Rapid progression to human immunodeficiency virus infection / acquired immunodeficiency syndrome (HIV/AIDS) correlates with variation in viral ‘tat’ sequences

Mary Bridget Nanteza², David Yirrell¹,⁵, Benon Biryahwaho², Natasha Larke³, Emily Webb³, Frances Gotch⁴ and Pontiano Kaleebu¹,²

¹MRC/UVRI Uganda Research Unit on AIDS, c/o Uganda Virus Research Institute, Plot 51-59 Nakiwogo Road, P. O. Box 49, Entebbe, Uganda.
²Uganda Virus Research Institute, Plot 51-59 Nakiwogo Road, P. O. Box 49, Entebbe, Uganda.
³MRC Tropical Epidemiology group, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom.
⁴Department of Immunology, Imperial College, Chelsea and Westminster Hospital 369 Fulham Rd, London SW10 9NH, United Kingdom.
⁵Department of Medical Microbiology, Ninewells Hospital, Dundee, DD1 9SY United Kingdom.

Accepted 3 August, 2013

INTRODUCTION

Disease progression among people living with human immunodeficiency virus infection/acquired immunodeficiency syndrome (HIV/AIDS) seems to be strongly associated with multiplicity of factors such as host HLA alleles and HIV-1 subtype. Genetic variation in individual viral sequence such as an insertion of two amino acids in the C-terminus of exon 2 of tat (Tzitzivacos et al., 2009) and deletions in nef were in some studies shown to correlate with slow disease progression. In gp41, low levels of antibodies to the epitope 'ELDKWA' have been associated with advanced HIV-1 disease (Srisurapanon et al., 2005). The tat protein trans-activates transcription by attaching to the trans-activating responsive (TAR) element of the 5' long terminal repeat (LTR). Studies have been performed to delineate the functional mechanisms of tat (Kuppuswamy et al., 1989, LeGuern et al., 1993). Peloponese et al.

Key words: HIV-1, progression, tat, nef, gp41.

Abbreviations: RP, Rapid progressors, LTS, long term survivors.
(1999) showed that HIV strains circulating in Africa were more virulent than strains in Europe and America and was attributed to variations in the tat protein. Furthermore, Niyasom et al. (2009) also showed that subtype B tat activity was associated with reduced disease progression. Humoral and cytotoxic T-cell responses to tat have also shown inverse correlation with slow and non-progressive HIV-1 disease (van Baalen et al., 1997, Zagury et al., 1998, Gupta and Mitra, 2007). However this association was not replicated in a study conducted among Ugandans (Senkaali et al., 2008). The trans-activating and immuno-responsive functions of tat have been attributed to the cysteine-rich region (region III) and the basic region (region IV) of tat exon 1. Tat has also been shown to enhance HIV-1 replication.

With regard to nef, several studies have demonstrated that nef deleted mutants of HIV-1 and simian immunodeficiency virus (SIV) were associated with diminished viral virulence and attenuated infection however some other work did not confirm this observation (Hofmann-Lehmann et al., 2003, Chakrabarti et al., 2003). The nef gene has been shown to exert its effect through acceleration of HIV-1 activation from latency and enhancement of viral replication. There are indications that nef is implicated in the downregulation of CD4 and MHC class I molecules (Jin et al., 2008), and Geriach et al., 2010) thus disabling the humoral and cytotoxic responses.

ApoBec3 cytidine deaminases are antiviral proteins that inhibit the replication of HIV-1. The ‘YXXL’ motif in the membrane-proximal cytoplasmic gp41 has been identified to mediate the binding to the human Apobec3 (Pery et al., 2009). It has been hypothesised that sequence variations in the ‘YXXL’ motif could interfere in the binding of Apobec3 and result in up regulation of viral replication. Such events could subsequently result in rapid disease progression. We report on DNA sequences of tat, nef and membrane-proximal gp41 and attempts to correlate the variation observed to HIV-1 disease progression.

METHODOLOGY

Study subjects

This was a retrospective cross sectional study conducted among HIV-infected adults from a natural history population-based cohort maintained by the MRC/UVRI Uganda Research Unit on AIDS in Uganda (Morgan et al., 1997). Rapid progressors were individuals who progressed to a CD4 count of <200 cells/µl (median 173) in a median time of 3.7 (range 1.3 to 4.9) years. This median time was calculated from mid-way between the last HIV sero-negative result and the index HIV sero-positive result, to the time of obtaining the study blood sample. Long-term-survivors were individuals who had a CD4 count of >500 cells/µl (median 689) in a median time of 8.8 (range 7.5 to 9.3) years. The median time was measured from the time of the index HIV sero-positive result to the time of obtaining the study blood sample. 64% of the long-term survivors were prevalent cases with no prior documentation of a negative result. The remainder were incident cases where the true length of infection could be documented (data not shown). The blood samples were obtained before anti-retroviral therapy (ART) was widely implemented in Uganda and participants were selected on the basis of having no previous exposure to anti-retroviral drugs. The Uganda Virus Research Institute Scientific and Ethical Committee approved the study.

CD4/CD8 count estimation

CD4/CD8 lymphocytes were quantified from 50 µl of fresh ethylene diaminetetraacetic acid (EDTA) blood using flowcytometry on a fluorescence activated cell sorting (FACS) count according to the manufacturer’s instructions (Becton Dickinson International, Belgium).

DNA extraction

DNA was extracted from 300 µl frozen uncultured PBMC using the Puregene kit (Gentra Systems Inc., North Carolina, USA) according to the manufacturer’s protocol.

DNA amplification

We designed tat and gp41 primers and the nef primers were adapted from (Jubier- Maurin et al., 1999) to suit HIV-1 subtype A and D that were dominantly present in Uganda. All the tat, gp41 and nef primers were synthesized by Oswe1 DNA, Southampton, UK. For the first round PCR, 5 µl of the DNA extract (~1.0 ×g DNA) was added to a 15 µl reaction containing x1 PCR buffer; 200 ×M of dCTP, dATP, dTTP and dGTP (Sigma, USA); 0.2 pmoles outer primer pairs (tat-1 and tat-2) or (nef-1 and nef-2) or (env-7 and env-8); 1.5 mM MgCl2 for (tat-1 and tat-2) and (nef-1 and nef-2); and 1.4 mM MgCl2 for (env-7 and env-8). Finally, 0.05 U of Taq DNA polymerase was added. DNA samples were cycled: (i) 94°C (1 min), 55°C (1 min) and 72°C (1 min); for three cycles; (ii) 94°C (30 s), 55°C (45 s), 72°C (1 min); for 30 cycles; and (iii) 72°C (5 min). Two microlitres of the first round PCR product was transferred to an 18 µl reaction mixture containing x1 PCR buffer; 200 ×M of dRhodamine dNTPs and 200 ×M of the inner primers (tat-3 and tat-4) or (nef-3 and nef-4) or (env-5 and env-6) or (env-5 and env-4); and 0.05U of Taq DNA polymerase. Amplification was performed using the cycling conditions stated above. The details of the primers used are shown in Table 1.

Sequencing

The template for sequencing was generated from a 120 ×l secondary PCR reaction containing 3 ×l of the corresponding primary PCR product. The generated product was cleaned using the QIAquick PCR Purification kit (QIAGEN, UK). A sequencing PCR reaction was carried out in a volume of 10 ×l which consisted of 1 ×l of 3.2 pmole/×l single secondary primer; 4 ×l of dRhodamine deoxy terminator mix (Applied Biosystems, Warrington, UK); and 5 ×l of cleaned PCR product. The mixture was subjected to thermal cycling...
RESULTS

We found major and minor sequence variations affecting the three target HIV-1 genes.

The ‘long’ tat was only found in RPs;

Tat sequence variations were observed in tat exons 1 and 2. A major variation was found in exon 2; the ‘long’ tat (115+ amino acids) was only found in RPs (seven out of 18 RPs, 36.8% compared to 0 out of 14 LTSs, 0%) whereas both leucine (79%) and isoleucine (I) (21%) (crude p = 0.002, adjusted p=0.008 chi square test (Figure 1)). Furthermore, there was some suggestion that positions 75, 79, and 86 of tat exon 2 seem to be associated with rapid disease progression (crude p = 0.032, adjusted p = 0.301) (Figure 2A). Position 82 also seems to be associated with rapid disease progression (crude p = 0.036, adjusted p=0.356). Whilst these adjusted p values do not provide strong evidence (p > 0.05), this was among a small sample size. Sequence variations were observed among “long” tat subtype A sequences as 28% sequences had proline (P) at position 75, glutamine (Q) at position 79 and glutamic acid (E) at position 86. Position 79 constitutes part of the cellular adhesion region of tat and variations at this position could play a role in HIV pathogenesis. In tat exon 1, lysine (K) at position 63 appears to be associated with rapid disease progression (crude p = 0.011, adjusted p = 0.073) and was also associated with the “long” tat in subtype A sequences. Position 8 in tat exon 1 was conserved in LTSs; leucine (L) was exclusively found in viruses isolated from LTSs whereas both leucine (79%) and isoleucine (I) (21%) were found in RPs (crude p = 0.034 adjusted p = 0.351).

Sequence analysis

Sequences from each region were separately aligned with homologous regions from consensus HIV-1 strains D and A obtained from the Los Alamos database [http://hiv-web.lanl.gov] using version 2.2 of the Genetic Data Environment [GDE] package (Smith et al., 1994). Neighbour joining phylogenetic trees for each region were generated using the PHYLIP set of computer programs (Felsenstein et al., 2003) implemented through the Treecon package (Van de Peer et al., 1994) employing a Kimura distance matrix (Kimura, 1980). The nucleotide sequences were submitted to the Genbank; tat exon 1 AF425936 – AF425974, nef AF425870 – AF425900 and gp11 HXB2 AF425901 – AF425935.

Statistical analysis

The frequencies of residues at each amino acid position were compared between RPs and long term survivors (LTSs) in an alignment with the corresponding consensus subtypes D and A using a Chi squared test. The crude association between presence/ absence of the long tat and clinical outcome (RP vs. LTS) was evaluated using a Chi square test. To allow for the large number of statistical tests and correlations between positions, p-values adjust- ted for multiple testing were also calculated using an empirical permutation procedure with 100,000 iterations (Churchill and Doerge, 1994). To investigate the possibility that predictor residues could be specific for subtype, frequencies of residues at each position were examined separately by subtype; however formal statistical analyses were not undertaken because this was among a small sample size. Nevertheless several crude p values were found and evidence of an association at p<0.05 was provided, many of the adjusted p values were considerably larger, this was as a conse-quence of adjusting for the multiple statistical tests conducted in a small sample size. All analyses were carried out in Stata 10.

Table 1. Details of primers used.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer location on HXB2</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tat1</td>
<td>5711-5730 outer</td>
<td>5’GGATAC YTGGGMAGGAG TTG 3’</td>
</tr>
<tr>
<td>Tat2</td>
<td>6227-6207 inner</td>
<td>5’CATTGCCACGTCTTCTGTC 3’</td>
</tr>
<tr>
<td>Tat3</td>
<td>5775-5795 inner</td>
<td>5’CAGAATTG GGTGYCWACATA G 3’</td>
</tr>
<tr>
<td>Tat4</td>
<td>6137-6116 inner</td>
<td>5’CTATRGTCACACAACTATT GC 3’</td>
</tr>
<tr>
<td>EnvVII</td>
<td>7932-7952outer</td>
<td>5’GTCTGG GGCAT AAACAG CT C 3’</td>
</tr>
<tr>
<td>EnvVIII</td>
<td>8782-8761 inner</td>
<td>5’CTTCTAAGCCCTGTCGATT C 3’</td>
</tr>
<tr>
<td>EnvV</td>
<td>8004-8032 inner</td>
<td>5’GGAATTGG GGGCTGCTG 3’</td>
</tr>
<tr>
<td>EnvVI</td>
<td>8707-8686 inner</td>
<td>5’CTA7CTRTCCMCYCAGCTACTG 3’</td>
</tr>
<tr>
<td>EnvV</td>
<td>8537-8516 inner</td>
<td>5’CAAKYGGTGTAGCTGAAAGG 3’</td>
</tr>
<tr>
<td>Nef1</td>
<td>8513-8533 outer</td>
<td>5’GTGCCTCTCCAGCTACC ACCG 3’</td>
</tr>
<tr>
<td>Nef2</td>
<td>9508-9488 inner</td>
<td>5’AGCATCTGAGGGYTAGCCACT 3’</td>
</tr>
<tr>
<td>Nef3</td>
<td>8696-8717 inner</td>
<td>5’GKGAYAGATA GGYTATA GAA 3’</td>
</tr>
<tr>
<td>Nef4</td>
<td>9467-9448 inner</td>
<td>5’CRCTCCCTGGAAGTCCC 3’</td>
</tr>
</tbody>
</table>

The second exon of tat from a different reading frame in gp41 was added to the first exon.
Figure 1. Tat sequence length variation and disease progression. LF, Long form; SF, short form. The figure shows the proportion of participants with ‘long’ tat and ‘short’ tat expressed as a percentage of total RPs and LTSs.

(Figures 2A and B). Among the “short” forms of tat, serine (S) at position 75, arginine (R) at position 79, and lysine (K) at position 86 were perfectly conserved in all LTSs and the amino acid residue at these positions was the same as the consensus subtype D sequence (Figure 2B). A neighbour joining phylogenetic tree for the tat exon 1 sequences is shown in Figure 3. Eight of the study isolates clustered with reference strains subtype A. Six were RPs and two were LTSs. Twenty six (26) clustered with reference strains subtype D. Nine were RPs and 17 were LTSs. Five study isolates did not identify with specific reference strains. These were 19RP, 04RP, 12RP, 14RP and 15LTS.

Nef sequence variation

The nef region exhibited a minor sequence variation at the protein kinase C binding site (data not shown). The alanine residue in the protein kinase C binding site of nef 102'PMTYKAA'108 was more common in RPs (56%) than LTSs (15%) whereas glycine (G) at the same site was more common in LTSs (85%) than RPs (44%) (crude p=0.023, adjusted p=0.381). A similar pattern was shown by Walker et al. (2007).

Gp41 sequence variation;

The epitope ‘55NWF'SI’—‘LW64’ of the membrane-proximal external region (MPER) for gp41 neutralizing antibody ‘4E10’ showed a minor sequence variation (data not shown). The serine residue within the epitope was more common in LTSs (76.5%) than RPs (27.8%) whereas aspartic acid (D) was more common in RPs (38.9%) than LTSs (17.6%) (crude p = 0.038, adjusted p = 0.502). The ‘96YXXL99’ of the membrane-proximal cytoplasmic gp41 was conserved in the RPs and LTSs.

DISCUSSION

Campbell et al. (2004) performed a functional study on two subtype D sequences from our study population 05RP and 11LTS. A short alpha helix was observed in tat 05RP. Tat from 05RP was more efficient than tat from 11LTS in the trans-activation function. The differences between 05RP and 11LTS were the minor sequence variations at positions 8 and 63 of tat exon 1. Position 8 in the acidic region of tat exon 1 was quite conserved among the subtype A and D isolates however it exhibited sequence variation with the isoleucine residue among the rapid progressors. The acidic region is a domain for the neutralization antibody epitope of tat (Sneham et al., 2012). Sequence variation at position 8 might represent neutralization escape mutants for the subtype A isolates. Position 63 is located in the glutamine rich region. The glutamic rich region is a vital domain and plays a role in tat mediated apoptosis of the T-cells. Sequence variation in this domain may affect the rate of disease progression. In this study, the proline residue at position 63 was found among the RPs. This might play a role in enhancing tat mediated apoptosis of the T-cells and thus disease progression.

Position 79 of tat exon 2 is part of the ‘79RGD80’ motif of tat exon 2 that is involved in cellular adhesion, uptake of extracellular tat, apoptosis and enhancing HIV replication. The ‘RGD’ motif enhances the adhesion of the extracellular tat via the αvβ3 and α5β1 integrins. Extracellular tat stimulates the HIV LTR and results in the up regulation of the transcriptional process. Sequence variations at this site might therefore have various effects. Variations could down regulate or up regulate the transcription process resulting in slow or rapid disease progression. The subtype A isolates exhibited sequence variation at position 79 tat exon 2. Position 79 contained the glutamine residue ‘79Q/R’ among the RPs. The variation could not
Tat Amino acid sequence for subtype A study isolates in RPs and LTSs

A

<table>
<thead>
<tr>
<th>Acidic region</th>
<th>Cysteine rich</th>
<th>Core region</th>
<th>Basic</th>
<th>Glutamine rich</th>
<th>Cell adhesion C-terminal regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>21</td>
<td>31</td>
<td>41</td>
<td>51</td>
</tr>
</tbody>
</table>

### CONSENSUS A1

```
               MDPVDPNLEPWNHPGTSQPCSKKYCKKCTYQCFLNKGLGSGKKKRRKQGOTTFSKSGDRQNP1RTPQ1PO1Q
```

### CONSENSUS A2

```
               EA....K...........
```

### CONSENSUS A3

```
               A....T V
```

#### Regions:

- **Region 1**
  - Acidic region
  - Cysteine rich
  - Core region
  - Basic
  - Glutamine rich
  - Cell adhesion C-terminal regions

#### Figure 2:

The amino acid alignments of tat sequences from RPs and LTSs of HIV-1 infection. The three published consensus sequences A1, A2, and D were obtained from the Los Alamos National Laboratory. The (.), denotes same identity with the respective consensus sequence; (-), deletion; (x), sequence residue was not clear; the letters underlined represent residue positions of potential importance to HIV-1 disease progression; letters in bold within the consensus represent the amino acid sequences of the cell adhesion signal motif of tat and the 'long' tat; (?), sequence that could not be classified into a subtype and (*), is a stop codon.
Figure 3. The phylogenetic tree of the first exon of TAT.
be studied in LTS because of the small sample size and short sequences. The ‘long’ *tat* was exclusively found in RPs in subtype A sequences. There was some suggestion that those residues with higher frequency in the sequences of RPs were likely to coexist with the ‘long’ *tat* in subtype A although this could not be formally evaluated due to the small study numbers. *Tat* exon 2 has been shown to improve the trans-activation process and induce HIV-1 pathogenic events (Lopez-Huertas et al., 2010). Thus variations within *tat* exon 2 could improve trans-activation function and contribute to disease progression. Although the cysteine-rich and basic regions have been reported to be critical for *tat* function. Functional studies of the ‘long’ and ‘short’ *tat* are required to understand better the role of the length of *tat* in disease progression. Future comparative studies using samples from European and American subjects could also give more insight on *tat* length and disease progression although the wide spread use of antiretroviral therapy makes this difficult.

**CONCLUSION**

In this study, the ‘long’ *tat* was associated with rapid HIV-1 disease progression. The ‘long’ *tat* could be a template for developing a prognostic screening assay as well as a therapeutic target for HIV disease control.

**ACKNOWLEDGEMENTS**

The study was funded by Rogers Research Fellowship through MRC/UVRI Uganda Research Unit on AIDS in Uganda. We are very grateful to Dr. Nicaise Ndemb for his useful comments, the MRC/UVRI Uganda Research Unit on AIDS in Uganda for its full support, and the ‘Study Participants’ who willingly donated blood and made it possible to perform this study.

**CONFLICT OF INTERESTS:** The authors have declared that no conflict of interest exists.

**REFERENCES**


van Baalen CA, Pontesili O, Huisman RC, Garetti AM, Klein MR, Miedema F, Gruterset RA, Ostahaus AD (1997). Human immunodeficiency virus type 1 Rev- and *Tat*-specific cytotoxic T lymphocyte...
frequencies inversely correlate with rapid progression to AIDS. J. Gen. Virol. 78:1913-1918.

