Diversity of HIV Type 1 Envelope (V3-V5) Sequence in HIV Type 1-Infected Indian Children

Somi Sankaran Prakash,¹ Rajesh Kalra,¹ Rakesh Lodha,² Sushil K. Kabra,² and Kalpana Luthra¹

Abstract

We assessed the viral envelope (V3-V5 region) sequence diversity from 13 HIV-1-infected Indian children from north India. All of the 13 children were found to be infected with subtype C viruses. One of the viral sequences exhibited usage of the CXCR4 coreceptor predicted by Web PSSM and Geno2pheno tools. This virus also had a longer V3 sequence with 37 amino acids, a GRGQ motif, and a methionine residue before it (AIIMS_307). A unique finding was the complete deletion of the V4 region of another virus (AIIMS_363). High sequence diversity was observed in the envelope of the HIV-1-infected Indian children.

ONE OF THE MAJOR CHALLENGES in designing a vaccine against human immunodeficiency virus (HIV) is its enormous genetic diversity. HIV-1 infection in India is predominantly due to clade C viruses, though other subtypes and recombinant forms have been reported to occur in a small percentage of infected people.¹ The molecular epidemiology of HIV-1 is changing in different parts of the world. One of the recent studies reported the occurrence of AG recombinant forms in HIV-1-infected children from north India.² In this context, it is important to sequence and acquire information on the envelope sequence variation in the viruses circulating in this region in order to design appropriate vaccine candidates against HIV-1.

Envelope glycoprotein gp120 of HIV-1 is divided into five constant and five variable regions. One of the variable regions, V3, is a highly conserved region in the HIV-1 envelope both in sequence and structure. This is because of the important role of the V3 region in binding to the coreceptors, either CCR5 or CXCR4, which occurs once HIV-1 gp120 is bound to the CD4 receptor on host cells.³ Our objective was to assess the envelope sequence diversity of HIV-1 infection in Indian children as limited information exists so far concerning this. We recently reported the neutralization antibody response elicited in HIV-1-infected children against HIV-1 subtype C primary isolates.⁴ We extend our findings in this study by assessing the sequence variability of the envelope (V3-V5 region) of the viruses infecting these children.

Thirteen HIV-1-infected children from north India living in New Delhi and the surrounding areas were recruited in this study. Written informed consent was obtained from the parents/guardians. The study was approved by the Institute ethics committee. Genomic DNA was isolated from the whole

blood/peripheral blood mononuclear cells (PBMCs) from these children. The HIV-1 envelope V3-V5 region was amplified by nested polymerase chain reaction (PCR) using the following primers: first round PCR (ENV1: TCAGCA CAGTACAATGTACACATGGAAT, HXB2 location: 6949–6976 and ENV2: GTGCTTCCTGCTCCCAAGAACCCA, HXB2 location: 7810–7784); second round PCR (ENV3: TGTAAATGGCAGTCTAGCAGAA, HXB2 location: 7003–7025 and ENV4: TTATATAATTCACCTCTCCAATT, HXB2 location: 7678–7656). The amplified products were purified and cloned into T/A cloning vector (RBC Biosciences, Taiwan). Random positive clones were picked up and sent for sequencing commercially (Macrogen Inc., Korea). Sequences were analyzed for the subtype using the REGA HIV-1 subtyping tool (Stanford drug resistance database; <http://hivdb.stanford.edu>) and Genotype tool available for retroviruses in the NCBI website (<http://ncbi.nlm.nih.gov/retroviruses>). Predicted coreceptor usage was determined using the web-based tools Web PSSM and Geno2pheno based on the sequence of the V3 region.^{5,6} The number of potential N-linked glycosylation sites (PNGS) was determined using the tool N-Glycosite available from the Los Alamos database (<http://hiv.lanl.gov>). Sequences were analyzed for the diversity of the V3, V4, and V5 regions by aligning the amino acid sequences using Clustal X. Phylogenetic analysis was done by using the HIV-1 sequence database (<http://hiv.lanl.gov>), Clustal X, Seaview, and MEGA4.0 softwares. Reference nucleotide sequences of different subtypes and circulating recombinant forms were downloaded from the Los Alamos database (<http://hiv.lanl.gov>). A phylogenetic tree was constructed by the neighbor-joining method by taking 1000 bootstrap replicates.⁷

¹Department of Biochemistry and ²Department of Pediatrics, All India Institute of Medical Sciences, New Delhi, India.

TABLE 1. CHARACTERISTICS OF HIV-1-INFECTED INDIAN CHILDREN

Patient ID	Age/sex	ART status	Clinical stage ^a	CD4 count (cells/ μ l)	Viral load (RNA copies/ml)
AIIMS_302	9/F	Treated	T1	555	2,450
AIIMS_304	10/M	Treated	T2	221	< 47
AIIMS_307	11/M	Treated	T2	106	20,500
AIIMS_328	6/F	Treated	T1	1140	< 47
AIIMS_352	10/F	Naive	Stage 1	290	11,400
AIIMS_353	12/F	Naive	Stage 1	379	60,400
AIIMS_356	11/M	Naive	Stage 1	308	104,000
AIIMS_358	5/M	Naive	Stage 1	1092	NA
AIIMS_359	10/M	Treated	T2	NA	29,500
AIIMS_361	1.5/F	Naive	Stage 1	NA	96,000
AIIMS_363	10/M	Treated	Stage 1	534	< 47
AIIMS_375	3.5/M	Naive	Stage 2	NA	< 47
AIIMS_376	8/M	Treated	T1	NA	28,400

^aClinical staging was done based on the WHO classification (2008). For antiretroviral-treated children, T-staging was done. NA, not available; M, male; F, female; ART, antiretroviral treatment.

with the GPGQ motif present, typical of subtype C viruses (Fig. 2). Interestingly, the sequence of the viral envelope from one patient (AIIMS_307) showed considerable variability in the V3 region with insertion of two amino acids (Fig. 2) with a GRGQ motif present. This virus also showed usage of CXCR4 in both the Web PSSM and Geno2pheno tools for the prediction of coreceptor usage. Insertion of a methionine residue before the GPGQ motif in subtype C viruses has been shown to be associated with CXCR4 usage.⁸ Also, a larger number of positive charges in the V3 region is known to be associated with CXCR4 usage.⁹ The V3 sequence of the above virus exhibited all these features, suggesting that it may be an X4 tropic virus. These changes in the V3 region are similar to reports already existing concerning subtype C viruses exhibiting CXCR4 usage.^{8,9} CXCR4 usage in subtype C viruses is rarely observed even in advanced stages of infection.^{10,11} This patient has been on antiretroviral therapy for the past 4 years, with a very low CD4 count and high viral load (Table 1). The envelope sequence of the virus infecting another patient (AIIMS_375) was also predicted to be showing CXCR4

usage by the Web PSSM program, while with Geno2pheno CCR5 usage was predicted.

The V4 and V5 envelope regions of these viruses were found to have insertion, deletion, and various amino acid substitutions at different places. Insertion, deletion, and substitutions are reported to occur with high frequency in the V4 region.^{12,13} Previous studies have shown that the insertions and deletions are common in the V4 region and occur due to a misalignment mechanism and were found to interfere with the processing of the HIV-1 envelope glycoprotein.^{12,14} The V4 region is also an important site of N-linked glycosylation. The glycosylation pattern within the V4 region shows high variation even within the same patient.¹⁵ This also helps the virus evade the host immune responses. A salient finding of this study is the complete deletion of the envelope V4 region of the virus infecting one patient (AIIMS_363) (Fig. 3). The number of potential N-linked glycosylation sites (PNGS) in the envelope (V3-V5) region of this virus was only eight, as compared to that in the viruses from the rest of the patients, which were in the range of 12–17 (Table 2).

ConsensusC_ENV	SSGGDLEITTHSFNCRGEFFYCNTSKLFNSTY---NST-----NST-----ITLPCRKQIINMWQEVGRAMYA
AIIMS302_ENV	SSGGDPEITTHSFNCGGEFFYCNTSRLFNAIY---NGTLIFNGTVFNDTSSSTNESITIPCRKQIINLWQEVGRAMYA
AIIMS304_ENV	SSGGDLEITTHSFNCRGEFFYCNTSGLFNATYNHTSGLFN--GTYSNGT---TITIPCRKQIINMWQEVGRAMYA
AIIMS307_ENV	PSGGDLEITTHSFNCAGEFFYCNTSGLFNSTY---NFTHNLNGTGGDPNS---TITLPCRKQIINMWQEVGRAMYA
AIIMS328_ENV	SSGGDLEITTHSFNCRGEFFYCNTSGLFNSTY---NFTHNLNGTGGDPNS---TITLPCRKQIINMWQEVGRAMYA
AIIMS352_ENV	SSGGDLEITTHSFNCRGEFFYCNTSGLFNSTY---NFTHNLNGTGGDPNS---TITLPCRKQIINMWQEVGRAMYA
AIIMS353_ENV	SSGGDLEITTHSFNCRGEFFYCNTSGLFNSTY---NFTHNLNGTGGDPNS---TITLPCRKQIINMWQEVGRAMYA
AIIMS356_ENV	SSGGDLEITTHSFNCRGEFFYCNTSGLFNSTY---NFTHNLNGTGGDPNS---TITLPCRKQIINMWQEVGRAMYA
AIIMS358_ENV	SSGGDLEITTHSFNCRGEFFYCNTSGLFNSTY---NFTHNLNGTGGDPNS---TITLPCRKQIINMWQEVGRAMYA
AIIMS359_ENV	SAGGDLEITTHSFNCRGEFFYCNTSGLFNSTY---NFTHNLNGTGGDPNS---TITLPCRKQIINMWQEVGRAMYA
AIIMS361_ENV	SSGGDLEITTHSFNCRGEFFYCNTSGLFNSTY---NFTHNLNGTGGDPNS---TITLPCRKQIINMWQEVGRAMYA
AIIMS363_ENV	SSGGDLEITTHSFNCRGEFFYCNTSGLFNSTY---NFTHNLNGTGGDPNS---TITLPCRKQIINMWQEVGRAMYA
AIIMS375_ENV	SSGGDLEITTHSFNCRGEFFYCNTSGLFNSTY---NFTHNLNGTGGDPNS---TITLPCRKQIINMWQEVGRAMYA
AIIMS376_ENV	NSGGDLEITTHSFNCRGEFFYCNTSGLFNSTYF--SNGTYMFNDTDMTNDTLRNGTITIPCRKQIINMWQEVGRAMYA

FIG. 3. Alignment of amino acid sequences of the V4 region of viral envelope from 13 HIV-1-infected Indian children. V4 sequences of viruses from 13 HIV-1-infected Indian children along with the adjoining constant regions were aligned using Clustal X software. The V4 sequence of consensus clade C HIV-1 was also included for comparison. The arrow is pointing to one patient's sequence (AIIMS363_ENV) that showed the complete deletion of the V4 region.

TABLE 2. GENETIC ANALYSIS OF THE *env* (V3–V5 REGION) IN HIV-1-INFECTED INDIAN CHILDREN

Patient ID	Subtype ^a	Predicted coreceptor usage ^b		Total number of positive charges in V3 sequence	Amino acids at 11/25 position of V3	Number of PNGS (potential N-linked glycosylation sites) ^c
		Web PSSM	Geno2pheno			
AIIMS_302	C	CCR5	CCR5	6	SD	17
AIIMS_304	C	CCR5	CCR5	6	SD	14
AIIMS_307	C	CXCR4 ^d	CXCR4	8	ID	15
AIIMS_328	C	CCR5	CCR5	6	SD	15
AIIMS_352	C	CCR5	CCR5	6	SD	14
AIIMS_353	C	CCR5	CCR5	6	SD	15
AIIMS_356	C	CCR5	CCR5	6	SE	13
AIIMS_358	C	CCR5	CCR5	5	SD	14
AIIMS_359	C	CCR5	CCR5	6	SG	16
AIIMS_361	C	CCR5	CCR5	6	SD	13
AIIMS_363 ^e	C	CCR5	CCR5	5	SD	8
AIIMS_375	C	CXCR4	CCR5	6	GE	16
AIIMS_376	C	CCR5	CCR5	6	SG	12

^aSubtyping was done by the NCBI HIV genotyping tool and REGA HIV-1 subtyping tool.

^bCoreceptor usage was determined using web-based tools Web PSSM and Geno2pheno.

^cNumber of PNGS was determined using the N-Glycosite tool available from the Los Alamos HIV database (<http://hiv.lanl.gov>).

^dCXCR4 usage was detected in 1 out of the 13 envelope sequences.

^eThe number of PNGS was significantly less in AIIMS_363, which had a complete deletion of the V4 region.

High variability of the V3–V5 regions of the viral envelope was detected in HIV-1-infected children in north India. A close phylogenetic relationship was observed among some of the sequences analyzed. One of the possible reasons for that could be due to a shared source of infection in these patients. The present study sheds more light on the molecular epidemiology of HIV-1 infection in Indian children.

GenBank Accession Numbers

The nucleotide sequences of 13 HIV-1-infected Indian children have been submitted to GenBank with the following accession numbers: HQ845212, HQ845213, HQ845224, HQ845231, HQ845232, and JN176191–JN176198.

Acknowledgments

We thank the study participants. This work was supported by a grant from the Indian Council of Medical Research (Grant HIV/50/137/2010-ECD-II).

Author Disclosure Statement

No competing financial interests exist.

References

- Neogi U, Sood V, Banerjee S, Ghosh N, Verma S, Samrat S, Sharma Y, Saxena A, Husain S, Ramachandran VG, Das S, Sreedhar KV, Goel N, Wanchu A, and Banerjee AC: Global HIV-1 molecular epidemiology with special reference to genetic analysis of HIV-1 subtypes circulating in North India: Functional and pathogenic implications of genetic variation. *Indian J Exp Biol* 2009;47:424–431.
- Neogi U, Sood V, Chowdhury A, Das S, Ramachandran VG, Sreedhar VK, Wanchu A, Ghosh N, and Banerjee AC: Genetic analysis of HIV-1 circulating recombinant form 02_AG, B and C subtype-specific envelope sequences from Northern India and their predicted co-receptor usage. *AIDS Res Ther* 2009;6:28.
- Hoffman TL and Doms RW: HIV-1 envelope determinants for cell tropism and chemokine receptor use. *Mol Membr Biol* 1999;16:57–65.
- Prakash SS, Chaudhary AK, Lodha R, Kabra SK, Vajpayee M, Hazarika A, Bagga B, and Luthra K: Efficient neutralization of primary isolates by the plasma from HIV-1 infected Indian children. *Viral Immunol* 2011 [Epub ahead of print]; DOI: 10.1089/vim.2011.0028.
- Jensen MA, Coetzer M, Wout AB, Morris L, and Mullins JJ: A reliable phenotype predictor for human immunodeficiency virus type 1 subtype C based on envelope V3 sequences. *J Virol* 2006;80:4698–4704.
- Sing T, Low AJ, Beerenwinkel N, Sander O, Cheung PK, Domingues FS, Buch J, Daumer M, Kaiser R, Lengauer T, and Harrigan PR: Predicting HIV coreceptor usage on the basis of genetic and clinical covariates. *Antivir Ther* 2007;12:1097–1106.
- Tamura K, Dudley J, Nei M, and Kumar S: MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007;24:1596–1599.
- Zhang H, Tully DC, Zhang T, Moriyama H, Thompson J, and Wood C: Molecular determinants of HIV-1 subtype C coreceptor transition from R5 to R5X4. *Virology* 2010;407:68–79.
- Cilliers T, Nhlapo J, Coetzer M, Orlovic D, Ketas T, Olson WC, Moore JP, Trkola A, and Morris L: The CCR5 and CXCR4 coreceptors are both used by human immunodeficiency virus type 1 primary isolates from subtype C. *J Virol* 2003;77:4449–4456.
- Kandathil AJ, Kannangai R, Abraham OC, Pulimood SA, Jensen MA, and Sridharan G: HIV-1 with predicted CXCR4 genotype identified in clade C from India. *Mol Diagn Ther* 2009;13:19–24.
- Cecilia D, Kulkarni SS, Tripathy SP, Gangakhedkar RR, Paranjape RS, and Gadkari DA: Absence of coreceptor switch with disease progression in human immunodeficiency virus infections in India. *Virology* 2000;271:253–258.
- Guglietta S, Pantaleo G, and Graziosi C: Long sequence duplications, repeats, and palindromes in HIV-1 gp120:

- Length variation in V4 as the product of misalignment mechanism. *Virology* 2010;399:167–175.
13. Castro E, Belair M, Rizzardi GP, Bart PA, Pantaleo G, and Graziosi C: Independent evolution of hypervariable regions of HIV-1 gp120: V4 as a swarm of N-Linked glycosylation variants. *AIDS Res hum Retroviruses* 2008;24:106–113.
 14. Wyatt R, Sullivan N, Thali M, Repke H, Ho D, Robinson J, Posner M, and Sodroski J: Functional and immunologic characterization of human immunodeficiency virus type 1 envelope glycoproteins containing deletions of the major variable regions. *J Virol* 1993;67:4557–4565.
 15. Belair M, Dovat M, Foley B, Mayerat C, pantaleo G, and Graziosi C: The polymorphic nature of HIV type 1 env V4 affects the patterns of potential N-glycosylation sites in proviral DNA at the intrahost level. *AIDS Res Hum Retroviruses* 2009;25:199–206.
- Address correspondence to:
Kalpana Luthra
Department of Biochemistry
Room No. 3002
All India Institute of Medical Sciences
New Delhi 110029
India
E-mail: kalpanaluthra@gmail.com