Human endogenous retroviruses (HERVs) make up a substantial part of the human genome. HERVs and solitary long terminal repeats (solo LTRs) are usually flanked by 4–6 nt short direct repeats through the well-known mechanism of their integration. A number of solo LTRs flanked by unusually long direct repeats were detected in the human genome. These unusual structures might be a product of an alternative virus insertion mechanism.

In this report we describe unusually long TSDs flanking some of the long terminal repeats of human endogenous retrovirus K in the human genome.

The LTR insertions with unusually long TSDs were first identified in a library of human-specific LTR HERV-K integrations, which was obtained using a new method described by us recently (Mamedov et al., 2002). Human specificity of individual LTRs was confirmed by comparison of PCR amplification products for human and great ape DNA samples. This stage was performed with primers targeted at unique sequences flanking the LTR integration sites at the 5’ and 3’ ends. The amplification product of a locus with an LTR insert is generally about 960 bp (the LTR length) longer than that derived from an orthologous locus lacking the LTR. However, this was not the case for two human-specific LTRs. For the fragments amplified from the LTR AC006035 insert located on human chromosome 7 and an orthologous site in chimpanzee lacking the LTR, the difference in length was approximately 250 bp higher than expected (i.e. ~1210 bp instead of 960 bp; see Fig. 1 for details). Here a standard PCR protocol with the Gibco PCR Reagents System and primers 5’-AACCACGTGAATTACACTTTCCTCA-3’ (forward) and 5’-GTCCAGTTA-GACCCCTCAACTTAG-3’ (reverse) was used. The samples were amplified for 28 cycles at 94 °C for 20 s, 65 °C for 20 s, and 72 °C for 40 s. A similar peculiar deviation was also observed for the LTR AC009132 located on chromosome 16. In this case the amplification profile was 28 cycles of 94 °C for 20 s, 53 °C for 20 s and 72 °C for 40 s. A similar peculiar deviation was also observed for the LTR AC009132 located on chromosome 16. In this case the amplification profile was 28 cycles of 94 °C for 20 s, 53 °C for 20 s and 72 °C for 40 s. A similar peculiar deviation was also observed for the LTR AC009132 located on chromosome 16. In this case the amplification profile was 28 cycles of 94 °C for 20 s, 53 °C for 20 s and 72 °C for 40 s. A similar peculiar deviation was also observed for the LTR AC009132 located on chromosome 16.

The sequences of both the LTR-containing loci were analyzed and no SDRs flanking the LTRs were detected. Sequences of 500 bp flanking these two LTRs in the human genome were subjected to a plot analysis using the VectorNTI program, which revealed unusually long TSDs of 250 and 61 bp in length for the LTRs AC006035 and AC009132, respectively. A long TSD flanking the former
LTR (AC006035) was nearly perfect except for a deletion of one T in a 13T track and a single G/T substitution (for alignments of TSDs, see data available at http://humgen.sibcb.ras.ru/supplement/suppl.html). It consisted of part of AluJo, a short non-repetitive genomic sequence, and part of MER67B. The LTR AC009132 was integrated into an L1MB3 element. Its long TSD was formed by a duplicated fragment corresponding to positions 6044–6106 of the L1 consensus sequence. This duplication was completely perfect.

Both the LTRs were specific to the human genome and belonged to the youngest subfamily amplified late in primate evolution (Table 1). Their pre-integration states were characterized by shorter PCR fragments amplified from orthologous loci in chimpanzee genomic DNA. The absence of the LTR insertions in orthologous loci of chimpanzee and gorilla suggested that the LTRs were integrated after the divergence of the chimpanzee and human lineages, which occurred 5–6 million years ago. A low divergence of TSD sequences in human DNA implies recent duplication events. To evaluate the duplication time, PCR products obtained from chimpanzee and gorilla genomic DNAs from the locus orthologous to the LTR AC009132 were cloned using a PCR products T-easy cloning kit (Promega) and sequenced using an Applied Biosystems 373 automatic DNA sequencer. An analysis of the obtained sequences and sequenced fragments of the chimpanzee genome revealed that the orthologous loci of the two closely related primate genomes lacked the duplication (see supplementary data), thus confirming that this duplication did occur after the divergence of the human and chimpanzee lineages.

A search through the non-redundant and high throughput genome sequences databases of GenBank for HERV-K LTRs using the NCBI BLAST program (http://www.ncbi.nlm.nih.gov/BLAST) and the RepeatMasker program by A. Smith and P. Green (http://ftp.genome.washington.edu) also revealed other long TSDs. We analysed SDRs of about 400 solo HERV-K LTRs and 35 proviruses randomly picked up from GenBank. Most of the analysed sequences appeared to have perfect SDRs of the usual length (4–6 bp). Some of the SDRs analysed were not perfect and contained one or two nucleotide substitutions. Some of the LTRs analysed were truncated at their 5’ or 3’ ends, which prevented

**Table 1. Location of LTRs flanked by unusually long TSDs**

<table>
<thead>
<tr>
<th>Accession no.</th>
<th>Chromosome</th>
<th>Position of LTR in the deposited sequence</th>
<th>TSD length (bp)</th>
<th>LTR family*</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC006035</td>
<td>7p21.2</td>
<td>180578–181546</td>
<td>250</td>
<td>II-L4</td>
</tr>
<tr>
<td>AC009132</td>
<td>16q22.3</td>
<td>153029–153985</td>
<td>61</td>
<td>II-L4</td>
</tr>
<tr>
<td>AC108063</td>
<td>4p15.32</td>
<td>49378–50352</td>
<td>96</td>
<td>II-V</td>
</tr>
<tr>
<td>Z95704</td>
<td>4p16.3</td>
<td>30541–31496</td>
<td>262</td>
<td>I-Y</td>
</tr>
</tbody>
</table>

*According to the classification reported in Lebedev et al. (2000). Timing of the insertion of the corresponding master genes was estimated as to be 4, 31 and 41 million years ago for II-L4, II-V and I-Y, respectively.
identification of SDRs. However, some full-length LTRs were flanked by SDRs of unusual size. Sequences of 500 bp flanking these LTRs were subjected to plot analysis. As a result, two more LTRs flanked by 96 and 262 bp long TSDs (see Table 1) were detected (AC108063 and Z95704, respectively). The former LTR was integrated into a unique genomic sequence and its direct repeats were more diverged than the two described above. The LTR sequence allowed us to assign it to an intermediate-age subfamily (Table 1) that was transpositionally active before the divergence of the gorilla and human/chimpanzee lineages.

To examine whether the duplication arose due to the LTR insertion, PCR assays of various primate genomic DNAs were performed using primers 5’-TCATAGATAGAAA-CAAGGTCCTCCT-3’ and 5’-CCCCCAGTGGCTCGTACTAGAG-3’ targeted at the LTR AC108063 flanks. The amplification was performed for 30 cycles at 94 °C for 20 s, 63 °C for 20 s and 72 °C for 40 s. The PCR fragments corresponding to the LTR AC108063 insertion site with 96 bp long TSDs were detected in DNA samples from human, chimpanzee and gorilla. With several gibbon DNA samples presumably lacking the LTR, the corresponding PCR fragments were shorter. A PCR product derived from the gibbon genome of a Hylobates lar individual was cloned and sequenced (accession no. AY536064), confirming the absence of duplication in the ‘pre-integration’ site.

The LTR Z95704 belongs to an old subfamily and is targeted at the LTR AC108063 flanks. The 96 bp long TSDs were detected in DNA samples from human, chimpanzee and gorilla. With several gibbon DNA samples we didn’t detect duplication (see Table 1) were detected (AC108063 and Z95704, respectively). The former LTR was integrated into a unique genomic sequence and its direct repeats were more diverged than the two described above. The LTR sequence allowed us to assign it to an intermediate-age subfamily (Table 1) that was transpositionally active before the divergence of the gorilla and human/chimpanzee lineages.

The LTR Z95704 belongs to an old subfamily and is integrated into an L1 retrotransponson. Its direct repeats are highly (~8%) diverged. The sequences of the sites corresponding to the human LTRs AC108063 and Z95704, taken from the recently published chimpanzee genome (available at http://genome.ucsc.edu), revealed similar duplications flanking the LTRs.

So far only a few examples of long TSDs flanking other transposable elements in the sites of their integration are known, among them a 214 bp long TSD produced due to the integration of a human L1 (Feng et al., 1996), 952 bp long repeats surrounding an IS476 element in a recombinant plasmid (Chen et al., 1999) and 82 bp long repeats flanking the intracisternal A particle (IAP) in the mouse genome (Tanaka & Ishihara, 1995). The long interspersed elements (LINE) retrotransposition mechanism differs from that of retroviruses and includes nicking of the target DNA. Feng et al. (1996) suggested that the formation of such long TSDs was due to peculiarities of helicase activity at the site of integration. In the case of the bacterial insertion sequence (IS) element, the duplication was suggested to be due to plasmid recombination at the site of insertion (Chen et al., 1999). Similar to retroviruses, the integrated form of an IAP element has gag, pol and env genes between two LTRs and uses the same mechanism of retrotransposition. In this context, the closest example is an 82 bp long TSD flanking the de novo integration of an IAP element into the IL-3 gene of myeloid leukaemia cells, generated by whole-body irradiation of mice. Tanaka & Ishihara (1995) suggested that such a long target site duplication was somehow associated with the impact of radiation, which might cause rearrangements or induce an unusual mechanism of retrotransposition.

It has been hypothesized (Morrish et al., 2002) that retrotransposons sometimes take part in the repair of DNA breaks caused by various reasons. A similar repair process coupled with HERV-K retrotransposition might form the LTRs described in this work. However, real forces and participants responsible for the integration-associated duplication remain unclear. Hopefully, they will be identified when other mammalian genomes are sequenced and more examples of such rare events are available. The study of the mechanism of the long TSD formation will give us a deeper insight into retroviral transposition and interactions of endogenous retroviruses with the host genome.

**Acknowledgements**

We would like to express our gratitude to Dr B. Glotov for helpful and constructive discussion and help in manuscript preparation. This work was supported by INTAS-2001-0759, the Russian Foundation for Basic Research 02-04-48614 and 2006.20034 grants, as well as by the Physico-Chemical Biological Program of the Russian Academy of Sciences.

**References**


