The genomic structure of a new simian T-lymphotropic virus, STLV-PH969, differs from that of human T-lymphotropic virus types I and II

Marianne Van Brussel,* Patrick Goubau, Robert Rousseau, Jan Desmyter and Anne-Mieke Vandamme

Rega Institute for Medical Research and University Hospitals, Katholieke Universiteit Leuven, Minderbroedersstraat 10, B-3000 Leuven, Belgium

A new simian T-lymphotropic virus, STLV-PH969, was recently isolated from a wild-born Hamadryas baboon. Previous analysis had revealed that it differs sufficiently from the other HTLV/STLVs to be considered a new type, provisionally designated primate T-lymphotropic virus-L. Here we analyse a 3850 bp cDNA fragment spanning the 3' part of the STLV-PH969 genome. The fragment encodes three major proteins: Env, Tax and Rex. Sequence comparison and phylogenetic analysis indicate that in general STLV-PH969 tends to be more closely related to HTLV-II than to HTLV-I, although separate gene regions might have evolved under different constraints. Detailed comparison of the Env, Tax and Rex proteins among the HTLV-I, -II and STLV-PH969 prototypes reveals that the amino acid sequence of each protein shows a preferential conservation of functionally important domains. RNA-PCR on cytoplasmic messengers demonstrated splicing between a splice donor site immediately downstream of the env start codon, and two splice acceptor sites identified in the pX region. The predominant spliced messenger encodes both Tax and Rex. The other messenger potentially encodes a new viral protein from the proximal part of the pX region that is similar in amino acid composition to p12 of HTLV-I and HTLV-II respectively. This genomic organization of the proximal pX region of STLV-PH969 is different from that found in HTLV-I and HTLV-II. Therefore, the distinct classification of this virus can be justified, not only in terms of sequence divergence but also in terms of its different genomic structure.

Introduction

Human and simian T-lymphotropic viruses (HTLV, STLV) are complex retroviruses that share the ability to transform T cells. In vivo, HTLV-I has been associated with leukaemia (Poiesz et al., 1980) and a neurological disorder (Gessain et al., 1985; Osame et al., 1987). Much less is known about the pathogenicity of HTLV-II and this virus has only recently been associated with chronic myelopathy (Murphy et al., 1993).

These viruses are disseminated worldwide and are characterized by their genetic stability and limited horizontal transmission. HTLV-I and STLV-I cannot be separated into distinct phylogenetic lineages according to the species of origin, indicating that several interspecies transmissions have occurred (Koralnik et al., 1994; Vandamme et al., 1994; Saksena et al., 1994; Song et al., 1994). Therefore, the HTLV/STLV group should more appropriately be designated as primate T-lymphotropic viruses (PTLVs). The origin of HTLV/STLV is still a matter of debate. The existence of very divergent HTLV-I and STLV-I strains in Asia (Gessain et al., 1993; Song et al., 1994) together with the presence of HTLV-II in Amerindian tribes (Ferrer et al., 1992, 1993; Gessain et al., 1995) support an Asian origin of HTLV/STLV. This hypothesis has been put into question by a phylogenetic analysis of the LTR regions of HTLV-I and STLV-I strains (Vandamme et al., 1994) which suggested a possible African origin of HTLV-I/STLV-I. Moreover, HTLV-II has been found in African pygmies (Goubau et al., 1992, 1993; Gessain et al., 1995), indicating that the highest genetic divergence among all HTLV/STLV strains is found in Africa.

Searching for new clues to the evolution of the HTLVs and STLVs, we identified a virus, STLV-PH969, that differs sufficiently from the known types to be considered as a new STLV type (Goubau et al., 1994). The virus was isolated from a African baboon (Papio hamadryas), born in Eritrea, with a divergent HTLV like antibody pattern. After establishment of a virus producing cell line (PH969), we isolated an 1802 bp cDNA fragment that extends from the env region to part of the tax/rex genes.

* Author for correspondence. Fax +32 16 332131.
e-mail vanbruss@uz.kuleuven.ac.be

The sequence reported in this paper has been deposited in the EMBL database under accession number X88852.

0001-3469 © 1996 SGM
Pending formal classification, the virus was provisionally designated PTLV-L with STLV-PH969 as the prototype strain. Our group and others have independently reported the discovery of yet another divergent STLV in a bonobo (Pan paniscus), which cannot readily be assigned to one of the existing types, although it seems more related to type II than to the others (Liu et al., 1994; Giri et al., 1994; A.-M. Vandamme, unpublished). The presence of both these new viruses in African simians further supports an African origin for the HTLV/STLVs as a group.

The genomic organization of HTLV-I, HTLV-II and bovine leukaemia virus (BLV), which all belong to the same genus (Sagata et al., 1984) differs from the other retroviruses in the presence of a region, called pX, downstream of the env gene, encoding the regulatory proteins Tax and Rex. Both proteins are encoded by two exons on a bicistronic messenger; one is situated immediately upstream of the env gene and the second is encoded by the distal part of the pX region (Sagata et al., 1985; Nagashima et al., 1986; Rosenblatt et al., 1988).

Additional messengers generated by complex alternative splicing were recently characterized for HTLV-I, HTLV-II and BLV. These messengers differ in the different virus types and potentially encode new proteins with unknown function. For HTLV-I, p121, p131, p301, Rof and p21rex have been described (Koralnik et al., 1992; Ciminael et al., 1992). Three novel proteins, p101v, p11v and p281v, and several isoforms of Rex have been identified for HTLV-II (Ciminael et al., 1995). One to three new proteins have been identified for BLV: Tof (Ciminale et al., 1993) and RIII and GIV (Alexandersen et al., 1993).

In this study we report the isolation of the 3' part of the STLV-PH969 genome and we characterize the genes of three major proteins which are encoded by this region: the complete env gene and the genes encoding Tax and Rex. We elaborate on the possible function in the encoded proteins of highly identical stretches of amino acids. In unravelling the genomic organization of the pX region, we provide further evidence for the existence of viral messengers generated through alternative splicing in the pX region.

**Methods**

**Cell lines.** An STLV-PH969 expressing cell line (PH969) was obtained by cocultivation of human cord blood lymphocytes with peripheral blood mononuclear cells from a Hamadryas baboon that exhibited an HTLV-like antibody pattern (Goubau et al., 1994). Cells were passaged in RPMI 1640 medium (Life Technologies) supplemented with 15% fetal calf serum, 1 mM-L-glutamine and 10 U/ml interleukin-2 (Boehringer Mannheim). Similarly, a second bleed from the same baboon was used to obtain an STLV-PH969 producing baboon cell line (PH1824) by cocultivation with blood lymphocytes from a seronegative baboon.

**cDNA library and screening.** Construction of the cDNA library from PH969 poly(A) RNA was described previously (Goubau et al., 1994). A first screening of the cDNA library, with a 5'P-labelled PCR fragment, resulted in the isolation of an 1802 bp cDNA insert (Goubau et al., 1994). After incorporation of [x-32P]dCTP using a Random Primed DNA labelling kit (Boehringer Mannheim), this cDNA insert was used as a specific probe in a second screening of the library (8 x 106 plaques). Positive unpurified clones with the longest cDNA insert were selected by a PCR based method: 50 µl of phage lysate was digested with proteinase K (Boehringer Mannheim) and used in a PCR with a vector specific (T7 or T3), and an STLV-PH969 specific primer. The reactions were performed under standard conditions (see "RNA-PCR"). After purification of selected clones, the cDNA insert was excised from the λZAP-II (Stratagene) vector, cloned into pBluescript and sequenced in both directions by primer walking.

**RNA-PCR.** Poly(A) RNA was isolated from PH969 and PH1824 cells using the Quickprep Micro mRNA purification Kit (Pharmacia) and reverse transcribed with SuperScript (Life Technologies), using random priming, as recommended by the manufacturer. cDNA from 106 cells was used in one PCR reaction with AmpliTaq (Perkin-Elmer) under the following conditions: 10 mM-Tris-HCl pH 8.3, 50 mM-KCl, 200 µM-dNTPs, 1 µM-primer. The MgCl2 concentration ranged from 1.5-3 mM, optimized for each primer set. The following primer sets were used for the analyses of the viral transcripts: (i) MYBREXex1, 5' AAATCAGAGACCATCAACTCC 3' (sense, nt 158-178 of the cDNA insert) and MYBREXex2, 5' TCCCCAG GGAATCTGATGTT 3' (antisense, nt 2503-2521); (ii) MYBREXex1 and MVBpX2, 5' GAT CAAAGTTGCATGCCTCAG 3' (antisense, nt 2112-2132) (see Fig. 1). Oligonucleotides were synthesized by Pharmacia Biosystems. A total of 35 cycles (95 °C for 1 min, 50 °C for 1 min and 72 °C for 30 s) were performed on a Trio-thermoblock (Biometra). Aliquots of the amplified products were analysed on 6 % polyacrylamide gels. PCR fragments were purified on 3 % agarose gels using Wizard (Promega) and sequenced directly with the dsDNA Cycle Sequencing system (Life Technologies) as described previously (Goubau et al., 1994).

**Sequence analysis.** Nucleotide and amino acid sequences were analysed using the GeneWorks software package (IntelliGenetics), including matrix comparison plots and hydrophobicity scores. The FASTA program (Lipman & Pearson, 1985), implemented in GeneWorks, scores each deletion as one event. When calculating percentage identities, deletions and insertions were scored as one event. Phylogenetic analysis of protein sequences was performed by the neighbour-joining method (Saitou & Nei, 1987) as implemented in the Phylip software package (Felsenstein, 1989). Distances were calculated with a PAM 001 matrix. For each protein analysed, we used the entire sequence except for some highly divergent parts of the sequence for which no proper alignment could be made, e.g. the signal peptide of the surface protein. The BLV surface protein, which is only distantly related to the HTLV/STLV proteins was aligned using a hydrophobic cluster analysis as a guide (Calletsba et al., 1994). For the analysis of the HTLV/STLV and BLV transmembrane proteins we used the N-terminal 167 aa, since on pairwise matrix comparison plots only this region showed significant identity between the BLV transmembrane protein and its HTLV/STLV counterparts. Insertions and deletions were not taken into account. Statistical analysis was done by bootstrapping (1000 samples).

**Results**

**Nucleotide sequence analysis and genomic organization of the 3' part of the STLV-PH969 genome**

We established human (PH969) and baboon (PH1824) cell lines that produced the recently described new type
of simian T-lymphotropic virus, STLV-PH969. From a cDNA library of the PH969 cell line, 20 clones containing cDNA fragments of STLV-PH969 viral messengers were isolated. The longest clone, 3850 bp, was further purified and sequenced. This cDNA fragment is homologous to the HTLV-I(ATK1) sequence spanning nt 5052–8638.

### Table 1. Nucleotide sequence comparisons of the HTLV/STLV prototypes HTLV-I(ATK1), HTLV-II(Mo) and STLV-PH969

<table>
<thead>
<tr>
<th>Type</th>
<th>General</th>
<th>env</th>
<th>SU</th>
<th>TM</th>
<th>Proximal pX</th>
<th>tax</th>
<th>tax/rex overlap</th>
</tr>
</thead>
<tbody>
<tr>
<td>STLV-PH969 vs HTLV-I</td>
<td>63%</td>
<td>66%</td>
<td>64%</td>
<td>67%</td>
<td>45%</td>
<td>73%</td>
<td>79%</td>
</tr>
<tr>
<td>STLV-PH969 vs HTLV-II</td>
<td>66%</td>
<td>68%</td>
<td>64%</td>
<td>74%</td>
<td>52%</td>
<td>74%</td>
<td>77%</td>
</tr>
<tr>
<td>HTLV-I vs HTLV-II</td>
<td>65%</td>
<td>66%</td>
<td>64%</td>
<td>68%</td>
<td>35%</td>
<td>76%</td>
<td>80%</td>
</tr>
</tbody>
</table>

* env, envelope gene; SU, surface protein coding sequence; TM, transmembrane protein coding sequence.

† The numbering is according to the first nucleotide of the cDNA fragment.

![](image.png)
Fig. 2. (a) Gel electrophoresis of PCR products obtained from identified spliced messengers with the following primer sets: MVBREXex1 and MVBREXex2, lanes 1-5; MVBREXex1 and MVBpX2, lanes 6-10. The amplification products were obtained from poly(A) RNA from PH969 cells (lanes 2 and 7) and from PH1824 cells (lanes 3 and 8). Negative controls were performed on genomic DNA from the same cell lines (lanes 4 & 9 and 5 & 10, respectively) and water (lanes 1 and 6). The positions of the primers are indicated in Fig. 1 (b). The estimated lengths of the PCR fragments obtained from spliced viral messengers were 211 and 192 nt respectively for each primer set. PCR fragments of the desired length (indicated with an arrow) were further analysed by sequencing. Lane 0, molecular mass marker (pBR322 DNA digested with HaeIII). (b) Nucleotide sequence of the PCR fragments around the splice junction site. Panel (i) is from the viral messenger generated by splicing between sd-tr and sa-pX2, panel (i/) by splicing between sd-tr and sa-tr. The positions of the splice junctions are indicated by arrows. Sequencing was performed with the reverse primer. (c) Amino acid and nucleotide sequences of RorfII. The N-terminal stretch of amino acids generated from the first coding exon of Rex is in bold. The position of the splice site is indicated with an arrow.

and to nt 5010-8751 of the HTLV-II(Mo) sequence, which include the first coding exon of rex, the complete env gene and the pX region extending into the 3'LTR. The percentage nucleotide identity of these different gene regions in the STLV-PH969 cDNA fragment and HTLV-I and -II prototypes ranges from 64% (surface protein encoding region) to 80% (tax/rex region) (Table 1).

Several ORFs were identified, three of which correspond to the ORFs of known proteins of the HTLV/BLV genus. The position of these ORFs with respect to the length of the cDNA fragment is given in Fig. 1 (a). The env gene (nt 180-1655) encodes a putative envelope precursor of 490 aa, which has the potential to be processed into a transmembrane (TM) protein and a surface (SU) protein, of 178 aa and 314 aa, respectively. Downstream of env, two overlapping ORFs (ORF III and ORF IV; Fig. 1a) are homologous to the major Rex and Tax ORFs, encoded in the pX region of HTLV-I and -II. A small ORF (nt 121-183, Fig. 1a), upstream of env, is homologous to the minor Rex ORF encoded in the corresponding region of the HTLV-I and -II genome. In addition, a putative splice donor sequence (sd-tr; Fig.
Table 2. Protein sequence comparisons of the HTLV/STLV prototypes HTLV-I(AKTI), HTLV-II(Mo) and STLV-PH969

The percentage identities were calculated with FASTA implemented in the Geneworks software package (for details see Methods).

<table>
<thead>
<tr>
<th>Type</th>
<th>Percentage amino acid identity in protein:</th>
<th>Env*</th>
<th>Tax†</th>
<th>Rex</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SU</td>
<td>TM</td>
<td>N-term.</td>
<td>C-term.</td>
</tr>
<tr>
<td>STLV-PH969 vs HTLV-I</td>
<td>59%</td>
<td>75%</td>
<td>78%</td>
<td>82%</td>
</tr>
<tr>
<td>STLV-PH969 vs HTLV-II</td>
<td>64%</td>
<td>88%</td>
<td>78%</td>
<td>86%</td>
</tr>
<tr>
<td>HTLV-I vs HTLV-II</td>
<td>62%</td>
<td>79%</td>
<td>77%</td>
<td>79%</td>
</tr>
</tbody>
</table>

* Env, envelope protein; SU, surface protein; TM, transmembrane protein.
† The N-terminal part of Tax is the region spanning two high identity regions, i.e. aa 1–176 according to the consensus sequences indicated in Fig. 4.

1b) was found immediately downstream of the env initiation codon. Together with a putative splice acceptor sequence immediately upstream of the major tax and rex ORF (sa-tr; Fig. 1b) in the distal part of the pX region, these data suggest that Tax and Rex might be encoded on a viral messenger generated by splice junction between sd-tr and sa-tr. To obtain experimental evidence for this hypothesis, we performed an RNA–PCR on poly(A) RNA isolated from PH969 and PH1824 cells. We chose primers (MVBREXex1 and MVBREXex2) flanking each side of the putative intron. The positions of the primers are indicated in Fig. 1(b). A 192 bp PCR fragment found in both cell lines (lanes 2 and 3, Fig. 2a) was purified and analysed by cycle sequencing. The sequence of this fragment (Fig. 2b) is evidence for the existence of a spliced messenger generated through the use of the splice donor downstream of the env start codon (sd-tr) and the splice acceptor preceding the tax and rex ORFs at pX (sa-tr). This messenger encodes a Tax protein of 350 aa, initiated from the env start codon, while a Rex protein, initiated upstream of env is 182 aa long. As in HTLV-I and -II, additional start codons are present in ORF III (asterisk, Fig. 1a).

In the proximal pX region, a large ORF (ORF I; Fig. 1a) has been found, immediately preceded by a putative splice acceptor site (Goubau et al., 1994). In this region, two nucleotide insertions were found between the clone described here and the 1802 bp clone in Goubau et al. (1994): in the new fragment a G was inserted at nt 2026 and a C at nt 2038. Both insertions were confirmed by direct sequencing of proviral DNA from cell line PH969. As a consequence, the ORF I originally described in Goubau et al. (1994) extends for another 18 aa. Computer searches of the protein databases revealed similarities between the C-terminal part of the predicted ORF I amino acid sequence and proline and serine rich proteins, including the p11Y sequence of HTLV-II. Downstream of ORF I, we identified yet another ORF (ORF II; Fig. 1a), immediately preceded by a sequence that corresponds to an additional splice acceptor consensus sequence (Fig. 1b). The predicted 64 aa protein sequence of ORF II is highly hydrophobic and leucine rich (Fig. 2c), two features also found in the p12I and p10X of HTLV-I and HTLV-II respectively (Ciminale et al., 1992, 1995).

The presence of two putative splice acceptor sequences upstream of ORF I and ORF II, together with the fact that none of the ORFs is initiated by an AUG start codon, suggested that putative proteins encoded by these ORFs might be initiated upstream through splicing. Therefore, we investigated the possible existence of splicing between the sd-tr splice donor and the putative splice acceptor sequences in the proximal pX region.

RNA–PCR was performed on poly(A) RNA isolated from the two STLV-PH969 producing cell lines, PH969 and PH1814. To detect a viral messenger generated by splice junction between the sd-tr splice donor and the putative splice acceptor sequence preceding ORF I, we performed several PCRs, each one using a primer upstream of the sd-tr splice donor and a primer in the ORF I region. Although different primer sets were used, no specific PCR fragment that would reveal the presence of such a messenger could be obtained. Liquid hybridization demonstrated that specific PCR fragments were generated only from unspliced messengers (data not shown).

The existence of splicing between the sd-tr splice donor and the splice acceptor site (sa-pX2; Fig. 1b) preceding ORF II was analysed by RNA–PCR with primers MVBREXex1 and MVBpX2 (Fig. 1b). A 211 bp PCR fragment was obtained from PH969 and PH1824 poly(A) RNA (lanes 7 and 8; Fig. 2a). Sequencing (Fig. 2b)...
(a) SU protein

HTLV-I (ATK1)  FLATLIL F.QFCP.IFG DYS.C V.K ... T.LA.SA A.Q ... VS.S G
HTLV-II (Mo)   .NY.-FL L.5LTHF.LA Q-O ... I.S.T. ... N.T. R.H ... K
STLV-PH969    -.GLYC. VHYIL. AS SGN ... F.A ... .GSSL.R ... F.R. S.S ... D.

Consensus  A.A ... T.N.G ... A.K ... A.Y ... QH ... RINH ... F.R ... Q.P.N ... A.SA ... H.Y ... H.N ... HS ... S...

Consensus  YHKTSYLFL PHNIEKPR. GLCYSASYS DPCSQCPYL GCQSWTCPTY GPVSSP.WKF ... DILEQEV S.VSL

(b) TM protein

HTLV-I (ATK1)  ...V ... M.A.V. RL.N ... G ... H ... Q ... K.L.KI ...
HTLV-II (Mo)   ...V ... M.A.V. RL.N ... G ... H ... Q ... K.L.KI ...
STLV-PH969    ...V ... M.A.V. RL.N ... G ... H ... Q ... K.L.KI ...

Consensus  A.VPIA.VWLVS ALAAGTVGAG GVTGSLSLAS SKSLL.EVDK DISHL~AIV KNH.NILRVA QYAAQNRRGL DLLFW

Consensus  SLSVPF ... R.RR

(b) TM protein

HTLV-I (ATK1)  ...V ... M.A.V. RL.N ... G ... H ... Q ... K.L.KI ...
HTLV-II (Mo)   ...V ... M.A.V. RL.N ... G ... H ... Q ... K.L.KI ...
STLV-PH969    ...V ... M.A.V. RL.N ... G ... H ... Q ... K.L.KI ...

Consensus  A.VPIA.VWLVS ALAAGTVGAG GVTGSLSLAS SKSLL.EVDK DISHL~AIV KNH.NILRVA QYAAQNRRGL DLLFW

Consensus  SLSVPF ... R.RR

(b) TM protein

HTLV-I (ATK1)  ...V ... M.A.V. RL.N ... G ... H ... Q ... K.L.KI ...
HTLV-II (Mo)   ...V ... M.A.V. RL.N ... G ... H ... Q ... K.L.KI ...
STLV-PH969    ...V ... M.A.V. RL.N ... G ... H ... Q ... K.L.KI ...

Consensus  A.VPIA.VWLVS ALAAGTVGAG GVTGSLSLAS SKSLL.EVDK DISHL~AIV KNH.NILRVA QYAAQNRRGL DLLFW

Consensus  SLSVPF ... R.RR

(b) TM protein

HTLV-I (ATK1)  ...V ... M.A.V. RL.N ... G ... H ... Q ... K.L.KI ...
HTLV-II (Mo)   ...V ... M.A.V. RL.N ... G ... H ... Q ... K.L.KI ...
STLV-PH969    ...V ... M.A.V. RL.N ... G ... H ... Q ... K.L.KI ...

Consensus  A.VPIA.VWLVS ALAAGTVGAG GVTGSLSLAS SKSLL.EVDK DISHL~AIV KNH.NILRVA QYAAQNRRGL DLLFW

Consensus  SLSVPF ... R.RR

(b) TM protein

HTLV-I (ATK1)  ...V ... M.A.V. RL.N ... G ... H ... Q ... K.L.KI ...
HTLV-II (Mo)   ...V ... M.A.V. RL.N ... G ... H ... Q ... K.L.KI ...
STLV-PH969    ...V ... M.A.V. RL.N ... G ... H ... Q ... K.L.KI ...

Consensus  A.VPIA.VWLVS ALAAGTVGAG GVTGSLSLAS SKSLL.EVDK DISHL~AIV KNH.NILRVA QYAAQNRRGL DLLFW

Consensus  SLSVPF ... R.RR

Protein sequence comparison

The predicted amino acid sequences of the STLV-PH969 proteins were compared with their HTLV-I and HTLV-II analogues. A general comparison based on amino acid

Fig. 3. Protein sequence alignment of HTLV/STLV prototypes. The consensus sequences are at the bottom. Shaded boxes represent regions of high identity (supported by PAM 250 matrix comparison plots). The numbering (in bold) refers to the first and last amino acid in each box, according to the consensus sequence. (a) SU protein: neutralizable regions (NTR 1, 2 and 3) in the SU protein of ATK1, discussed in the text, are indicated under the consensus sequence (filled bars). The region corresponding to the HTLV-II specific K55 protein (Lipka et al., 1991) is indicated (open bar). Further details are given in the text. The potential glycosylation sites are indicated by circled asparagine residues. (b) TM protein. (c) Tax protein: the nuclear localization domain (open bar) and the zinc finger domain (filled bar) are indicated. (d) Rex protein: the nucleolar targeting signal (open bar) and the putative activation domain (filled bar) are indicated.

indicated that this fragment was obtained from a viral messenger generated through splicing between sd-tr and sa-pX2. This transcript has the potential to encode an 84 aa long protein, designated Rorf II, initiated at the first coding exon of Rex and extending to ORF II (Fig. 2c).
sequence identity is shown in Table 2. To obtain information about local differences in similarity, we performed a more detailed analysis based on matrix comparison plots with a PAM 250 score (not shown). As we focused on major differences we restricted this analysis to the HTLV-I, -II and STLV-PH969 prototypes. Protein alignments of the envelope proteins, Tax and Rex of HTLV-I (ATK-I), HTLV-II (Mo) and STLV-PH969 are shown in Fig. 3. Within each protein alignment, regions of higher and lower identity are outlined (regions of high identity are shaded in Fig. 3).

The surface and transmembrane glycoproteins

The surface (SU) proteins of HTLV-I, -II and STLV-PH969 (SU-I, -II and -PH969, respectively) have similarities ranging from 59–64%, with the highest identity between the SU proteins of types II and PH969. High identity was observed in a region around aa 65–131, with about 78–80% identity between SU-I, -II and -PH969. A linear type specific neutralization epitope (NTR 1; solid bar, Fig. 3a) has been mapped in this region of SU-I and SU-II (aa 88–98 and aa 84–94 respectively; Palker et al.,...
1992). The same region has been shown by site-directed mutagenesis to be involved in syncytium formation (aa 75–101; Delamarre et al., 1994). The hydrophobicity of this possible epitope in SU-PH969 is similar to the one found in SU-II (data not shown).

A number of neutralization epitopes of SU-I have been mapped in a central part of the molecule around aa 175–199 (Baba et al., 1993). Similarly, a neutralization epitope has been characterized in the corresponding region of the SU-II protein (aa 186–192; Tanaka et al., 1994). In contrast with the NTR 1 region, this NTR 2 region (Fig. 3a) is highly divergent between the three types. Between HTLV-I and HTLV-II, this has been demonstrated by the fact that several linear type specific epitopes which overlap the NTR 2 region have been identified for both types (Lal et al., 1992). An HTLV-II type specific recombinant protein, K55 (Fig. 3a), that overlaps the NTR 2 region (aa 162–205; Lipka et al., 1991) cross-reacts with STLV-PH969 positive serum (Goubau et al., 1994) whereas no reaction was observed with MTA-1, the corresponding HTLV-I specific peptide. We characterized a third region of high similarity at aa 216–241 (shaded box, Fig. 3a). For SU-I, a third linear neutralizing B-cell epitope (NTR 3; see Fig. 4a) has been mapped within this region (aa 213–236; Baba et al., 1993).

The sequence of the transmembrane (TM) protein (Fig. 3b) seems to be highly conserved among HTLV-I, -II and STLV-PH969 (TM-I, TM-II and TM-PH969, respectively) (see also Table 2). This is in agreement with the cross-reactivity observed between HTLV-II and STLV-PH969 positive sera and the TM-I protein (Goubau et al., 1994), suggesting that the structure of these proteins is highly similar. The amino acid substitutions between TM-PH969 and TM-II are concentrated around aa 36–54 and in the C-terminal 41 aa. According to a human immunodeficiency virus (HIV) based model of the TM-II protein (Gallagher et al., 1989), the former stretch of amino acids might belong to an α-helical region, while the latter region partially overlaps the putative membrane spanning domain and adjacent cytoplasmic tail.

**The regulatory proteins Tax and Rex**

The Tax protein is highly conserved (Table 2, Fig. 3c). We noticed a difference in local similarity between the N-terminal and C-terminal parts of the protein alignment. In the N-terminal part of the protein (aa 1–176), the Tax proteins of STLV-PH969 (Tax-PH969) and HTLV-II (Tax-II) have the highest identity (86%). For Tax-I, several domains important for Tax function and stability have been mapped in this region. In the nuclear localization signal (Smith & Greene, 1992) and in the overlapping zinc finger domain (Semmes & Jeang, 1992), only a few differences are found between Tax-PH969 and Tax-I.

The C-terminal part of the protein (aa 176–353) is rather less conserved between Tax-PH969 and Tax-II than in either of these compared with Tax-I (Table 2). Site-directed mutagenesis has indicated that the C-terminal part of Tax is involved in interaction with the basal transcription machinery (Adya & Giam, 1995). Two regions of high similarity were found in this part of Tax: aa 242–274 and aa 296–316. The latter overlaps a recently defined minimal activation domain (Semmes & Jeang, 1995). As with HTLV-I and HTLV-IIb (Pardi et al., 1993), STLV-PH969 has an extended Tax compared to HTLV-Ila. Although divergent, the highly hydrophilic character of this extension remains conserved.

In contrast with Tax, the Rex protein is highly divergent among the three types (Table 2, Fig. 3d). Rex-PH969 is more divergent from Rex-I and Rex-II than the latter are from each other (Table 2). However, several functional features remain conserved in Rex-PH969. In common with Rex-I and Rex-II, the N-terminal part of Rex-PH969 is highly basic, which is a major characteristic of the nucleolar targeting signal, identified in this region for Rex-I (Siomi et al., 1988). Moreover, for all three proteins there is a striking concentration of arginine residues in this region of Rex, which has been proven to be important for binding of Rex-I to the Rex responsive element in the LTR (Hammes & Greene, 1993). For Rex-I, a putative activation domain (Fig. 3d) that also interacts with the Rev responsive element of HIV was mapped in the central part of the Rex protein (aa 79–99; Weichselbraun et al., 1992). Apart from a small interruption in the middle, this whole region and adjacent residues are highly conserved in Rex-II and Rex-PH969 (shaded in Fig 3d). By analogy with the HIV Rev activation domain (Malim et al., 1991), it has been suggested that a leucine rich core (aa 81–91) might play a major role in Rex-I activation (Hope et al., 1991). The conservation of these leucine residues in Rex-I and Rex-II supports this hypothesis. In addition, a clustering of proline residues, also found around the Rev activation domain, is completely conserved among Rex-I, -II and -PH969.

**Phylogenetic analysis**

A phylogenetic analysis was performed on three functional gene regions of the HTLV/STLV genome encoding the envelope proteins (SU, Fig. 4a; TM, Fig. 4b) and the Tax protein (Fig. 4c). We included only the most divergent strains within HTLV-I and HTLV-II to obtain information on how STLV-PH969 relates to these types. To root the trees with BLV (arrow, Fig. 4), which is only
Genomic analysis of the 3' part of STLV-PH969

Distantly related to the PTLV family (40% identity at the nucleotide level), we focused on protein sequences, which are generally better conserved than nucleotide sequences. The unrooted trees produced by the neighbour-joining method are supported by high bootstrap values, ranging from 83–100%. Analyses based on the nucleotide sequences of these proteins gave similar results. In the tree based on the Tax protein, no pronounced differences between STLV-PH969, HTLV-I and HTLV-II in relative branch lengths are found, whereas in the other two trees, STLV-PH969 is clearly closer to HTLV-II than to HTLV-I, as already suggested by the similarity study (Table 2). The topology of the root was different in the tree based on the SU protein than in the other two trees. In this tree, HTLV-I separated from the common ancestor before HTLV-II and STLV-PH969 diverged from each other, with a high bootstrap value (87%). In the analyses of the TM and Tax proteins, the root node was placed on the Tax-PH969 branch, although with less confidence since the bootstrap value on the root was only 73% for both trees. The alternative topology for the TM and Tax tree places the root on the HTLV-I branch (26% and 23%, respectively) as with the SU protein.

Discussion

In this study we report the isolation of a 3850 bp cDNA fragment that represents the 3' part of the genome of a new simian T-lymphotropic virus, STLV-PH969. Since the cDNA fragment was isolated from a cDNA library of poly(A)RNA from the virus producing cell line PH969, it was probably reverse transcribed from an unspliced gag/pol or singly spliced env viral messenger. The fragment contains the ORFs for three major proteins, Env [the precursor of the surface (SU) and transmembrane (TM) proteins], Rex and Tax (ORF III and ORF IV respectively). In addition, we identified two ORFs (ORF I and ORF II) in the proximal part of the pX region. Although no clear identity was found with any known HTLV or BLV protein, the amino acid sequences of ORF I and ORF II have some similarities in amino acid composition with the proteins of the proximal part of the pX region of HTLV-I and HTLV-II.

Based on their similarity with the consensus sequence (Shapiro & Senapathy, 1987), one potential splice donor (sd-tr; Fig. 1b) was identified downstream of the env start codon and three putative splice acceptors preceding ORFs I, II, III and IV in the pX region. Using PCR and

TM protein and Tax, an alternative topology for which Tax-I clusters with Tax-II, placing the root node on the Tax-I branch, is supported by 26% and 23%, respectively, of the bootstrap replicates.
sequence analysis on two STLV-PH969 producing cell lines, we showed the existence of viral messengers generated by the use of two of these sites (sa-tr and sa-pX2; Fig. 1b). A splice junction occurs between the sd-tr splice donor and the major sa-tr splice acceptor, generating a viral messenger encoding Tax and Rex. Splicing was also demonstrated between the sd-tr splice donor and sa-pX2, a splice site in the proximal part of the pX region. The viral messenger generated by this splice junction potentially encodes a new viral protein, Rorf II, initiated at the AUG start codon of Rex and extended into ORF II. Together, our data indicate that STLV-PH969, in common with the other T-lymphotropic viruses, uses complex splicing to encode additional viral proteins. It remains unclear whether ORF I is used as an exon through alternative splicing, since no viral messenger encoding this ORF could be detected.

Cells infected by HTLV-I, HTLV-II and BLV express additional proteins (e.g. p21rex of HTLV-I) encoded on messengers that are generated by splicing between the splice donor in the 5'LTR and splice acceptors in the pX region (Koralnik et al., 1992; Ciminale et al., 1992, 1995). The presence of AUG codons in ORF III of STLV-PH969 suggests that truncated isoforms of Rex might be expressed in STLV-PH969 infected cells also. As with HTLV-I and -II, these isoforms might be encoded on a singly spliced messenger. Since our present data on the STLV-PH969 genome did not allow detection of these viral messengers, further analysis is needed in this context.

The proximal part of the pX region is one of the most divergent regions of the HTLV genome. Although similarities might be found, the genomic organization of the proximal pX region differs between HTLV-I, HTLV-II and BLV. Our previous analysis of HTLV-I, -II and STLV-PH969 by matrix comparison plots showed significant similarity between STLV-PH969 and HTLV-II in a 100 nt stretch (Goubau et al., 1994) of this region. Because of important deletions in the HTLV-II pX region, the proximal pX region remains very different in HTLV-II and STLV-PH969. Moreover, the Rorf II amino acid sequence shows only poor similarity with the corresponding gene products of HTLV-I and HTLV-II. Although partial, our data indicate that the genetic organization of the STLV-PH969 pX region differs from that in the HTLV-I and HTLV-II genomes (Fig. 5). Hence, this further justifies the distinction of this virus as a new HTLV/STLV type.

Comparison of the general similarity and the identities among the partial gene regions confirms previous results that differences between STLV-PH969 and the known HTLV types are as great as the differences between HTLV-I and HTLV-II. Together with the phylogenetic analysis of three gene regions of the HTLV/STLV prototypes, our data indicate that, although phylogenetically distant, STLV-PH969 tends to be closer to HTLV-II than to HTLV-I. The evolutionary distance depends on the gene region, suggesting that some regions might have evolved under different constraints and at different evolutionary rates. When the phylogenetic trees were rooted with the BLV sequence, the topology of the tree based on the SU protein differed from that of trees based on the Tax and TM proteins. Although these data do not predict the phylogenetic relationship for the entire virus, the fact that in all analyses the root was placed on either the HTLV-I branch or the STLV-PH969 branch suggests that the HTLV-II branch would be the youngest.

Detailed comparison of the amino acid sequences of the SU, TM, Tax and Rex proteins of STLV-PH969 with their HTLV-I and -II analogues revealed regions of higher and lower identity within these sequences. For all four proteins, some of these regions overlap with domains important for function and/or maturation. This is particularly obvious in the Rex protein of STLV-
PH969: despite the high divergence of Rex-PH969 compared to Rex-I or Rex-II, several features which were previously mentioned in structure–function studies of Rex-I and Rex-II remain conserved in Rex-PH969. The conservation of the Tax protein in the HTLV family suggests that Tax has critical constraints with regard to its function. The extension of Tax-PH969, Tax-I and Tax-Iib compared to Tax-Ila suggests that Tax-Ila has lost this C-terminal end because of the generation of a premature termination codon, which means that the longer Tax protein more closely resembles the ancestral Tax.

For HTLV-I, the involvement of three neutralizable regions of the SU protein (NTR1, NTR2 and NTR3; Fig. 4a) in the syncytium forming capacity of infected cells was recently demonstrated by site-directed mutagenesis (Delamarre et al., 1994). Whether these domains have similar functions in SU-L remains to be established. Since no clear identity to the amino acid sequence of K55 (Lipka et al., 1991) was found at the corresponding part of the STLV-PH969 SU protein, the cross-reactivity of STLV-PH969 positive antiserum with this HTLV-II type specific recombinant protein cannot be explained by general sequence conservation. The hypothesis that STLV-PH969 cross-reacting epitopes within K55 might be present, however, is supported by the existence of at least one cross-reacting epitope within this region between SU-I and SU-II (aa 183–191; Edouard et al., 1994). On the other hand, the large number of amino acid differences within this part of the surface protein might allow the development of type specific antibodies raised against this region, for which STLV-PH969 reacts differently compared to HTLV-I and HTLV-II.

The regulatory proteins Tax and Rex, the TM protein and especially the SU protein are important in virus–cell interactions. The amino acid differences in the STLV-PH969 proteins will influence the function and interaction of these proteins with cellular components. The large divergence of the proteins of this third type of primate T-lymphotropic virus is useful for comparative structure–function relationship of the proteins and the biology of the HTLV/BLV genus.

The identification of STLV-PH969 has provided better insight into the evolution of the HTLVs/STLVs as a group. The characterization of its proteins and genomic structure could allow the development of improved diagnostic methods in the search for unknown T-lymphotropic viruses in simians and in man.

We thank Josette Costermans, Martine Michiels, Sonia Van Dooren and Kurt Beuselinck for excellent technical assistance. This work was supported by grant no. 3.0098.94 from the Belgian National Foundation for Scientific Research (NFWO/FGW0) and by a special grant from the Collen Research Foundation.

References


