Mutation rates in LTR of HTLV-1 in HAM/TSP patients and the carriers are similarly high to Tax/Rex-coding sequence

Mineki Saito, Yoshitaka Furukawa, Ryuji Kubota, Kouichiro Usuku, Shuji Izumo, Mitsuhiro Osame and Mitsuaki Yoshiha

Department of Cellular and Molecular Biology, The Institute of Medical Science, The University of Tokyo, Shirokanedai, Minato-ku, Tokyo 108; Third Department of Internal Medicine and The Division of Molecular Pathology, Center for Chronic Viral Disease, Faculty of Medicine, Kagoshima University, 8-35-1 Sakuragaoka, Kagoshima 890, Japan

The genomic sequence of human T-cell leukemia virus type 1 (HTLV-1) is highly conserved, although minor sequence variations enable classification of the isolates into several subgroups. We previously reported, however, that the Tax-coding sequence of HTLV-1 genome is highly variable in a random fashion within individuals with HAM/TSP and asymptomatic carriers. Here, we describe frequent base substitutions in the LTR sequence similarly to those in Tax-coding sequence. These observations indicate that frequent mutations are not unique to the sequence encoding the most effective antigen for cytotoxic T lymphocytes, but also seen in the LTR, a non-coding sequence. Thus, frequent mutations seem to occur during the viral replication process rather than the selection of rare mutants by immune surveillance.

Keywords: HAM/TSP; HTLV-1 instability; intrastain variability; LTR variation

Introduction

Infection by human T-cell leukemia virus type 1 (HTLV-1) (Poiesz et al., 1980; Hinuma et al., 1981; Yoshida et al., 1982) is associated with adult T cell leukemia (ATL) (Yoshida et al., 1984). HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP) (Gessain et al., 1985; Osame et al., 1986), and other specific diseases (Mochizuki et al., 1992). Unlike other retroviruses, the HTLV-1 genome is known to be stable, and thus is highly conserved among inhabitants of endemic areas who are distantly related (Gessain et al., 1992). In contrast to these findings, we (Saito et al., 1995) and others (Kira et al., 1994; Niewiesk et al., 1994) previously reported highly frequent variations of the pX sequence within infected individuals including HAM/TSP patients and asymptomatic carriers. In these infected individuals, the sites of provirus integration were almost random, indicating proviral spread through viral replication.

The pX region has two overlapping frames to encode Tax and Rex proteins (Nagashima et al., 1986), both of which are essential for efficient viral replication (Seiki et al., 1986; Inoue et al., 1987; Yoshida and Seiki, 1987). Tax protein is a transcriptional activator required for viral gene expression (Sodroski et al., 1984; Fujisawa et al., 1985; Felber et al., 1985). Therefore, inactive Tax mutants would be unable to express the viral antigens; thus cells infected with such mutants could accumulate by escaping from the host immune response. Furthermore, a peptide of Tax protein, which localizes mainly in the nucleus (Kiyokawa et al., 1985), was identified as the most efficient epitope for cytotoxic T cells against HTLV-1 (Jacobson et al., 1990; Kannagi et al., 1992; Elovaara et al., 1993). Therefore, variations of the Tax sequence might result in accumulation of cells carrying the mutated proviruses.

To evaluate the specificity of the high frequency of variations in the pX region, we analyzed sequence variations in a non-coding sequence, the long terminal repeat (LTR), and here report that the LTR sequence is also highly variable within infected individuals. The frequency of the mutations in LTR was similar to that observed in the pX region in each patient with HAM/TSP and carriers, clearly indicating that sequence variations are not restricted to the Tax-coding sequence.
Results

Frequent mutations in LTR

To compare the frequency of sequence variations in two different regions in HTLV-1 genome, we chose the long terminal repeat (LTR) to analyze the sequence variations. The patients we used here were the same as used previously for the analysis of the pX sequence (Saito et al., 1995), so that we can compare the mutation rates in the LTR sequences directly with those in the pX region. Although LTR contains many regulatory elements that play critical roles in viral replication and gene expression, it does not encode any protein.

For sequence analysis of LTR, cellular DNA was isolated from peripheral blood lymphocytes from patients with HAM/TSP (cases 1-3 and 5), and the LTR sequence was amplified by PCR. The primers indicated in Figure 1 were used for amplification of the 738 base pairs (bp) corresponding to 9 to 746 of the HTLV-1 LTR, including almost the entire sequence of the LTR. The amplified sequences were cloned into Bluescript and ten independent clones from each sample were sequenced.

The sequences determined were compared with that of ATK-1, a prototype sequence of HTLV-1 (Seiki et al., 1983), and the nucleotides that differed from ATK-1 are summarized in Figure 2. In all clones analyzed, nucleotides at 128, 209, 210 and 316 were different from ATK-1 in all four cases. These observations suggest that the viruses originally infected into these individuals had such base substitutions and had no mutation at these bases during the replication within the individuals. Therefore, it is suggested that these variants are endemic in the southern part of Kyushu island where all these patients originated. Such substitutions probably present in the original viruses were also seen at position 146 in case 1, at 277 and 292 in case 2, at 45 and 713 in case 3, and at 660 in case 5. These original substitutions were rather frequent in LTR when compared with the sequences of the pX region.

Setting apart from these original differences of the sequence, sequence variations were observed in a significant number of clones (Figure 2). Clones marked with gray shading are the dominant species in each specimen and these were the same as ATK-1. The rest of the clones had a mutation or mutations and the sites and nature of the mutations did not show any particular pattern. Therefore, these base substitutions appeared to be random. The frequencies of these variant clones are summarized in Table 1. The frequency of the mutant clones in LTR was almost identical that in the pX sequence (Saito et al., 1995), clearly indicating that frequent mutation was not restricted to the pX sequence, but also taken place in LTR.

We also performed similar analyses on an ATL patient (ATL) and an asymptomatic carrier (AC case 1) and the results are shown in Figure 2 and Table 1. The frequencies of the sequence variations in the

![Figure 1](image_url)  
**Figure 1** Structure of LTR and distribution of the mutations. Locations of cis-acting elements are indicated by boxes including the 21 bp enhancer, Tax-responsive element-2 (TRE-2), TATA box, polyadenylation signal (AATAAA) and the Rex-responsive element (RXE). The positions of mutations in all samples are cumulatively indicated by vertical bars. The sequence of the primers (L-S and L-AS) for the PCR and their locations are also indicated. The nucleotide numbers correspond to those of prototype ATK-1 sequence (Seiki et al., 1983).
LTR were significantly lower than those in HAM/TSP patients (Table 1) and were again very similar to those observed in the pX sequence. As previously shown, both of these cases showed clonal expansion of HTLV-1-infected T cells in peripheral blood (Saito et al. 1995), thus the viral loads in these cases increased by clonal replication of HTLV-1-infected T cells, but not by viral replication. Therefore, these results clearly support further a conclusion that the high frequencies of base substitutions in the LTR are observed in cases in which the proviral doses increased by the viral replication and re-infection.

**Sites of frequent mutation in LTR**

When mutations were plotted on the LTR sequence, the mutations distributed almost randomly over the entire LTR and no clustering at particular regions was observed including essential and non-essential sequences for the LTR functions (Figure 1). For example, some mutations were observed in the elements that are important for efficient replication of HTLV-1 including the 21 bp enhancer and the Rex-responsive element. A mutation at 116 was in the 21 bp sequence and those at 347, 359, 387, 425, 449, 459, 479, 498, 512, 513, 517, 518, 528, 531 were in the Rex-responsive element. The 21 bp enhancer is essential for efficient transcription of the viral genome (Fujisawa et al., 1986; Saito et al., 1986; Shimotohno et al., 1986; Paskalis et al., 1986) and the R region forms a secondary structure, which is required for Rex binding and expression of the unsliced forms of the viral mRNA and the genomic form of HTLV-1 (Toyoshima et al., 1990). Many other mutations were also observed in the non-essential region for the LTR function. These frequencies of the base substitutions were not particularly different between these essential and non essential sequences. These findings suggest that frequent mutations in the LTR sequence are not conserved for preservation of the LTR activities. Such non-selective mutations in the LTR are similar to those observed in the pX region in which the mutations were not conservative for the function of Tax protein (Saito et al., 1995). Like the findings in the Tax-coding sequence, no identical mutation was detected in each specimen, clearly suggesting randomness of the mutations and incapability of mutants to undergo replication.

**Discussion**

In this study, we demonstrated that sequence variations of HTLV-1 are also frequently observed.
in LTR, a non-coding sequence, within infected individuals in which the proviruses showed a random integration. Furthermore, the frequency of the mutations in the LTR every 10,000 nucleotide (Table 1) was similar to that observed in the Tax-coding sequence in pX region (Saito et al., 1995) and much higher than those in an ATL patient who showed a clonal expansion of infected cells. Thus, high mutation frequencies observed previously in the HTLV-1 genome of HAM/TSP patients or carriers are not restricted to the Tax coding sequence.

The pX region has at least two overlapping genes, tax and rex, which are both essential for viral replication (Selki et al., 1986; Inoue et al., 1987). Thus, the region was mostly conserved (Tsunimoto et al., 1988; Shirabe et al., 1990). Because of this unique status of the pX region, very highly frequent base substitution in these overlapping genes was not expected since most of these mutations may inactivate one or both genes (Saito et al., 1995). Furthermore, the high base substitution was unexpected also since the HTLV-1 genome was highly conserved among infected people and among the endemic areas. One might postulate a specific or unique reason for the high rate of mutations in the Tax-coding sequence. One possibility is as follows: Tax is essential for the activation of transcription of the viral genome (Sodroski et al., 1984; Fujisawa et al., 1985; Felber et al., 1985) and some mutations inactivate its function; thus, cells infected with these variants would not express the viral antigens. This would result in accumulation of infected cells infected with these mutants since these cells would be able to escape from host immune rejection. Alternatively, Tax protein is the most potent antigen for induction of cytotoxic T lymphocytes (CTL), which play critical roles in rejection of infected cells in vivo (Jacobson et al., 1990; Kannagi et al., 1992; Elovaara et al., 1993; Parker et al., 1994). Therefore, it would be possible that mutation in the Tax coding sequence may alter the target amino acid sequence of the peptide or modify the peptide-forming capacity. These alterations would enable cells infected with these mutants to escape from immuno-surveillance mediated through CTL. This would also lead to accumulation of mutant proviruses. If either or both possibilities are the case, it would be expected that such high frequency of the sequence variations in infected individuals is specific to the Tax coding sequence. However, highly frequent variations in LTR observed in this paper do not support such possibilities discussed above. On the other hand, one may argue that mutations in the LTR suppress the viral gene expression, thus the cells infected with the mutants could escape the immune response of the host. As the result of this selection, specific mutations in the LTR sequence could be accumulated. This possibility is however unlikely since no particular mutation dominated in the LTR, moreover it is not expected that all random mutations in the LTR reduced the viral gene expression equally.

Furthermore, the frequencies of mutations in the LTR in infected individuals were similar to those observed in the Tax coding sequence. Thus, frequent mutations are not restricted to the Tax coding sequence and the frequency of mutations was very similar to that in the Tax-coding sequence. Furthermore, the sites of the mutations distributed all over the LTR sequence and were not restricted to conserve the functions of the LTR. These properties of the mutations were also similar to those in the Tax-coding sequence. Therefore, these observations suggest that frequent and random mutations would be seen in other sequences in the HTLV-1 genome.

In the case of ATL or some carriers in which infected cells are clonally expanded, such high frequency of mutations was not observed. These results in turn suggest that frequent mutations may occur during viral replication. However, if the mutant progenies could replicate with equal efficiency to the wild type virus, the HTLV-1 genome should be highly variable and should not be conserved among inhabitants of endemic areas who are distantly related. Therefore, we propose that the vast majority of mutants would be defective or less efficient in their replication.

The frequency of sequence variations in the pX and LTR sequences was six to 10 bases every 10,000 base (Table 1). These mutations should not be the artifact due to the replication errors of Taq polymerase, since the rate was much lower in an ATL case that showed a clonal integration of the proviruses. If the mutations take place randomly as we propose, these numbers imply that every viral progeny has six to 10 mutations per genome. Thus, the vast majority of the viral progenies would be inactive or less competent in their replication. PCR of proviruses in peripheral blood lymphocytes probably amplifies these variants, most of which are not inherited by the progenies and are thus not transmitted to recipients. The contribution of these viral progenies, which are defective in replication, to the pathogenesis or viral latency within infected individuals might be an interesting subject of investigation.

Materials and methods

Patients

The four patients with a clinical diagnosis of HAM/TSP (HAM cases 1–3, 5) were the same as those previously analyzed for the pX sequence (Saito et al., 1995), and the pathological characteristics of these patients were previously described (Umehara et al., 1993). Cases 1, 3 and 5 were from Kagoshima and case 2 was from Miyazaki, both of which are endemic areas of HTLV-1 infection in Japan.
Cellular DNA was prepared from peripheral blood lymphocytes (PBL) according to Maniatis et al. (1982).

**PCR**
PCR was carried out similarly to the method previously described (Saito et al. 1995). Briefly, 1 µg cellular DNA was used to amplify the pX sequence by 35 cycles of PCR using reaction buffer containing 50 mM KCl, 10 mM Tris (pH 8.3), 1.5 mM MgCl2, 0.2 mM deoxynucleotide triphosphate, 2.5 unit Taq polymerase (Takara, Japan) and 1 µM primers. Primers used were L-S and L-AS and the sequences were 5'-ACCATAGGCCCCAAAATATCCCCT-3' (L-S, position 9–31) and 5'-AATTCTTCCTCTGAGAGTGCTATAG-3' (L-AS, position 772–746), respectively. Each PCR cycle was consisted of denaturation at 95°C for 60 s, annealing at 56°C for 75 s and extension at 72°C for 90 s and the final cycle with extension at 72°C for 10 min. After the PCR reaction, 5 µl aliquots of the reaction mixtures were analyzed by electrophoresis in 2% Nusieve (FMC Bioproducts, ME) agarose gel.

**Cloning and sequencing**
Amplified DNA products with the expected size were excised from the agarose gel and the DNA was precipitated with ethanol after phenol extraction. The isolated DNA was subcloned into pBluescript KS' and sequenced using a Cycle Sequencing Kit (Applied Biosystems, CA) and an automatic sequencer (373A DNA Sequencer, Applied Biosystems).

**Acknowledgements**
This work was supported in part by a special grant for Advanced Research on Cancer from the Ministry of Education, Culture and Science of Japan.

**References**


