Molecular Characterization of the env Gene of Two CCR5/CXCR4-Independent Human Immunodeficiency 2 Primary Isolates

Quirina Santos-Costa,1 Ricardo Parreira,2 José Moniz-Pereira,1 and José Miguel Azevedo-Pereira1*

1Centro de Patogénese Molecular – Unidade dos Retrovírus e Infecções Associadas (CPM-URIA), Faculdade de Farmácia da Universidade de Lisboa, Lisboa, Portugal
2Unidade de Virologia/Unidade de Parasitologia e Microbiologia Médicas (UPMM), Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisboa, Portugal

Human immunodeficiency virus 2 (HIV-2) infection is characterized by a slower disease progression and lower transmission rates. The molecular features that could be assigned as directly involved in this in vivo phenotype remain essentially unknown, and the importance of HIV-2 as a model to understand pathogenicity of HIV infection has been frequently underestimated. The early events of the HIV replication cycle involve the interaction between viral envelope glycoproteins and cellular receptors: the CD4 molecule and a chemokine receptor, usually CCR5 or CXCR4. Despite the importance of these two chemokine receptors in human immunodeficiency virus 1 (HIV-1) entry into cells, we have previously shown that in some HIV-2 asymptomatic individuals, a viral population exists that is unable to use both CCR5 and CXCR4. The goal of the present study was to investigate whether possible regions in the env gene of these viruses might account for this phenotype. From the molecular characterization of these env genes we could not detect any correlation between V3 loop sequence and viral phenotype. In contrast, it reveals the existence of remarkable differences in the V1/V2 and C5 regions of the surface glycoprotein, including the loss of a putative glycosylation site. Moreover, in the transmembrane glycoprotein some unique sequence signatures could be detected in the central ectodomain and second heptad repeat (HR2). Some of the mutations affect well-conserved residues, and may affect the conformation and/or the dynamics of envelope glycoproteins complex, including the SU–TM association and the modulation of viral entry function.

KEYWORDS: Env glycoproteins; HIV-2; CCR5; CXCR4; amino acid motifs; coreceptor usage

INTRODUCTION

According to the universally accepted model, human immunodeficiency virus (HIV) entry into host cells involves specific interactions between the virion heterotrimeric envelope complex, formed by surface (SU) and transmembrane (TM) Env glycoproteins, and two cellular proteins: CD4 and a coreceptor. Conformational changes in the SU glycoprotein, induced by SU-CD4 binding, cause the exposure, or formation of, a specific chemokine receptor (coreceptor) binding site in SU. After binding of SU to the coreceptor molecule, a second conformational change occurs which exposes the N-terminal fusion peptide in TM, allowing it to then be inserted into host cell membrane [Clapham and McKnight, 2002]. The gp41 trimers then undergo another conformational change where the N-terminal and C-terminal heptad repeat regions come together to form a six-helical coiled-coil structure [Chan et al., 1997; Weissenhorn et al., 1997]. The formation of this helical bundle brings the viral envelope and cellular membrane into proximity, facilitating the fusion of the two membranes and the release of the viral nucleocapsid into the cell cytoplasm.

Nineteen chemokine receptors have been considered, in vitro, as coreceptors for HIV type 1 (HIV-1), type 2 (HIV-2), and simian immunodeficiency virus (SIV) [Broder and Jones-Trower, 1999; Simmons et al.,...]

The authors declare that they do not have any financial or other conflict of interest that might be construed to influence the contents of the manuscript, including the results or interpretation of publication.

Grant sponsor: Fundação para a Ciência e Tecnologia (FCT); Grant number: POCI/SAU-IMI/55726/2004.
*Correspondence to: José Miguel Azevedo-Pereira, CPM-URIA, Faculdade de Farmácia de Lisboa, Av Prof Gama Pinto, 1649-003 Lisboa, Portugal. E-mail: miguel.pereira@ff.ul.pt
Accepted 9 July 2009
DOI 10.1002/jmv.21619
Published online in Wiley InterScience (www.interscience.wiley.com)
However, despite the extensive range of molecules that could potentially act as viral coreceptors, CCR5 and CXCR4 are not only the major coreceptors for HIV-1, but they also seem to be major determinants in the HIV-1 pathogenesis [Zhang et al., 1998; Simmons et al., 2000]. CCR5-dependent (R5) strains are predominant during the early stages of HIV-1 infection and they seem to play a pivotal role for its successful establishment. Indeed, several epidemiologically based studies have shown that individuals who are homozygous for the Δ32 ccr5 mutation are virtually protected from HIV-1, while those heterozygous for this allele display slower disease progression [for a review see Regoes and Bonhoeffer, 2005].

Moreover, only in ∼40% of infected humans does a viral population arise that can use CXCR4 in addition to CCR5 (R5/X4 strains), or instead CCR5 (X4 strains) [Berger et al., 1998, 1999; Simmons et al., 1996]. The appearance of such strains is usually associated with accelerated CD4+ T-lymphocyte loss and disease progression [Connor and Ho, 1994; Richman and Bozzette, 1994; Bjorndal et al., 1997; Connor et al., 1997]. Thus, Env glycoprotein structure determines early viral life cycle events, and directly influences cellular tropism, transmission, and pathogenesis of HIV-1 infections.

On the other hand, HIV-2 infection in vivo is characterized by a less virulent phenotype and, in general, is associated with a slower development of immunodeficiency when compared to HIV-1 [Marlink et al., 1994; Norrgren et al., 2003]. Differences in the natural history of HIV-1 and HIV-2 infection also include a lower plasma viremia and, consequently, significantly lower rates of sexual and vertical transmission. One of the factors that could account for this reduced virulence is the less efficient way with which HIV-2 interacts with cellular receptors [Azevedo-Pereira et al., 2005].

While CCR5 and CXCR4 are the major HIV-1 coreceptors, HIV-2 entry into host cells is frequently characterized by a distinct spectrum of coreceptor usage [Bron et al., 1997; Guillon et al., 1998; McKnight et al., 1998; Azevedo-Pereira et al., 2003a], including those rarely engaged by HIV-1. In addition, some non-adapted HIV-2 strains are able to enter susceptible cells through a CD4-independent way [Reeves et al., 1999; Azevedo-Pereira et al., 2003b] as a consequence of direct interaction with chemokine receptors. Both features are unequivocally associated with the oligomeric structure of envelope glycoproteins [Reeves et al., 1997; Dumonceaux et al., 2001; Azevedo-Pereira et al., 2003b], suggesting that this structure should be more flexible in HIV-2 than in HIV-1, and stressing the potential relevance of HIV-2 env gene sequence studies to further extend the knowledge of the initial interactions between viruses and target cells.

In this report we describe the cloning and sequence analysis of the env gene from two HIV-2 isolates obtained from asymptomatic individuals that do not use CCR5 or CXCR4 coreceptors to infect primary peripheral blood mononuclear cells (PBMCs) [Azevedo-Pereira et al., 2003a]. From this analysis, we found that these two viruses show remarkable differences in primary amino acid sequence, particularly in the V1/V2 and C5 regions of surface glycoprotein. Moreover, some unique sequence signatures were also detected in the central ectodomain and in the second heptad repeat (HR2) of the transmembrane glycoprotein. Some of the mutations affect well-conserved residues, and may influence the conformation and/or the dynamics of the envelope glycoprotein complex, including the SU–TM association and the modulation of viral entry function.

**MATERIALS AND METHODS**

**PCR Amplification and Cloning of the Env Coding Region**

The *env* gene of HIV-2<sub>MIC97</sub> (MIC97) and HIV-2<sub>MJC97</sub> (MJC97) was amplified by polymerase chain reaction (PCR) as previously described [Azevedo-Pereira et al., 2003b] using 250 ng of chromosomal DNA from low-passage infected PBMC. Briefly, PHA-stimulated PBMC were infected with either MIC97 or MJC97 strains and the cultures were maintained for 15 days. Cells were then washed, lysed, and total chromosomal DNA was extracted by a DNA isolation kit (Genta Puregene Cell Kit (8x10<sup>e8</sup>); QIAGEN, Sao Paolo, Brazil) according to manufacturer’s instructions. PCR was performed on this template using the Expand Long Template PCR System (Roche Applied Science, Indianapolis, IN). The primer pair used was QSC3 (5′-AGCTAGTGCAATGCTATATTTG-3′; forward) and QSCBam-R (5′-GGCCGGAACGGT-CTTGGATCCTACTGCCC-3′; reverse), located between coordinates 6,729–6,752 and 9,120–9,148, respectively (nucleotide positions have been numbered according to the HIV-2<sub>ALI</sub> sequence; accession number: AF082339). This primer pair enables the amplification of the *env* gene spanning from the signal peptide region to 17 nt downstream the BamH I restriction site. Amplified fragments were cloned into the pCR<sup>®</sup>-4 cloning vector (Invitrogen, Sao Paolo, Brazil) and used to transform ultracompetent *E. coli* DH5α-T1 cells (Invitrogen). Doubled-stranded plasmid DNA was purified (Endofree Plasmid Maxi kit–Qiagen) and sequenced.

**DNA Sequencing and Sequence Analysis**

Purified DNA was sequenced using the BigDye terminator cycle sequencing kit (Applied Biosystems, Foster City, CA) and an ABI PRISM 3100 Genetic Analyser (Applied Biosystems). Seven clones of each amplified fragment were sequenced, aligned using CLUSTALX software. After manual edition and correction, a complete consensus sequence from MIC97 and MJC97 env genes was obtained using BioEdit program.

DNA and putative amino acid sequences obtained were aligned with HIV-2 and SIV reference sequences obtained from the Los Alamos Sequence database (available online at http://www.hiv.lanl.gov/). Multiple
Sequence alignments were made using CLUSTAL W [Thompson et al., 1994] and the phylogenetic trees generated with the neighbor-joining method using Mega 4.0 software [Tamura et al., 2007] and genetic distance matrixes corrected with the Kimura 2-P formula, using a transition/transversion ratio of 2 and assuming a homogeneous distribution of nucleotide substitution rates across the sequences analyzed. Their robustness was assessed by performing 1,000 bootstrap replicates of the original sequence data. Edited sequences were submitted to GenBank (accession numbers, AY168925 and EU021092).

Recombinant analysis was carried out using Simplot 3.5.2 software [Lole et al., 1999] and the definition of putative N-linked glycosilation sites was made using N-Glycosite (available at http://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html) or NetNGlyc 1.0 (available at http://www.cbs.dtu.dk/services/NetNGlyc/). Transmembrane domains were defined through Sosui 1.11 software [Hirokawa et al., 1998] available at http://bp.nuap.nagoya-u.ac.jp/sosui/. The net charge of variable regions 1, 2, and 3 (V1, V2, and V3, respectively) was calculated using ProtMW2 (available at http://www.mbi.ufl.edu/~shaw/ProtMW2.htm). Putative protein secondary structure predictions were carried out using the PSIPRED software (http://bioinf.cs.ucl.ac.uk/psipred/) and further assessed with HNN and SOPMA secondary structure prediction methods (accessible online at http://pbil.ibcp.fr/htm/index.php).

The coreceptor usage phenotype of each viruses included in the sequence analysis is summarized in Table I.

**RESULTS**

**Phylogenetic Analyses**

The envelope glycoproteins coding regions of the CCR5/CXCR4-independent HIV-2 isolates, MIC97, and MJC97, were cloned into the pCR³ 4 vector. Seven of each env clones were sequenced and aligned revealing a high degree of similarity between clones. Due to this similarity, a consensus sequence was constructed for MIC97 and MJC97, and their nucleotide and derived amino acid sequence aligned with other HIV-2 and SIV references from Los Alamos HIV sequence database.

The phylogenetic relationship of MIC97 and MJC97 with HIV-2 and SIV strains revealed that they belong to the phylogenetic group A, although they clustered distinctly from the majority of the sequences analyzed (data not shown). Nonetheless, recombination analysis, assessed by Simplot 3.5.2 software showed that none of the strains were intergroup recombinants (data not shown).

**Sequence Analysis of the V3 Region**

To understand the molecular basis of the unusual coreceptor usage profile of these isolates [Azevedo-Pereira et al., 2003a], we first examined the molecular features of their putative V3 region, as the V3 loop significantly influences CCR5 and CXCR4 coreceptor usage of HIV-1 isolates [Choe et al., 1996; Speck et al., 1997; Hartley et al., 2005].

The V3 loop of MIC97 and MJC97 contains 34 amino acid residues and a net charge of +6. An alignment of the V3 region of envelope proteins revealed that these two viruses do not show any major divergence when compared to other HIV-2 isolates (Fig. 1), and both extremities and tip of the V3 loop were particularly well preserved. From this alignment it was also evident that HIV-2 sequences corresponding to the V3 loop showed limited variability, in contrast with HIV-1. This observation emphasizes the notion that this region, in HIV-2, is probably less exposed, and under minor selective pressure than its equivalent in HIV-1 [Damond et al., 2001].

However, it is important to note the existence of two divergent sequence motifs outside the V3 loop region. One is located upstream the V3 loop sequence, just before a putative glycosilation site: the sequence motif KYY, present in HIV-2ROD (ROD) and many other strains, was substituted by T272F273N274 and T272Y273N274 in MIC97 and MJC97, respectively, implicating the replacement of the Lys-272 (present in the large majority of strains) by the threonine (Fig. 1). In this particular motif it is also noticeable the presence of an asparagine (Asp-274) replacing the highly conserved tyrosine. Although both are non-charged amino acids, asparagine is more hydrophilic than tyrosine and is frequently involved in inter- and intra-chain interactions with other amino acid residues through hydrogen bonds.

The substitution of a positively charged residue is also observed just before the N-terminal cysteine of the V3 loop where a histidine (His-279) was replaced by a tyrosine (Fig. 1). It should be noted that all but one (HIV-2ISY) of the analyzed sequences, show a positively charged amino acid at this position, and none has simultaneously lost both positively charged residues in the sequence motif K272XXNLXXH279, which was only observed in MIC97 and MJC97. The other noticeable difference is located in a region downstream of the V3 loop, where the motif G217DWQG221 is absent from all the other sequences analyzed.

The study of putative N-linked glycosilation sites in the V3 region revealed a similar glycosilation pattern for both viruses, similar to that of other HIV-2 isolates, regardless of coreceptor usage. This suggests that in HIV-2, as in HIV-1, the role of N-linked glycan in the V3 loop-coreceptor binding is uncertain [Nakayama et al., 1998; Li et al., 2001] and dependent on the overall amino acid composition of a given Env glycoprotein.

**Analyses of Env Regions Outside the V3 Loop**

Since the V3 loop sequence could not be associated with any particular molecular signature evocative of the unique coreceptor usage of MIC97 and MJC97, we attempted to determine if regions outside V3 loop might possess distinct molecular characteristics that
could account for CCR5 or CXCR4 independence. Remarkably, the V1 and V2 regions, including those of the MIC/MJC showed high degree of divergence, particularly evident in the highly variable V1 loop. These distinctive features include several non-conservative amino acid substitutions and loss of putative N-linked glycosylation sites (Fig. 2). Noteworthy, there are two non-conservative amino acid substitutions, located at the base of the V2 loop (amino acids 132 and 133 in Fig. 2), that eliminate one or two charged residues (depending on the strain) that are present in all the viruses used in our comparisons, irrespective of R5, X4, or R5/X4 phenotype. However, none of these resulted in a predicted change in this region’s secondary structure as defined by PSIPRED (data not shown) in comparison with HIV-2ROD (R5/X4) or HIV-2ALI (R5).

Sequence Analysis of Transmembrane Glycoprotein Coding Region

Since MIC97 and MJC97 showed a marked impairment in replication capacity [Azevedo-Pereira et al., 2003a] we decided to analyze the potential role played by the TM subunit of Env. This rationale stems from our

<table>
<thead>
<tr>
<th>Viruses</th>
<th>GenBank accession number</th>
<th>Phenotype</th>
<th>Refs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MJC97</td>
<td>EU021092</td>
<td>R5 – X4</td>
<td>Azevedo-Pereira et al. [2003a]</td>
</tr>
<tr>
<td>MIC97</td>
<td>AY166925</td>
<td>R5 – X4</td>
<td>Azevedo-Pereira et al. [2003a]</td>
</tr>
<tr>
<td>A.CAM6</td>
<td>U05355</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>A.MDS</td>
<td>Z48731</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>A.CAM1</td>
<td>U05359</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>A.CAM3</td>
<td>U05355</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>A.CAM2CG</td>
<td>D00835</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>A.FG</td>
<td>J03654</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>A.ROD</td>
<td>M15390</td>
<td>R5X4</td>
<td>Broder et al. [1999]</td>
</tr>
<tr>
<td>A.CBL23</td>
<td>U05352</td>
<td>24</td>
<td>Broder et al. [1999]</td>
</tr>
<tr>
<td>A.CBL24</td>
<td>U05353</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>A.ALI</td>
<td>AF082339</td>
<td>R5</td>
<td>Reeves et al. [1999]</td>
</tr>
<tr>
<td>A.PEI2</td>
<td>U22047</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>A.CAM5</td>
<td>U05357</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>A.CBL22</td>
<td>U05351</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>A.ISY</td>
<td>J04498</td>
<td>R5X4</td>
<td>Broder et al. [1999]</td>
</tr>
<tr>
<td>A.ST</td>
<td>M31113</td>
<td>R5</td>
<td>Broder et al. [1999]</td>
</tr>
<tr>
<td>A.CBL21</td>
<td>U05350</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>A.CAM4</td>
<td>U05356</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>A.GH1</td>
<td>M30895</td>
<td>R5</td>
<td>Isaka et al. [1999]</td>
</tr>
<tr>
<td>A.BEN</td>
<td>M30502</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>A.UC2</td>
<td>J03693</td>
<td>R5X4</td>
<td>Broder et al. [1999]</td>
</tr>
<tr>
<td>A.D194</td>
<td>J04542</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>AB.7312A</td>
<td>L36874</td>
<td>R5X4</td>
<td>Broder et al. [1999]</td>
</tr>
<tr>
<td>B.EHO</td>
<td>U27200</td>
<td>R5X4</td>
<td>Hatse et al. [2007]</td>
</tr>
<tr>
<td>B.KR020</td>
<td>AB100245</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>B.UC1</td>
<td>L07625</td>
<td>R5</td>
<td>Broder et al. [1999]</td>
</tr>
<tr>
<td>B.96200</td>
<td>AF170057</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>B.D205</td>
<td>X61240</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>SIVsm62B</td>
<td>U04983</td>
<td>R5</td>
<td>Broder et al. [1999]</td>
</tr>
<tr>
<td>SIVsm62D</td>
<td>U04987</td>
<td>55</td>
<td>Broder et al. [1999]</td>
</tr>
<tr>
<td>SIVsmm239</td>
<td>M33262</td>
<td>R5</td>
<td>Broder et al. [1999]</td>
</tr>
<tr>
<td>A.P1-1986</td>
<td>DQ213026</td>
<td>58</td>
<td>Shi et al. [2005]</td>
</tr>
<tr>
<td>A.P1-1991</td>
<td>DQ213027</td>
<td>R5</td>
<td>Shi et al. [2005]</td>
</tr>
<tr>
<td>A.P1-1997</td>
<td>DQ213028</td>
<td>R5X4</td>
<td>Shi et al. [2005]</td>
</tr>
<tr>
<td>A.P2-1994</td>
<td>DQ213029</td>
<td>R5</td>
<td>Shi et al. [2005]</td>
</tr>
<tr>
<td>A.P2-1996</td>
<td>DQ213030</td>
<td>R5</td>
<td>Shi et al. [2005]</td>
</tr>
<tr>
<td>A.P2-1998</td>
<td>DQ213031</td>
<td>R5</td>
<td>Shi et al. [2005]</td>
</tr>
<tr>
<td>A.P2-2002</td>
<td>DQ213032</td>
<td>R5</td>
<td>Shi et al. [2005]</td>
</tr>
<tr>
<td>A.P4-1992</td>
<td>DQ213033</td>
<td>R5</td>
<td>Shi et al. [2005]</td>
</tr>
<tr>
<td>A.P4-2000</td>
<td>DQ213034</td>
<td>X4</td>
<td>Shi et al. [2005]</td>
</tr>
<tr>
<td>A.P4-2001</td>
<td>DQ213035</td>
<td>R5</td>
<td>Shi et al. [2005]</td>
</tr>
<tr>
<td>A.P4-2002</td>
<td>DQ213036</td>
<td>X4</td>
<td>Shi et al. [2005]</td>
</tr>
<tr>
<td>A.P8-1998</td>
<td>DQ213037</td>
<td>X4</td>
<td>Shi et al. [2005]</td>
</tr>
<tr>
<td>A.P8-2001a</td>
<td>DQ213038</td>
<td>X4</td>
<td>Shi et al. [2005]</td>
</tr>
<tr>
<td>A.P8-2001b</td>
<td>DQ213039</td>
<td>X4</td>
<td>Shi et al. [2005]</td>
</tr>
<tr>
<td>A.P8-2002</td>
<td>DQ213040</td>
<td>X4</td>
<td>Shi et al. [2005]</td>
</tr>
</tbody>
</table>

R5 – X4 –, unable to use CCR5 and CXCR4 coreceptors; R5X4, usage of both CCR5 and CXCR4; X4, usage of CXCR4; R5, usage of CCR5; ND, not determined.
recent report [Santos-Costa et al., 2009] showing a direct involvement of TM in the replication rate of chimeric viruses derived from ROD molecular clone pROD10 [Peden, 1992]. Indeed, we showed that the high replication rate of ROD is significantly reduced if isogenic recombinant viruses are fitted with a TM glycoprotein from a less-replication competent virus (MIC97 or MJC97). This suggests that viral replication events mediated by TM also contribute to the less efficient replication of ROD-derived chimeric viruses. Accordingly, we decided to investigate if the TM coding region of the MIC97 or MJC97 Env might possess any particular molecular characteristics that could be associated with the lower replication capacity of both viruses. As shown in Figure 3, the subdomains involved in major TM functions (fusion peptide, the two heptads
repeat regions, HR1 and HR2, and the transmembrane helix) are preserved with minor changes in both MIC97 and MJC97. Nonetheless, some peculiar changes deserve specific emphasis. Between HR1 and HR2 domains, two unique substitutions (P583T and V585K) were observed on both MIC97 and MJC97, altering the amino acid sequence PWV to TWK, and noteworthy, introducing a positively charged amino acid (lysine) in a position where all the sequences analyzed show either a neutral (valine, alanine, threonine, or proline) or, exceptionally (in PEI2 sequence), an acidic residue. These substitutions lie adjacent to the disulfide-bonded loop region of the TM ectodomain.

Since in HIV-1 this region and particularly the motif PWV, is involved in the TM–SU association through the interaction with C1 and C5 regions of SU glycoprotein [Wyatt et al., 1997; Binley et al., 2000] we analyzed these two conserved regions of MIC97 and MJC97. Our data revealed the presence of a non-conservative amino acid substitution in the C5 region (Fig. 4), where a glycine (Gly-468) was replaced by a negatively charged amino acid (aspartic acid).

In a particularly well-conserved region of the HR2 heptad repeat, an M627T substitution was observed in both MIC97 and MJC97 (Fig. 3). The replacement of methionine by a threonine is also observed in the sequence of the HIV-2ALI strain that was reported as being a slow-replicating isolate in primary PBMC [Azevedo-Pereira et al., 2003a; Santos-Costa et al., 2009]. However, this phenotype is compensated if the target cell display cellular receptors at higher densities, such as the case of the GHOST-CD4 CCR5 cell line [Azevedo-Pereira et al., 2003b]. At the end of the HR2 region, the sequence motif GNWF, which is found in the majority of group A HIV-2 sequences, was replaced by SNWL. In between the HR2 and the transmembrane regions, an insertion of a serine at position 647 creates a SS doublet (also observed at positions 750 and 751 replacing the motif RD). This insertion is expected to slightly alter the predictive secondary structure of TM.

---

**Fig. 2.** Characterization of the V1–V2 region from MIC97 and MJC97 envelope glycoproteins. Amino acid numbering starts at the first amino acid after the signal peptide of MIC97 and MJC97 Envs. V1 and V2 loops are delimited by the two boxes. Shadowed amino acids represent the putative N-linked glycosylation sites.
Fig. 3. Characterization of the transmembrane (TM) region from MIC97 and MJC97 envelope glycoproteins. Amino acid numbering starts at the first amino acid after the signal peptide of MIC97 and MJC97 Envs. The dotted box delimits the fusion peptide (FP) while the heptad repeat region 1 (HR1) and 2 (HR2) are defined by the two solid line boxes. Underlined amino acids refer to the transmembrane domain (TMD). The arrowed denote the insertion of a serine in the position 647 (downstream the HR2). Asterisks indicate the GYRPV motif (amino acids 682–686) that has been associated with the Vpu-like activity of HIV-2 TM glycoprotein. The R/K750→S and D/N/T751→S substitutions are indicated by a red circle while the HGG sequence motif is denoted by a red box. Shadowed amino acids represent the putative N-linked glycosilation sites.
Fig. 4. Characterization of the C5 region from MIC97 and MJC97 envelope glycoproteins. Amino acid numbering starts at the first amino acid after the signal peptide of MIC97 and MJC97 Envs. Underlined amino acids represent the G/F substitution.
glycoprotein in this particular region, as assessed by PSIPRED, interrupting the typical alpha-helix structure and introducing a small β-sheet stretch (data not shown). Finally, since the TM glycoprotein of HIV-2 has been linked to a Vpu-like activity [Bour et al., 1996; Ritter et al., 1996; Iida et al., 1999; Abada et al., 2005], required to ensure high-level production of HIV-2 progeny by enhancing viral particle release [Barnett et al., 1996; Bour et al., 1996; Ritter et al., 1996], we objectively analyzed the TM domains that have been attributed to this Vpu-like function: the membrane-anchored subunit, the extracellular region and the cytoplasmic domain (CD) [Bour et al., 1996, 1999; Ritter et al., 1996; Abada et al., 2005; Noble et al., 2006]. In none of these regions could we detect major differences that might account for any relevant genomic signatures potentially linked to the lower replication rate phenotype of MIC97 and MJC97. These include the absolute conservation of the GYRPV motif (Fig. 3, amino acids 682–686) that has been associated with the Vpu-like activity of HIV-2 TM glycoprotein [Abada et al., 2005]. Nevertheless, in the CD, it is important to note the substitution of the RD sequence motif detected in the majority of HIV-2 sequences (residues 750 and 751) by two serines, implicating the lost of two charged amino acids. Noteworthy is also the substitution of the highly conserved sequence YGC at the end of the TM cytoplasmatic domain by HGG (amino acid 780–782, Fig. 3).

**DISCUSSION**

We previously reported the identification of two HIV-2 primary isolates (HIV-2MIC97 and HIV-2MJC97) unable to productively infect both GHOST and U87 cell lines coexpressing CD4 and different coreceptors [Azevedo-Pereira et al., 2003a]. More recently, we constructed a set of chimeric viruses and investigated the contribution of Env glycoproteins to the unique coreceptor usage, and replication efficiency, of both strains. From this analysis we found that the segment, delimited by the conserved regions C1 and C4, was determinant for CCR5/CXCR4-independent infection of target cells, while the TM subunit was directly implicated in the lower replication rate of both viruses [Santos-Costa et al., 2009]. In the present study we extended that work, and decided to investigate the viral molecular determinants of this phenotype, focusing on the amino acid sequences of the Env glycoproteins. Our aim was the identification of specific sequence signatures that could eventually be assigned to the unique in vitro phenotype presented by these two strains.

Since HIV-1 coreceptor usage is mainly determined by the V3 region, our first goal was to examine this region in both sequences, and compare them with those of other HIV-2 strains. The V3 region of both MIC97 and MJC97 presented a net charge of +6. In HIV-1, a net charge of +5 or less is indicative of CCR5 usage, while above +5 it usually indicates that CXCR4, or both CCR5 and CXCR4, can be used as coreceptors [Briggs et al., 2000; Delobel et al., 2007]. In HIV-2, the C-terminal half of the V3 loop region (corresponding to the sequence between aa-18 and aa-36) has been directly related with coreceptor usage [Isaka et al., 1999]. Those few reports addressing the involvement of the V3 region in HIV-2 coreceptor usage indicate that a net charge below, or equal to, +6 is associated with CCR5 usage, whereas isolates with higher V3 charge (above +7) preferred CXCR4 as their main coreceptor [Isaka et al., 1999; Shi et al., 2005]. CXCR4 usage is also related with positively charged amino acids (arginine or lysine) at position 19 (related to the V3 loop sequence), while R5 viruses display neutral amino acids (valine or isoleucine) at this position. Moreover, absence of positively charged residues at positions 11 and 25 [De Jong et al., 1992; Hoffman et al., 2002] were also predictive of an R5-phenotype for both viral strain. Therefore, from the analyses of V3 sequences of MIC97 and MJC97, and applying the rules indicated above, we should assume, although erroneously [Azevedo-Pereira et al., 2003a; Santos-Costa et al., 2009], that both viruses would efficiently use the CCR5 coreceptor. This discrepancy between phenotypic data and amino acid signatures restricted to the V3 loop underscores the notion that, besides V3, additional regions of the Env glycoproteins are directly/indirectly involved in cellular receptors usage by HIV [Wyatt et al., 1995; Ross and Cullen, 1998; Labrosse et al., 2001; Saunders et al., 2005; Pastore et al., 2006; Huang et al., 2008]. Furthermore, and substantiating this model, the analysis of a HIV-1 Env chimera containing the V3 loop of an HIV-1 hybrid virus indicated that the V3 loop is the major determinant of CXCR4 usage, whereas the V1–V2 region are more important for binding to coreceptors other than CXCR4 [Hoffman et al., 1998]. In HIV-2 a correlation between V3 loop sequence and viral phenotype is even more difficult to establish. Contradictory results on the impact of V3 sequences in coreceptor usage by HIV-2 has been reported [Kulkarni et al., 2005; Shi et al., 2005; Lin et al., 2007], supporting a model in which cooperative subunit interactions, both within the Env glycoproteins, and between the latter and cellular receptors are in HIV-2, more than in HIV-1, both cell type and virus strain dependent.

In contrast to the V3 region, the V1 and V2 regions, including those of the MIC/MJC, showed high degree of divergence, which was particularly evident in the highly variable V1 loop, and included the loss of a putative glycosilation site. In HIV-1, genetic analysis by site-directed mutagenesis revealed that carbohydrate moieties in the V1, V2, and V3 regions directly influence the capacity of Env glycoproteins to mediate membrane fusion. This probably results from a differential structure in the CD4 or coreceptor binding sites present in those mutants, which may prevent the efficient engagement of cellular receptors [Ogert et al., 2001]. Remarkably, in this latter report, one of the N-linked glycosilation mutants, lacking a glycosilation site in a region located between V1 and V2, had completely lost...
the ability to use both CCR5 and CXCR4 coreceptors in a membrane fusion assay. We lack experimental evidence on the influence of putative glycosylation site loss on MIC97 and MJC97 coreceptors usage, but the similarities are striking, therefore, tempting us to suggest that a possible correlation might exist between their glycosylation profile and CCR5/CXCR4-independence.

From the V1/V2 sequence analysis we also found two non-conservative amino acid substitutions, located at the base of V2 loop, that eliminate one or two charged residues that are present in all other viruses, regardless of their R5, X4, or R5/X4 phenotype. The few data available indicate that this region is highly variable, and apparently hidden (either by glycosylation or sequestration) in the tertiary structure of the oligomeric envelope complex, as assessed by epitope mapping of the ROD SU glycoprotein [McKnight et al., 1996]. The importance of this region in envelope glycoprotein interaction with cellular coreceptors is unknown. However, the density of positively charged residues in this particular region, regardless the coreceptor usage phenotype of the strain, is remarkable. Moreover, it was suggested that amino acid substitutions at the base of V2 loop might be critical, resulting in local distortions that could influence V2 conformation, and the exposure of the V2 tip [McKnight et al., 1996]. This tip is particularly well conserved in HIV-2 sequences which may indicate that, similarly to HIV-1 [Moore et al., 1994], it could be functionally important.

From previous results regarding the characterization of MIC97 and MJC97, we detected a marked impairment in the replication capacity in both viruses [Azevedo-Pereira et al., 2003a], and identified a direct involvement of the TM subunit in this phenotype [Santos-Costa et al., 2009]. However, the results presented here show that the subdomains involved in major TM functions (namely the fusion peptide, the two heptads repeat regions and the transmembrane region) are preserved in MIC97 and MJC97. Nevertheless, some minor changes deserve a deeper discussion. In the central region of the ectodomain, two unique substitutions downstream of the disulfide-bonded loop alter the consensual sequence motif PWV to TWK, and introduce a positively charged amino acid (lysine) in a position where all other sequences show either a neutral (valine, alanine, threonine, or proline) or, exceptionally, an acidic residue. This central region, highly conserved among HIV isolates, is thought to act as a hinge during the formation of the hairpin structure needed for fusion activity, and in the interaction between the TM and SU subunits within the native envelope glycoprotein complex [Helseth et al., 1991; Binley et al., 2000; York and Nunnberg, 2004]. Of particular note is the fact that in HIV-1 the proline residue of the PWV motif seems to be involved in the TM–SU association, as revealed by cysteine scanning mutagenesis of the C1 and C5 regions [Binley et al., 2000]. The assessment of which HIV-2 Env residues may be involved in the TM–SU association has not yet been carried out. Nonetheless, it is reasonable to infer that non-conservative amino acid substitutions either within, and/or proximal to, the disulfide-bonded loop region, such as those detected in our analysis of the MIC97 and MJC97 sequences, are likely to affect the conformation and/or the dynamics of the envelope glycoproteins complex, including the SU–TM association and in modulating its function in viral entry, as demonstrated in HIV-1 [Helseth et al., 1991; Poubia et al., 2003].

A distinguishable feature of lentivirus envelope glycoproteins is the unusual length of their CD sequences. With the exception of feline immuno-deficiency virus (FIV), the CDs of all lentiviruses are over 120 amino acids in length, while those of other types of retroviruses typically range between 20 and 40 amino acids. This suggests that CD may fulfill some important and unique function in lentiviral life cycle. Accordingly, several cellular proteins have been referred to bind to the CD of HIV-1 TM glycoprotein (Srinivas et al., 1993; Berloz-Torrent et al., 1999; Zhang et al., 1999; Evans et al., 2002; Blot et al., 2003, 2006), although their role in the viral life cycle remains to be fully elucidated. In HIV-2, the TM glycoprotein has been associated with a Vpu-like activity enhancing the budding and release of viral particles [Bour et al., 1996; Ritter et al., 1996; Iida et al., 1999; Iida et al., 1999; Abada et al., 2005]. Although we did not detect any major alteration in the TM regions associated with Vpu-like activity, namely in the membrane-anchored subunit or in the extracellular domain, we did find two unique substitutions in the CD of MIC97 and MJC97 TM glycoproteins. In both cases these substitutions lead to alterations in physical properties of the involved sequence motifs (loss or introduction of charged residues) that may alter the way the CD interact with cellular and viral proteins.

The analysis of C1 and C5 conserved regions of MIC97 and MJC97 unveil the presence of a non-conservative amino acid substitution in the C5 region where a glycine was replaced by aspartic acid. Whether this mutation modifies the way TM and SU interact is at this moment merely speculative. Even so, and in accordance to the results obtained for HIV-1 [Helseth et al., 1991; Maerz et al., 2001], the highly conserved residues Leu-567, Trp-570, and Gly-571, that precede the TM disulfide-bonded loop, seem to interact in HIV-1 with conserved, and predominantly hydrophobic, residues in the C1 or in C5 regions such as Ile-464 and Gly-468 observed in HIV-2 sequences. The replacement of Gly-468 by aspartic acid detected in both MIC97 and MJC97 sequences will probably influence the SU–TM stability. Thus, it may account for some of the phenotypic characteristics of MIC97 and MJC97, since this region is crucial for SU–TM contact, and may link the SU conformational changes, induced by CD4 and chemokine receptor binding, to TM fusion activation.

From inspection of the MIC97 and MJC97 TM sequences, we also detected an insertion of a serine residue (S647) in between HR2 and the transmembrane region, creating a SS doublet. The possible implications of this alteration are difficult to foresee. Yet, this region

is directly involved in the formation of the helical-hairpin complex needed to bring the two lipid bilayers into close proximity during the fusion step. According to this idea, the conformational stability of the fusogenic TM core is likely to be critical in determining the membrane fusion properties of the envelope glycoprotein [He et al., 2008]. This stability is directly dependent on hydrophobic interactions involving packing of non-polar side chains present in this region [Shu et al., 2000; Dong et al., 2001; Follis et al., 2002; Wang et al., 2002]. The introduction of a polar hydrophilic amino acid residue and the predicted alteration of secondary structure of this domain are likely to decrease the stability of the TM core, and thus the fusogenic activity of the envelope complex.

In conclusion, the data generated in this study have shown that HIV-2MIC97 and HIV-2MC97 amino acid sequences have differences that could be determinants of the unique phenotype presented by these two isolates. However, the envelope glycoprotein activity, during HIV replication, is determined by its secondary and tertiary structure and results from diverse, and complex, interactions between several regions within the native envelope glycoprotein complex. Without in vitro direct mutagenesis analysis it remains to be determined if the changes here identified in the putative primary structure of the Env proteins have any impact on their conformational and functional properties.

REFERENCES


Ross TM, Cullen BR. 1998. The ability of HIV type 1 to use CCR-5 as a coreceptor is controlled by envelope V1/V2 sequences acting in conjunction with a CCR-5-tropic V3 loop. Proc Natl Acad Sci USA 95:7802–7806.


virus type 1 isolates are dual-tropic and most can use either Lestr or CCR5 as coreceptors for virus entry. J Virol 70:8355–8360.


