http://indra.mullins.microbiol.washington.edu/webpssm/), (ds) Kernel (http://genome.ulaval.ca/hiv-dskernel), Geno2pheno (http://coreceptor.bioinf.mpi-inf.mpg.de/).
Antigenic tip: Deduced amino acid sequences from the nucleotide sequences, using MEGA software, was used to study antigenic tip.

Sequence variability: Sequence variability was determined in the V3 loop for CCR5/CXCR4 co-receptor utilizing subtypes B & C sequences using “Entropy two tool” in Los Alamos Database. This tool was used to construct sets of aligned sequences and determine any positions with significant variation at the corresponding position in the other set (18).


RESULTS
25 HIV-1 sero-positive subjects were enrolled in the study from various risk groups, 117 including heterosexual male, heterosexual female and antenatal risk group.

Amplification of samples by PCR
All 25 samples were found to be positive for the ~142 bp region of gag gene of HIV. The samples were then amplified for hyper-variable (V3-V5) region of env gene of HIV using nested PCR. Only 16 samples exhibited amplification for the env gene.

Sequencing and GenBank Submission
The sequences obtained were submitted to GenBank and accession numbers were obtained (HM630277- HM630292).

BLAST search and phylogenetic analysis
BLAST search revealed that our isolates belonged to HIV-1 env gene (V3-V5). The samples were found to belong to subtype C. Samples were aligned using Bioedit version 7.0.9.0. Phylogenetic analysis was carried out using the MEGA software version 5 with Maximum likelihood method (Figure 1).

N-linked glycosylation
In the CXCR4 utilizing V3 sequences, loss of the glycosylation site was observed in 37% sequences in subtype B sequences and 23% sequences in subtype C sequences.

Antigenic tip
Central domain of the V loop identified as the principal neutralizing determinant comprising GPGR motif determined the cross clade neutralizing activity of a human monoclonal antibody. Our isolates showed GPGQTFY to be the antigenic tip most common in subtype C V3 loop sequences in about 87.5% cases. GPGQAFTY motif commonly found in subtype B isolates was found in 12.5% of our subtype C isolates.

Co-receptor prediction
Fourteen sequences were found to be predictive of utilizing CCR5 co-receptor by the three methods.

V3 loop sequence variation
In both subtypes B and C, increase in sequence variability was observed in the V3 loop sequences in CXCR4 co-receptor utilizing sequences when compared with CCR5 co-receptor utilizing sequences (Figure 2). The subtype C exhibited increased sequence variability for CCR5 and CXCR4 co-receptor utilizing sequences when compared with subtype B.

Signature Residues
Signature residues include arginine at position 15 (in 96.6% isolates), glutamine 146 at position 23 (in 100% isolates) in the CCR5 co-receptor utilizing
sequences in subtype C. A signature residue in the CXCR4 co148 receptor utilizing sequence was at position 15 (in 86%) in subtype C. [Figure 3, Table 2].

DISCUSSION

Out of the 25 HIV-1 gag gene positive samples, only 16 (~64%) successfully amplified the V3-V5 region of the env gene. The unsuccessful amplification in the 9 samples could be due to mutations in the primer binding region in our Indian isolates. All sequences from this study clustered with previously published subtype C sequences indicating that subtype C is the subtype currently circulating in the Indian isolates. Regarding the epidemiology and transmission of the epidemic of HIV, with reference to subtypes B and C, sequences of subtype B, subtype C and recombinants clustered separately with the bootstrap values of the major nodes being very high. The neighbouring countries of India such as China and Myanmar exhibited clustering of isolates with our sequences. This is indicative of transmission of the infection between India, China and Myanmar and thus is essential to monitor the risk behaviours through surveillance in order to prevent/reduce transmission. N-glycosylation of variable loops, such as the V3 loop, often restricts access to conserved host receptor binding sites. This limits their exposure to the host immune system, and confers a survival advantage for the virus. We observed in the CXCR4 utilizing V3 sequences loss of glycosylation site 37% sequences in subtype B sequences and loss of 23% glycosylation sites in subtype C sequences when compared with their respective glycosylation sites in CCR5 sequences. The loss of glycosylation sites in CXCR4 sequences suggests that possibly in the later stages of infection the loss of glycosylation sites providing a survival advantage to the virus. This advantage is observed in a higher percentage of B subtype B as compared with subtype C. This possibly may be due to the fact that the virus in subtype B mainly utilizes both co-receptors (CCR5 in early stages and CXCR4 in late stages of HIV infection), whereas in subtype C the virus utilizes mostly CCR5 co-receptor even in late stages. Thus, this viral escape strategy is not of much significance in subtype C, but is certainly important for the B subtype. The antigenic tip in the subtype C isolates from our study reveals the predominant antigenic tip to be GPGQTFY in 87.5% of our isolates, the rest showed GPGQAFY. Antigenic tip analysis is important for epitope-specific neutralizing antibody preparation and vaccine development. As has already been reported, the numbers of CXCR4-using isolates in subtype C are limited, and so was the case in our study, where all isolates were predictive of utilizing CCR5 co-receptor. It is possible that there might be a viral adaptation within the subtype C V3 loop that allows for a limited CXCR4 co-receptor usage. Signature residues in the V3 loop were therefore studied. The signature residues possibly could account for a high pathogenicity, since these occur in high percentages in subtype C CCR5/CXCR4 co-receptor utilizing isolates and the corresponding residue was replaced in subtype B. Effect of these amino acid residues needs to be studied in order to better understand the subtype’s high pathogenicity.

During the initial asymptomatic stage of infection or the CCR5 co-receptor utilizing stages, the sequence variability is less as the immune system exerts a selective pressure on the virus and therefore its sequence remains comparatively conserved. Eventually, as the infection progresses towards symptomatic phase or towards CCR4 utilizing co-receptor stages, the sequence variability increases as immune system of the body is now immune-compromised and is not able to exert selective pressure and allows the virus to become more variable. When comparing the CCR5 utilizing sequences in subtype B and C, significantly increased variability in the subtype C and similar is the case with CXCR4 utilizing co-receptor sequences of subtype C. The subtype C HIV-1 virus is infectious and the body is not so efficient in exerting a selective pressure on the virus, thus leading to comparatively increased entropy. In conclusion, our results on molecular characteristics of subtypes B and C show that these
possibly may be involved in the subtype C’s high pathogenicity.

**Funding**
Partial funding was received from Council of Scientific Research

**Conflict of Interest**
Authors declare no conflict of interest in this manuscript.

**Ethical Approval Declaration**
Ethical approval was obtained from the ethical committee of Maulana Azad Medical College & National Centre for Disease Control, Delhi.

**ACKNOWLEDGEMENTS**
Author SS acknowledges financial support from the Council of Scientific and Industrial Research (CSIR), Delhi, India, during the course of this study. We thank SS Negi, RS Rautela, Hanu Ram, Poonam Gupta and Sachin Kumar for helpful discussions. Authors also thank the staff of the HIV/AIDS Division of NCDC and for their assistance in collecting samples, and providing serological and clinical data.

**REFERENCES**
GENETIC CHARACTERISTICS OF THE HYPER-VARIABLE REGION


Table 1: Table showing the risk group, predicted co-receptor, and clinical stage of infection.

<table>
<thead>
<tr>
<th>Accession</th>
<th>Number</th>
<th>Risk group</th>
<th>Age/sex</th>
<th>VHS (E/H)</th>
<th>Score</th>
<th>Genotype</th>
<th>d/c/sex</th>
<th>Clinical</th>
<th>mandement</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV1-07_BC-CN.2005.CF135</td>
<td>ANC</td>
<td>FF</td>
<td>40-50</td>
<td>R5</td>
<td>R5</td>
<td>E5</td>
<td>E5</td>
<td>Autoprotic</td>
<td></td>
</tr>
<tr>
<td>HIV1-07_BC-CN.2005.CF135</td>
<td>ANC</td>
<td>FF</td>
<td>40-50</td>
<td>R5</td>
<td>R5</td>
<td>E5</td>
<td>E5</td>
<td>Autoprotic</td>
<td></td>
</tr>
<tr>
<td>HIV1-07_BC-CN.2005.CF135</td>
<td>BM</td>
<td>2PM</td>
<td>20-30</td>
<td>R5</td>
<td>R5</td>
<td>E5</td>
<td>E5</td>
<td>Autoprotic</td>
<td></td>
</tr>
<tr>
<td>HIV1-07_BC-CN.2005.CF135</td>
<td>BF</td>
<td>1PM</td>
<td>20-30</td>
<td>R5</td>
<td>R5</td>
<td>E5</td>
<td>E5</td>
<td>Autoprotic</td>
<td></td>
</tr>
<tr>
<td>HIV1-07_BC-CN.2005.CF135</td>
<td>BM</td>
<td>2PM</td>
<td>20-30</td>
<td>R5</td>
<td>R5</td>
<td>E5</td>
<td>E5</td>
<td>Autoprotic</td>
<td></td>
</tr>
<tr>
<td>HIV1-07_BC-CN.2005.CF135</td>
<td>BF</td>
<td>1PM</td>
<td>20-30</td>
<td>R5</td>
<td>R5</td>
<td>E5</td>
<td>E5</td>
<td>Autoprotic</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: (a) Table showing signature residues in the CCR5 co-receptor utilizing sequences of V3 loop. (b) Table showing the signature residues in the CXCR4 co-receptor utilizing sequences of V3 loop (Subtype C-Query; Subtype B-Background)

Arvind Rai, et al. 110
GENETIC CHARACTERISTICS OF THE HYPER-VARIABLE REGION

Arvind Rai, et al.

(a) 

(b) 

(c) 

(d) 

(e) 

(f) 

(g)
Figure 2: The above figure shows the entropy plots, constructed using ‘Entropy’ tool of the Los Alamos Database. On the x-axis is the amino acid position, whereas, the y-axis shows the entropy/sequence variability. The above entropy plots show sequence variability in (a) CCR5 utilizing V3 sequences of subtype C (b) CCR5 utilizing V3 sequences of subtype B (c) entropy difference between ‘a’ and ‘b’ (d) CXCR4 utilizing V3 sequences of subtype C (e) CXCR4 utilizing V3 sequences of subtype B (f) entropy difference between ‘d’ and ‘e’ (g) entropy differences between the CCR5 and CXCR4 utilizing co-receptors in subtype B sequences (h) entropy differences between the CCR5 and CXCR4 utilizing co-receptors in subtype C sequences. The sites with a significant ($p \leq 0.05$) difference in the entropy are shown in red.

Figure 3: Sequence logos of (a) CCR5 co-receptor utilizing subtype C sequences (b) CCR5 co-receptor utilizing subtype B sequences (c) CXCR4 co-receptor utilizing subtype C sequences (d) CXCR4 co-receptor utilizing subtype B sequences. Size of each character displays the proportion of sequences in which a particular amino acid occurs.