Spliced human endogenous retroviral HERV-H env transcripts in T-cell leukaemia cell lines and normal leukocytes: alternative splicing pattern of HERV-H transcripts

Mats Lindeskog1 and Jonas Blomberg2

1 Department of Medical Microbiology, Section of Virology, Lund University, Solvegatan 23, S-22362 Lund, Sweden
2 Department of Infectious Diseases and Clinical Microbiology, Section of Virology, Uppsala University, Dag Hammarskjölds väg 17, S-75237 Uppsala, Sweden

The majority of human endogenous retroviral HERV-H elements in the human genome have large deletions in pol and lack most of env, 5–10% are more or less complete with a potentially immunosuppressive transmembrane protein-encoding env region. Spliced HERV-H env transcripts were detected in T-cell leukaemia cell lines and lymphocytes from healthy blood donors by using RT–PCR. The transcripts all contained a splice donor in the leader region downstream from the primer-binding site and a previously unreported splice acceptor in the integrase-encoding region of pol, absent in the HERV-H deletion elements. In singly spliced transcripts the leader and integrase regions were joined directly whereas in multiply spliced transcripts they were joined with an alternative exon from the protease-encoding region located between the two regions. env transcripts from three different HERV-H elements were identified; one element similar to a HERV-H consensus sequence was primarily amplified from the T-cell leukaemia cell lines and two other more defective elements were amplified from normal lymphocytes. One of these elements was shown to be a reintegrated spliced transcript where the protease and integrase regions were joined, removing most of pol but leaving gag intact. Other spliced transcripts, joining the protease region and the 3′-LTR, were also amplified. The fact that HERV-H elements with an intact env splice acceptor also use the splice sites in the protease-encoding region suggests that this unusual multiple splice pattern could have a biological function in the intact HERV-H.

Introduction

The human genome contains thousands of endogenous retrovirus-like elements (HERVs; reviewed by Wilkinson et al., 1994). Many of these are expressed as RNA in both normal and malignant tissue. A relatively high level of HERV RNA is generally seen in placenta cells (Rabson et al., 1983; Kato et al., 1987; Franklin et al., 1988; Johansen et al., 1989) whereas the transcript level in other normal tissues, e.g. PBMCs, is low and often detectable only by sensitive methods such as PCR (Medstrand et al., 1992; Lindeskog et al., 1993).

The significance of HERV RNA expression is still far from clear but there is accumulating evidence that HERV protein expression occurs in normal tissue, especially in the placenta (Kitamura et al., 1994; Venables et al., 1995), and in tumour cell lines (Löwer et al., 1995; Sauter et al., 1995). It has been suggested that the env-encoded surface (SU) and transmembrane (TM) proteins have the potential to participate in normal cell function. SU proteins may have a fusogenic function in the creation of the syncytiotrophoblast layer of the placenta (Rabson et al., 1983). TM proteins containing sequences similar to the immunosuppressive peptide in murine leukaemia virus (MLV) p15E protein (Cianciolo et al., 1985) may suppress maternal immunological rejection of the foetus (Venables et al., 1995) and may also be involved in tumour development (Cianciolo et al., 1984).

In most exogenous retroviruses, the gag and pol products are translated from full-length mRNA whereas the env products and, in certain retroviruses, regulatory proteins are translated from spliced subgenomic mRNAs. A balance between spliced and unspliced RNA (usually occurring in equal amounts in mammalian and avian type C viruses) is vital for the retroviral
life cycle (Katz & Skalka, 1990) and may be regulated by different mechanisms (Coffin, 1985). The main splice donor (SD) site is typically located in the leader region between the 5'-LTR and gag, whereas the splice acceptor (SA) site for the env transcripts is within the terminal part of pol or in the region between pol and env. In lentiviruses and human T-lymphotropic viruses (HTLV), the regulatory proteins are translated from multiply spliced transcripts, using several alternative splice sites in the pol/env border region and in the env/5'-LTR border region (Muesing et al., 1985; Seiki et al., 1985). Temporal regulation of alternative splicing is seen in human immunodeficiency virus (HIV) infection, where the multiply spliced regulatory protein mRNAs are produced early in the infection cycle and the gag, pol and env mRNAs (unspliced and singly spliced, respectively) appear at a later stage. This is facilitated by the Rev protein, which mediates cytoplasmic transport of incompletely spliced transcripts (Felber et al., 1989).

The frequency and pattern of RNA splicing varies among the different types of HERVs but an SD site in the leader region seems to be ubiquitously used in spliced subgenomic mRNAs. Spliced env transcripts are commonly observed, e.g. in HERV-E and ERV3 where they constitute the main type of transcripts (Rabson et al., 1983; Kato et al., 1987). In addition to the 'normal', singly spliced env transcripts, more complex splice patterns are sometimes observed. In the HERV-K family, a subset of elements produce a 1.8 kb multiply spliced mRNA containing the mRNAs of HIV and HTLV regulatory proteins, both in splicing pattern and amino acid composition of its translation product (Lower et al., 1993). In HERV-H, a cluster of SA sites is located in the gag/pro border region, resulting in transcripts where gag is removed and pol retained (Johansen et al., 1989; Wilkinson et al., 1990).

We have previously identified the presence of HERV-H transcripts containing gag and env sequences in normal lymphocytes by RT-PCR (Medstrand et al., 1992; Lindeskog et al., 1993). The majority of HERV-H elements (800–900 copies) have several large deletions in the pol region and lack almost the entire env region, whereas a fraction (50–100 copies) are full-length elements (Hirose et al., 1993; Wilkinson et al., 1993). The HERV-H env region contains sequences that are similar to the coding region for the immunosuppressive peptide of the MLV TM protein and are present in a 2.5 kb mRNA in placenta and lung cells (Lindeskog et al., 1993). env-containing HERV-H mRNAs of different lengths have also been detected in testicular tumour cell lines (Hirose et al., 1993).

Since all retroviral envelope proteins seem to be translated from spliced subgenomic mRNAs, greater knowledge of the patterns and dynamics of HERV RNA splicing would be valuable in the study of HERV env-encoded proteins. In this paper we report the presence of spliced HERV-H env transcripts in T-cell leukaemia cell lines and blood donor lymphocytes. The splice sites in the protease region are also used in the presence of a functional env SA site and contribute to alternative splicing of a number of subgenomic mRNAs.

**Table 1. Synthetic oligonucleotides used in the experiments**

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Sequence*</th>
<th>Origin†</th>
<th>Position‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBHIS</td>
<td>GACATTTGGTGCTGCTGACTCAGAT</td>
<td>(RTVLH2)</td>
<td>447–470</td>
</tr>
<tr>
<td>RHenR1</td>
<td>AAGTACTGGTCTCTTTTACATCGGG</td>
<td>(RGH2)</td>
<td>5804–5826 (as)</td>
</tr>
<tr>
<td>RHenR2</td>
<td>AACAAAGACTGATCGTCCAGTCTAG</td>
<td>D1</td>
<td>252–275 (as)</td>
</tr>
<tr>
<td>HPer</td>
<td>GATGAAAGGTAGTATTAGCCTCA</td>
<td>D1</td>
<td>513–536 (as)</td>
</tr>
<tr>
<td>HSU1R</td>
<td>AAGTAAAAGAAATGACCTATCGAGG</td>
<td>D1</td>
<td>836–860 (as)</td>
</tr>
<tr>
<td>HSU12</td>
<td>GGAAGAAAAGGAAAATCGAAAGCCA</td>
<td>RGH2</td>
<td>6492–6515 (as)</td>
</tr>
<tr>
<td>D1PR1</td>
<td>CGACCTCCTGGAATACCTCAAA</td>
<td>D1</td>
<td>200–222</td>
</tr>
<tr>
<td>D1PR2</td>
<td>AGCTGGCATCCTGATGCTCT</td>
<td>D1</td>
<td>184–205 (as)</td>
</tr>
<tr>
<td>D2PR1</td>
<td>ACCACAGCTTGGAGTAACTTCT</td>
<td>D2</td>
<td>199–221</td>
</tr>
<tr>
<td>D2PR2</td>
<td>TAAGATTTACCCAAGCTCGGTG</td>
<td>D2</td>
<td>200–222 (as)</td>
</tr>
<tr>
<td>C7PRL2</td>
<td>CAGAAGCTTACACGGACATCACA</td>
<td>(D3)</td>
<td>171–194</td>
</tr>
<tr>
<td>D3PRL</td>
<td>CATTGAGACTCCTTGGAGTTC</td>
<td>D3</td>
<td>201–223 (as)</td>
</tr>
<tr>
<td>HU3R</td>
<td>GTGGTGAAATGTCATCAGTAAG</td>
<td>RGH1</td>
<td>4534–4557 (as)</td>
</tr>
</tbody>
</table>

* I, Inosine.
† Origin of sequence. Previously published elements are RTVLH2 (Mager & Freeman, 1987), and RGH1 and RGH2 (Hirose et al., 1993). The parentheses denote that the oligonucleotides are not identical to the origin.
‡ Position in origin; as, antisense string.
(ATCC CCL 136); and glioma cells (Lindeskog et al., 1993). Total RNA was always treated with RNase-free DNase (Promega). Genomic DNA was extracted from one of the PBMC preparations using standard methods (Ausubel et al., 1987).

Oligonucleotide primers and probes (Table 1) were synthesized by Scandinavian Gene Synthesis or Pharmacia Biotech. Reverse transcription of RNA and PCR on cDNA and genomic DNA were done as previously described (Lindeskog et al., 1993), using a DNA Thermal Cycler (Perkin-Elmer-Cetus) and 50 ng of each primer at an annealing temperature of 55 °C.

PCR products were cloned in a plasmid vector using the TA Cloning Kit (Invitrogen) or the pT7 Blue T-Vector Kit (Novagen). Clones were sequenced in both directions with M13 primers or internal primers with the Sequenase kit, version 2.0 (USB). Sequencing reaction products were separated on 6% denaturing polyacrylamide gels. Computer sequence analyses were done with the PCGENE programs (Intelligenetics). The BLAST programs (Altschul et al., 1990) were used for searching the nucleotide and protein databases.

Southern blotting of PCR products was done on Hybond-N+ membranes (Amersham). Hybridization to [γ-32P]ATP-labelled oligonucleotide probes was done at 60 or 65 °C as described by Ausubel et al. (1987) using 0.5 pmol (1 μCi) of probe per ml. Washes were done in 5 × SSPE, 0.1% SDS at 65 or 70 °C for 20 min.

Multiple tissue Northern blot II (Clontech) was hybridized, as recommended by the manufacturer, at 45 °C using 0.5 pmol (1 μCi) of oligonucleotide probe per ml. Washes were done in 5 × SSPE, 0.1% SDS at 45 °C for 15 min.

Results

Amplification of spliced HERV-H env transcripts from lymphocytes and leukaemia cell lines

To study spliced sub-genomic HERV-H transcripts, we designed PCR primers from the histidine-tRNA primer-binding site (PBS) and a sequence from the integrase-encoding region of pol (amino acid motif WKGP). This sequence is conserved among most retroviruses and situated just downstream from the SA sites used in env transcripts of many retroviruses (Table 2). In the HERV-H family the sequence is located in a region (designated D) found only in full-length elements (Hirose et al., 1993; Mager & Freeman, 1995), so only spliced transcripts from elements containing the env region should be detected by this strategy. The primer pair PBHIS/RHENR1 was used in RT–PCR amplifications of RNA from the T-cell leukaemia cell lines H9 and MOLT4, lymphocytes from seven healthy blood donors, cultured embryonic lung cells, the lung carcinoma cell line A549, the rhabdomyosarcoma cell line RD and glioma cells. Amplification products were obtained from RNA from the T-cell leukaemia cell lines (250 and 350 bp) and all seven lymphocyte preparations (350 and 400 bp) (Fig. 1a) whereas RT–PCR on the other cell types did not produce bands visible by gel staining (data not shown). PCR on lymphocyte genomic DNA gave no products shorter than 2.3 kb (Fig. 1a). Amplification products of the three different sizes from H9 and two PBMC preparations (#3 and #5) were cloned separately into plasmid vectors and the insert nucleic acid sequences of at least five different clones obtained from each ligation, designated c4 (H9, 250 bp), c5 (H9, 350 bp), c14 (PBMC5, 350 bp), c25 (PBMC3, 350 bp) and c7 (PBMC5, 400 bp), were determined.

All c7 clones were derived from a spliced hybrid transcript between a HERV-H and a HERV-E element which will be discussed in a further paper. All other clones contained sequences that were derived from three different HERV-H genomic elements (here designated D1, D2 and D3) with 89–93% nucleotide sequence identity in the amplified regions (Fig. 2a). Deviations by less than 1% from a ‘consensus’ sequence of several clones are likely to have been generated by the process of RT–PCR.

The different fragment lengths corresponded to two different types of spliced transcripts, both utilizing the SD in the leader region downstream from the PBS and a previously unreported SA site in the D region (Fig. 2b). In the 250 bp D1 PCR product (represented by clone c4.2) these sites were joined directly. In the 350 bp multiply spliced PCR products of D1 (clone c5.11), D2 (clone c14.7) and D3 (clone c25.1), a

Table 2. Position of retroviral 3’ pol SA sites

<table>
<thead>
<tr>
<th>Virus</th>
<th>3’ IN†</th>
<th>SA position‡</th>
<th>mRNA ORF‡</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLV</td>
<td>WKGP</td>
<td>−159</td>
<td>env</td>
<td>Mann &amp; Baltimore (1985)</td>
</tr>
<tr>
<td>HTLV</td>
<td>WKGP</td>
<td>−24</td>
<td>rex, lax, env</td>
<td>Seiki et al. (1985)</td>
</tr>
<tr>
<td>HIV</td>
<td>WKGP</td>
<td>−19</td>
<td>vif</td>
<td>Muesing et al. (1985)</td>
</tr>
<tr>
<td>MMTV</td>
<td>WKGP</td>
<td>+59</td>
<td>env</td>
<td>Moore et al. (1987)</td>
</tr>
<tr>
<td>HERV-E</td>
<td>WKGP</td>
<td>−50§</td>
<td>env</td>
<td>‒</td>
</tr>
<tr>
<td>HERV-H</td>
<td>WTGP</td>
<td>−80</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Conserved amino acid sequence in the C-terminal part of integrase.
† Position of the SA site relative to the first base of the W codon (TGG) in the conserved 3’ integrase-encoding sequence.
‡ mRNA ORF using the SA site.
§ Our unpublished observation.
|| New SA site presented in this article.
The D2 element is a reintegrated spliced transcript

To detect possible spliced reintegrated origins of the cloned HERV-H transcripts, PCR was done on genomic lymphocyte DNA using protease primers from the three elements (D1PRL & D1PRR, D2PRL & D2PRR and C7PRL2 & D3PRR) in combination with PBHIS and integrase primers. No genomic leader/protease junctions were found but sequences containing the protease/integrase junction were detected with the D2PRL primer when used with either primers RHENR1 or RHENR2 (data not shown). The PBHIS/RHENR1 2.3 kb genomic amplification product corresponded in length to a HERV-H element in which the second intron (pol) is removed (Fig. 1a). To verify this, we cloned this fragment into a plasmid vector and partially sequenced five clones (g19 clones). The five clones were derived from a single genomic element that was identical to the c14 transcripts in the leader, protease and integrase regions (Fig. 2). The protease and integrase regions were joined at the same splice sites as in c14 but the gag ‘intron’ was intact.

The D1, D2 and D3 transcripts are differentially amplified

The presence of the different types of spliced transcripts from the three HERV-H elements in blood donor lymphocytes

Fig. 1. (a) Agarose (1.8%) gel electrophoresis of PBHIS/RHENV1 amplification products from 1 µg reverse transcribed total RNA (1/3 of the reaction volume) from two human T-cell leukaemia cell lines (MOLT4 and H9) and PBMC preparations from seven healthy blood donors and 1 µg genomic PBMC DNA. The positions of the HERV-H amplification products are indicated by arrowheads. On the left-hand side is molecular size marker VI (Boehringer Mannheim). (b) Autoradiograms of Southern blots of the amplification products hybridized to a HERV-H integrase probe (RHENV2), and element-specific protease probes (D1PRR, D2PRR and D3PRR).
HERV-H alternative splicing

**Fig. 2.** (a) Alignment of nucleotide sequences of cDNA PCR clones from PBHIS/RHENR1 amplifications (c5.11, c14.7 and c25.1), part of the genomic amplification product g19.2 and corresponding regions (positions in parentheses) from the full-length element RGH2 (Hirose et al., 1993). Identical positions are indicated with dashes and gaps with a dot. The leader region I, the protease-encoding region II and the integrase-encoding region III are joined in cDNA clones but separated by intervening sequences in the genomic sequences, except for clone g19.2, where regions II and III are joined. Splice sites in the genomic sequences are shown in lower-case letters and the bases conforming to the consensus of splice sites (Shapiro & Senapathy, 1987) are underlined. The number of similar PCR clones is shown next to the sequences with the number of differing nucleotide positions in parentheses. (b) Regions joined in spliced transcripts are shown with the corresponding clones above a full-length genomic element (positions of deletion regions A–E are boxed) and a spliced reintegrated element. The PCR-generated genomic clone g19.2 is also shown. PCR primers used in cloning are shown by arrows (↑) and splice sites are indicated by flags (▲).
and the T-cell leukaemia cell lines was determined by Southern blots of the PBHIS/RHENR1 amplification products and hybridization to a HERV-H integrase oligonucleotide probe RHNRE2 (located between the env SA site and RHNRE1) and the element-specific protease oligonucleotide probes D1PRR, D2PRR and D3PRR (Fig. 1b). RT–PCR with primer PBHIS and the HERV-H-specific integrase 3′-primer RHNRE2, and Southern blot hybridization of the amplification products to the protease probes gave very similar but generally more intense corresponding bands (data not shown).

Probe RHNRE2 detected both singly and multiply spliced HERV-H transcripts in all blood donor lymphocytes (250 and 350 bp bands, respectively) although the singly spliced transcript amplification products were only visible by gel staining in the leukaemia cell lines. The element-specific protease probes detected differences between the leukaemia cell lines and also between lymphocytes from different individuals. Element D1 spliced RNA was most strongly amplified from the two leukaemia cell lines whereas element D3 amplification products were almost absent from these cell lines. Elements D2 and D3 were differentially amplified from PBMCs from different blood donors, e.g. #5 is high in D2 transcripts but low in D3 transcripts whereas the opposite is the case in #6. Additional bands (530 bp), visible by gel staining only in the PBHIS/RHNRE2 amplifications, hybridized to probes RHNRE2 and D1PRR. Very weak signals were obtained from Southern blots of RT–PCR amplification products of RD and glioma cells with the probes RHNRE2 and D2PRR (350 bp) and from PBHIS/RHNRE2 amplification products of A549 with the probe D2PRR (data not shown).

Attempts to detect spliced HERV-H poly(A)+ env transcripts in different cell types by hybridization of the integrase probe RHNRE2 to Multiple tissue Northern blot II (spleen, thymus, prostate, testis, ovary, small intestine, colon and PBMCs) in duplicate were unsuccessful. Only a 5±5 kb band in the small intestine sample was observed (data not shown).

**Amplification of longer fragments of the env transcripts**

To amplify longer fragments of the spliced transcripts, primer PBHIS and a 3′-primer (HSUR2) from a region in the SU-encoding part of env (amino acid motif CWLCL, conserved in mammalian type C and type D exogenous retroviruses)
were used in RT–PCR on RNA from the H9 and MOLT4 cell lines and three blood donor PBMCs (#1, #2 and #4) (data not shown). Amplification products of the expected sizes (950 and 1050 bp) were obtained from H9, MOLT4, PBMC1 and PBMC2. Other products (350–500 bp) were also amplified. Southern blots of the amplification products were hybridized to the protease probes (D1PRR, D2PRR and D3PRR), the integrase probe RHENR2 and the env probe HSUR1 (located approximately 150 bp upstream of HSUR2). The distribution of the three HERV-H elements in the different cDNA amplification products was the same as in the PBHIS/RHENR1 RT–PCR. None of the shorter bands hybridized to any of the probes, making it probable that these bands were non-specific viral amplification products. RT–PCR with the primer PBHIS and HSUR1 or HPER (located just downstream from the end of the pol region) gave only bands of the expected size (data not shown).

The 950 and 1050 bp PBHIS/HSUR2 amplification products from H9 (c30) and PBMC1 (c31) (Fig. 2b) were cloned into a plasmid vector and clones hybridizing to the element-specific protease probes were sequenced (three clones from each element). For each of the elements the clones differed by no more than 3 bp from the consensus, which was used in sequence analyses. Clones hybridizing to the integrase probe RHENR2, but not to any of the protease probes (one clone from H9 and six from PBMC1) were partially sequenced. These were all element D1 amplification products of the singly spliced leader/integrase junction type.

**Primary structure of the spliced env transcripts**

None of the three HERV-H env elements have been described before. Nucleotide similarities of elements D1, D2 and D3 to the genomic element RGH1 (Hirose et al., 1993) in the 750 bp region downstream from the env SA site were 96.8, 92.0 and 91.5%, respectively. The inter-element similarities in the same region were: D1/D2, 94.0%; D1/D3, 91.9%; and D2/D3, 90.9%.

The D1 element was relatively close to the consensus sequence of ten structurally intact HERV-H elements in this region (1.7% difference after discarding positions without a clear consensus nucleotide and CpG dinucleotides; unpublished data) and had the open reading frames (ORFs) that are present in the consensus sequence: the last 272 bp of pol (ORF1), the first 191 bp of env (ORF3) and the 281 bp (ORF2) partially overlapping env (Fig. 3). The ORF2 putative translation product was not similar to any sequences in the protein databases. The D2 element (3.4% difference from the consensus sequence) had frame-shifts in ORF2 and env, whereas the D3 element (4.8% difference from the consensus sequence) had a frame-shift in the pol region and a stop codon in ORF2.

The integrase-encoding ORF1 was in the spliced transcripts joined to ORFs with initiation codons upstream from the splice site in the leader region (D1) or in the protease region (D1, Fig. 3 and D3). In the element D2 transcripts however, there were no corresponding initiation codons.

**Detection of protease/3′-LTR junction transcripts in lymphocytes**

To detect short multiply spliced transcripts using an SA site in the proximity of the 3′-LTR – as seen in regulatory protein transcripts of HIV, HTLV and HERV-K – a 3′-primer in the HERV-H U3 region (HU3R) was used with PBHIS in RT–PCR on leukaemia cell line and lymphocyte RNA (PBMC1–4). Multiple bands were amplified from all samples (Fig. 4a). The integrase probe RHENR2 and the env probe HSUR1 did not hybridize to any bands on a Southern blot of the PBHIS/HU3R amplification products, indicating the absence of SD sites in the pol/env border region. There were, however, bands hybridizing to the protease probe D1PRR in PBMC1 and -4 and the probe D3PRR in PBMC1, -2 and -3 (Fig. 4b). The 420 and 500 bp PCR products from PBMC1 were cloned into a plasmid vector (c33 clones). Three clones hybridizing to the D1PRR and D2PRR probes (c33.28, c33.42 and c33.109) were amplification
products of spliced transcripts derived from three different HERV-H elements, none of them D1 or D2. The leader region was joined to the protease region at three different SA sites (all described by Wilkinson et al., 1990), and the protease region to two different SA sites, 10 and 146 bp upstream from the 3′-LTR, respectively (Fig. 5). A BLAST search of the EST (expressed sequence tags) database found sequences with these protease/3′-LTR junctions (GenBank accession numbers T65683 and H39731) and also a leader/3′-LTR junction (AA121606). The clones hybridizing to D3PRR had the same leader/protease sequence as the D3 env transcripts. In the 420 bp fragment (c33.6 and c33.14), the protease region was joined to an SA site 6 bp downstream from the beginning of a HERV-H LTR. The 500 bp fragment (c33.16) had an additional 72 bp region from another HERV-H LTR (SA at the end of the U3 region and SD in the R region) spliced in between the protease and LTR regions. We did not determine which of the two LTRs belonged to the D3 element (the other one being situated within or downstream from the D3 element). There were no significantly long ORFs in any of the sequenced leader/protease/LTR clones.

**Discussion**

In this study, we have shown that spliced env transcripts from three different HERV-H elements (D1, D2 and D3) are present in normal blood donor lymphocytes. We observed a difference in the spliced env transcript levels of the three elements between the T-cell leukaemia cell lines and lymphocyte preparations. Element D1 was primarily amplified from the leukaemia cell lines whereas element D3 could only be amplified from normal lymphocytes. A difference among blood donors was also apparent: the relative amounts of D2 and D3 amplification products varied considerably. Whether this reflects different transcription levels or a differential utilization of splice sites is unknown. Individual differences in the transcript levels of different class II human MMTV-like (HML) endogenous retroviral subfamilies have been shown in lymphocytes from blood donors (Andersson et al., 1996) whereas a single point mutation can completely alter the extent to which a certain splice site is used (Katz & Skalka, 1990). The non-lymphocyte cell types tested contained much fewer amplifiable transcripts of these HERV-H elements, which may indicate lymphocyte-specific transcription or splicing. HERV-H LTR promoter activity has been found to vary between different cell types (Feuchter & Mager, 1990) and tissue-specific splicing has been found in the troponin T gene (Breitbart et al., 1987). The level of these transcripts seems to be too low to be detectable in Northern blots even in lymphocyte mRNA or there could be a deficient polyadenylation of the transcripts.

There have been reports of antigens with cross-reactivity to MLV TM protein (p15E), e.g. in human tumour cell lines and mitogen-treated lymphocytes (Cianciolo et al., 1984), and such
Antigens have been shown to be associated with immunosuppressive activity (Cianciolo et al., 1981; Wang et al., 1986). Since HERV-H TM sequences are quite similar to immunosuppressive sequences of infectious mammalian type C retroviruses, they are strong candidates for the origin of these antigens. However, most of the approximately 100 structurally intact HERV-H elements have probably lost the env ORF by mutations since the time of entry of infectious proviruses in the germ line. The endogenization of HERV-H in primates is at least 40 million years old (Mager & Freeman, 1995). The presence of stop codons was shown in the published HERV-H env sequences (Hirose et al., 1993) and in eight different genomic PCR-generated HERV-H env sequences, of which only two have an ORF in the 400 bp SU/TM region (unpublished data). Interestingly, one of these (g10.34), which has an ORF throughout the whole of the env region and codes for the minimal peptide for immunosuppression LQNRRGLDLL (Ruegg et al., 1989), is quite similar to the D1 sequence in the pol/env region (97.7% identity) (unpublished data). It is possible that this D1 element, which is very similar to a structurally intact HERV-H consensus sequence, belongs to a subset of HERV-H elements capable of producing immunosuppressive-peptide-containing envelope proteins.

The utilization of an SA site in the terminal part of the pol region in env transcripts by structurally intact HERV-H elements is analogous to the situation in most infectious retroviruses. The fact that the same SA site is used by three distinctly different elements suggests that this is the original SA site for spliced env transcripts in HERV-H. The same SA site is also used in two EST clones (GenBank accession numbers T67812 and R38435), which are singly and multiply spliced transcripts, probably from a single HERV-H element (unpublished observation).

The splice sites in the protease region have been proposed to be cryptic sites activated after the loss of the original env SA site (Goodchild et al., 1995). However, our finding that elements with a functional env SA site also use protease splice sites supports the hypothesis that this unusual splice pattern has a biological function in the infectious HERV-H progenitor. There are several alternative SA sites in the protease region (Wilkinson et al., 1990) that different HERV-H elements may use to varying extents. In contrast, the protease SD site and the integrase SA site are the same in all identified protease/integrase junctions, indicating that this junction could be important in the biological functions of these transcripts. The association of the terminal integrase-encoding region ORF with a start codon in the protease exon, as seen in transcripts from the D1 and D3 elements, could produce protease/integrase fusion proteins. These kinds of proteins could have regulatory functions, using the unspecific nucleic acid-binding properties of the terminal part of the integrase protein (Katz & Skalka, 1994). Another possible explanation of this splice pattern could be the existence of distinct, incompletely spliced, gag and pol transcripts retaining either of the two ‘introns’, which contain most of gag and pol regions, respectively. The human foamy virus pol gene has been shown to be expressed as a Pro–Pol polyprotein (Löchelt & Flügel, 1996) transcribed from a spliced pol transcript (Yu et al., 1996).

Clone g19 represents a new type of reintegrated spliced transcript, where the gag region is intact and the pol region is removed. Transcripts from the reintegrated element (D2) are further spliced removing the gag region (clone c14). It has recently been shown that spliced HERV-H transcripts with gag removed and pol retained have reintegrated in the genome both by viral retrotransposition and as processed pseudogenes (Goodchild et al., 1995). The mechanism of D2 element reintegration is unknown but since the packaging signal, needed for efficient packaging and reintegration of genomic retroviral RNA (Mann et al., 1983), is located downstream from the leader SD site in murine retroviruses (reviewed by Linial & Miller, 1990), these kinds of reintegrated spliced transcripts may be more common than reintegrated elements lacking the gag region.

The region between the pol and env ORFs is approximately 300 bp in all sequenced structurally intact HERV-H elements. The short ORF2 begins 100 bp upstream from the env ORF and overlaps with env by 180 bp in another reading frame. This atypical organization may be the result of frame-shifts or other mutations at an early stage in HERV-H before the amplification of these elements in the genome. This could account for the fact that the CWLCL motif is located much closer to the N terminus of the putative Env protein than in those proteins from exogenous type C and D viruses. When PCR was performed with primers from the PBS, the env region, and the region between pol and env no further SA sites were found. A search for SD sites in this region using PCR primers from the PBS and the LTR, which would detect junctions with SA sites upstream from the 3′-LTR as seen in HIV and HTLV, was negative. However, the latter PCR amplified HERV-H transcripts where the usual protease SD site was joined to SA sites either upstream from or within the 3′-LTR. Apart from the D2 element, where the protease/integrase junction is fixed in the genome, only D3, not D1, spliced to the LTR. The other examples of this splice pattern (including two EST sequences) were derived from five different HERV-H elements, which may lack the env SA site and thus be more prone to using LTR SAs. None of the protease/LTR junction transcripts contained ORFs so the significance of this splicing is uncertain but could be a way of producing short gag transcripts where both pol and env regions are excised. Leader/3′-LTR junctions are also present in HERV-H transcripts as seen in one EST clone.

Alternative splicing in HERV-H is illustrated both by the D1 element and the EST sequences mentioned above with leader/protease or leader/integrase junctions and by the D3 element with protease/integrase or protease/LTR junctions. The absence of D1 protease/LTR and D3 leader/integrase junctions could reflect unique affinities between different splice-site pairs (Breitbart et al., 1987). The ‘extra’ PCR fragments...
hybridizing to probes C5PRR and RHENR2 are probably another example of the use of alternative splice sites by transcripts from element D1 or a similar element. The alternative inclusion of the protease exon in both env transcripts and transcripts utilizing SA sites near the 3'–LTR seems to be a general feature of spliced HERV-H transcripts.

A pattern of complex splicing of HERV RNA is evolving, at least in full-length elements which are often only a fraction of the elements within a HERV family and are the closest we can get to visualizing the functions of the infectious progenitor. This could mean that complex life cycles including multiple splicing and an associated regulation have been more common in other retrovirus genera such as lentiviruses, HTLV-like viruses and spumaviruses than has previously been thought to be the case. The multiple splicing of HERV-K transcripts and the putative regulatory protein could be a HERV analogue to the complex retrovirus genera mentioned above (Löwer et al., 1993). Whether the alternative splicing of HERV-H transcripts has had a similar function in infectious HERV-H remains to be determined.

We thank Elzierta Vincic for technical assistance. This work was in part supported by the European Commission project GENE-CT930019; funds at the Medical Faculty of Lund, Sweden; the Royal Physiographic Foundation; the Crafoord Foundation, Lund, Sweden; and the Österlund Foundation, Malmö, Sweden.

References


Received 21 March 1997; Accepted 21 May 1997