

ORIGINAL ARTICLE

Takashi Shiroma · Jun Sugimoto · Takaya Oda
Yoshihiro Jinno · Fuminori Kanaya

Search for active endogenous retroviruses: identification and characterization of a HERV-E gene that is expressed in the pancreas and thyroid

Received: June 1, 2001 / Accepted: July 23, 2001

Abstract To elucidate possible physiological functions of human endogenous retroviruses (HERVs) and their role in the pathogenesis of human diseases, we have developed a strategy to identify transcriptionally active HERV genes. By this approach, we have identified and isolated an active *HERV-E* gene that was mapped to 17q11. Although the gene was predicted to produce no intact viral particles due to the presence of stop codons, long open reading frames were retained in each *gag* and *pol* region. Northern blot analyses revealed in the pancreas (and thyroid) two major transcripts, 3.3 and 4.1 kb in size, associated with 500- to 600-nucleotide-longer minor bands. Preferential expression in pancreas and thyroid gland tissues may suggest a role for this gene in physiological functions common to these tissues.

Key words HERV-E · Structure · Expression · Pancreas · Thyroid · Northern hybridization

Introduction

Human retrotransposons or retroelements, prominent examples being the L1 and Alu families, have generally been considered “junk.” Recent progress has revealed their evolutionary role by demonstrating one after another their functions, such as the remodeling of the genome through transduction (Moran et al. 1999), formation of processed-type pseudogenes (Esnault et al. 2000), and possible participation in Alu retrotransposition (Boeke 1997). Most notable is the insight that retrotransposon reverse tran-

scriptase (RT) was recruited to acquire telomerase RT, which in turn enabled eukaryotes to evolve from prokaryotes (Kazazian 2000, Nakamura and Cech 1998). Furthermore, inherited diseases and cancers caused by transpositions of L1 and Alu have been well documented in hemophilia A (Kazazian et al. 1988), neurofibromatosis (Wallace et al. 1991), and breast cancer (Miki et al. 1996), to name a few examples (reviewed by Kazazian and Moran 1998 and by Miki 1998).

In contrast, no disease or physiological function was known to be related to human endogenous retroviruses (HERVs), one of three major groups of human retrotransposons, until recently. Syncytin, a member of the HERV-W family, is an exceptional example that has been shown to mediate cell fusion and to be possibly involved in directing placental morphogenesis (Mi et al. 2000). Compared to the retrotransposons L1 and Alu, HERVs are in a unique position in both evolutionary and biological respects: they are not so old as L1 is, and are between viruses and host genomes in their behaviors. This seems to result in dual effects of HERVs on the host genome. Like the mouse mammary tumor virus, IDDMK_{1,2}/HERV-K18 has been shown to code for a superantigen and it may be implicated in the pathogenesis of autoimmune diabetes (Conrad et al. 1997). On the other hand, the Friend murine leukemia virus-resistant gene, *FvI*, consists of the *gag* gene of the mouse endogenous retrovirus L family (Best et al. 1996).

To elucidate the physiological, pathological, and evolutionary roles of HERVs, we previously attempted to isolate active HERVs and HERV-related genes, and we developed a strategy to identify such genes (Sugimoto et al. 2001). In the present study, we applied this approach to the HERV-E family and identified an active *HERV-E* gene that expressed abundant transcripts in normal pancreas and thyroid tissues. Structural and Northern blot analyses suggest that the gene may play a role in physiological functions common to these gland tissues. In addition, while the activity of these retrotransposons is usually suppressed by DNA methylation and/or other mechanisms, this active *HERV-E* gene may serve as an appropriate control for investigations of such mechanisms.

T. Shiroma · J. Sugimoto · T. Oda · Y. Jinno (✉)
Department of Molecular Biology, University of the Ryukyus,
Faculty of Medicine, 207 Uehara, Nishihara, Okinawa 903-0215,
Japan
Tel. +81-98-895-1203; Fax +81-98-895-1438
e-mail: sirius@med.u-ryukyu.ac.jp

T. Shiroma · F. Kanaya
Department of Orthopedic Surgery, University of the Ryukyus,
Okinawa, Japan

Materials and methods

Nucleic acids

RNA was extracted from human placenta as described previously (Jinno et al. 1994). Poly (A)⁺RNA was isolated using Oligotex-dT30(Super) (Takara, Kyoto, Japan).

Identification of active *HERV-E* genes by reverse transcriptase-polymerase chain reaction (RT-PCR) and subsequent sequencing of cloned products

Active *HERV-E* genes were identified following the strategy employed by Sugimoto et al. (2001). Briefly, 4 µg of DNase I-treated poly(A)⁺RNA was reverse-transcribed in a 20-µl reaction volume, and 0.5 µl of the reverse transcription product or 50 ng of genomic DNA was amplified by PCR with 0.5 µM of primers ENV1 and ENV2 in 25 µl of standard reaction buffer for 26 cycles of 94°C, 30s/51°C, 30s/72°C, 30s extension. Two micro liters of the 10-fold-diluted first PCR product was subjected to 16 cycles of nested PCR using primers ENV1 and ENV3 in 50 µl of the same reaction buffer and under the same thermal cycle conditions as for the first PCR. A parallel reaction was carried out in the absence of reverse transcriptase to monitor the residual DNA. Amplified products were separated in a 4% polyacrylamide gel, and the 469-bp fragment was eluted and cloned to the pGEM-T easy vector (Promega, Madison, WI, USA). The *env* sequence was determined by reading both strands with an automated DNA sequencer, model IR4200L (Li-Cor, Lincoln, NE, USA). Primer sequences were as follows: ENV1, 5'-CGAGAGTTGG TGCCTACTGA; ENV2, 5'-CCYATARTACTGTATGA TCCT; ENV3, 5'-GCTCCTGAYATAATTCCART; and ENV3, 5'-CCACTCATYATCTTTCCARTT.

Isolation of cosmid clones carrying the active *HERV-E* gene and structural analysis

The full-length active *HERV-E* gene, *ERVE1*, was isolated from a human cosmid library using an *ERVE1*-specific antisense oligonucleotide probe, EOLP18 (AGCGGGAG CTGCCAGTC), that differed by at least two nucleotides from other *env* sequences of RT-PCR and genomic PCR clones. Hybridization was carried out for 18 h at 55°C in a solution containing 6X sodium saline citrate (SSC), 5X Denhardt's solution, 0.01 M sodium phosphate, 1 mM ethylenediamine tetraacetate (EDTA), 0.5% sodium dodecyl sulfate (SDS), and 100 µg/ml denatured sheared salmon sperm DNA. After hybridization, membranes were washed in 6X SSC and 0.1% SDS three times for 15 min each time at room temperature and a final time with the same solution for 20 min at 55°C. Washed membranes were exposed to X-ray film (Fuji, Tokyo, Japan) for 24 h at 80°C. The sequence identity was confirmed by sequencing the PCR products of the *env* region from the isolated clones. Cosmid DNA containing the *HERV-E* gene was restriction-mapped

and subcloned into the pBluescript II vector (Stratagene, La Jolla, CA, USA). Sequences of *HERV-E* and its flanking regions were determined by reading both strands with an automated DNA sequencer, model IR4200L (Li-Cor).

Expression analysis by Northern blot hybridization

Multiple Tissue Northern blots (Clontech, Palo Alto, CA, USA) were used for Northern analyses.

Prehybridization and hybridization were carried out according to the manufacturer's instructions. Blots hybridized with an oligoprobe (EOLP18) were washed using the same procedure as for the cosmid library screening. Otherwise, the membrane was washed three times at room temperature with 2X SSC and 0.05% SDS for 15 min and once at 55°C with 0.1X SSC, 0.1% SDS for 30 min and exposed to X-ray film (Fuji) for 3 to 5 days. Probe DNA fragments were prepared either directly from a plasmid clone (*pol* probe) or by PCR amplification (*env*, *gag*, and LTR probes). Primers and thermal cycle conditions were as follows: for the *gag* probe, 5'-CACCAGTTGCTCATGTTTCCT/5'-TTCYCTG AATGACCTCAGAG, 94°C, 30s/51°C, 30s/72°C, 40s extension; for the LTR probe, 5'-CTAGAYGACCTTGG CRCCAC/5'-TAGCAGGAYRACYCACAGAC, 94°C, 30s/52°C, 30s/72°C, 40s extension. The *env* fragment was amplified with primers ENV1 and ENV3 as described above.

Results

Identification of an active *HERV-E* gene

Using poly(A)⁺RNA extracted from a placenta, RT-PCR was performed and the amplified product was cloned. Among 29 clones sequenced, 19 had identical sequences in the 428-bp fragment (excluding primer sequences) of the *env* region, and 5 had one or two mismatches to their sequences at random sites. Because our previous study had confirmed that 22% of the clones had a single mismatch caused by reverse transcription, PCR, or sequencing errors (Sugimoto et al. 2001), these 24 clones were considered to derive from a single locus or multiple loci with an identical sequence in the analyzed region. Database searches identified a draft sequence with accession number AC061975 that showed 100% homology with these transcripts. This *env* sequence differed by more than 6% from those of other loci. Three clones seemed to be transcripts from another locus found in a database sequence (GenBank accession no. AC087342). The remaining two clones corresponded to two independent loci (AC079804 and AC010329) (Fig. 1).

In contrast, sequencing of 28 genomic PCR clones revealed a variety of clones that seemed to represent at least 14 loci. Among them, 9 loci could be identified in database sequences, including the major transcript and λ4-1 loci (AC061975 and AC010329, respectively) (Repaske et al. 1985). Thus, the RT-PCR and subsequent sequencing

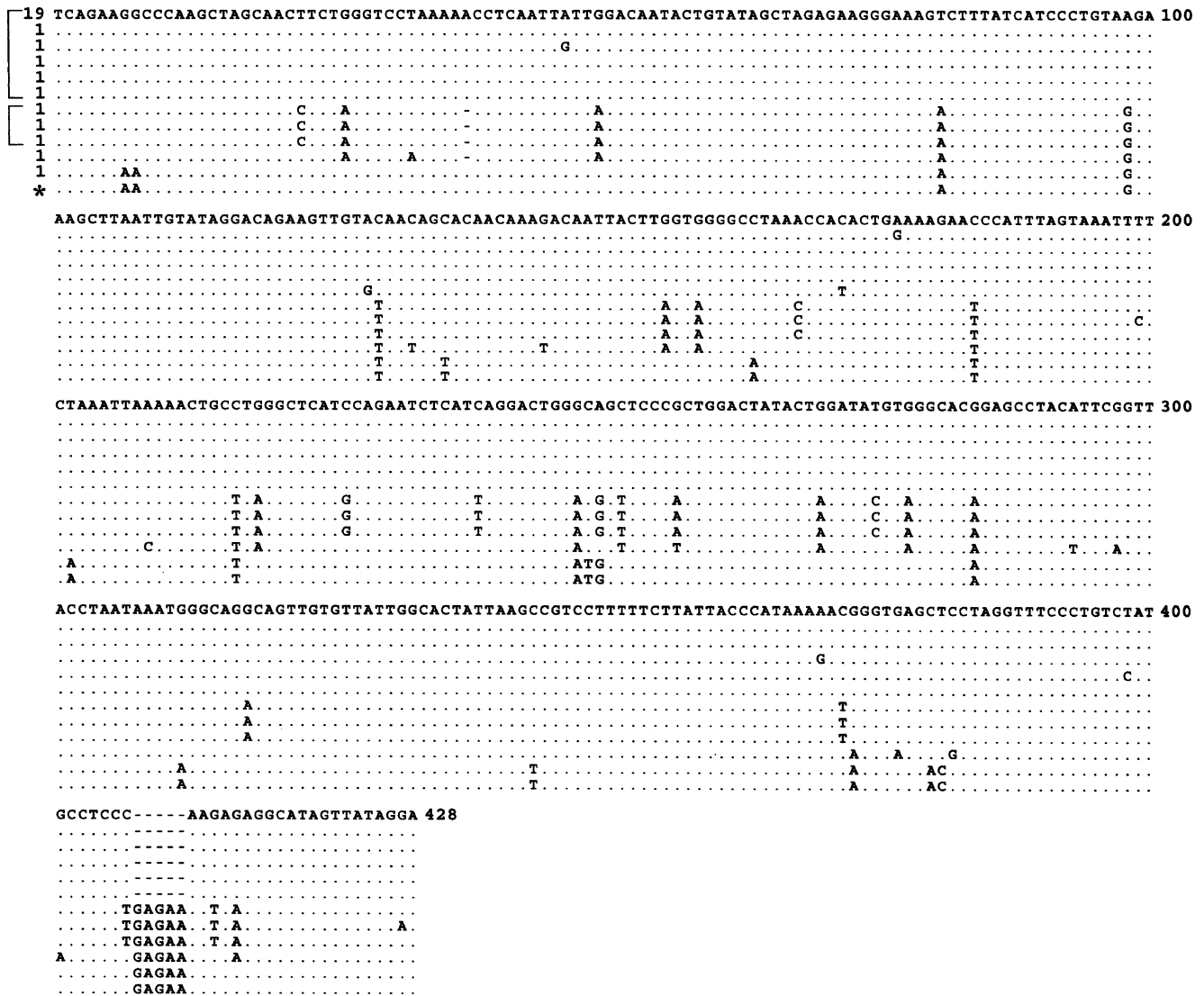


Fig. 1. Expression analysis by reverse transcriptase-polymerase chain reaction (RT-PCR) and subsequent sequencing of cloned products. The *env* region of *HERV-E* was reverse-transcribed from placental poly(A)⁺RNA, PCR-amplified, cloned, and sequenced. The 428-bp *env* sequence of the major transcript is shown, and only substituted

nucleotides are indicated in the other transcripts. Clone numbers that carried the indicated sequences are shown on the left, and clones considered to derive from a single locus are semi-bracketed. As a reference, the sequence of λ 4-1 (Repaske et al. 1985) is shown (asterisk)

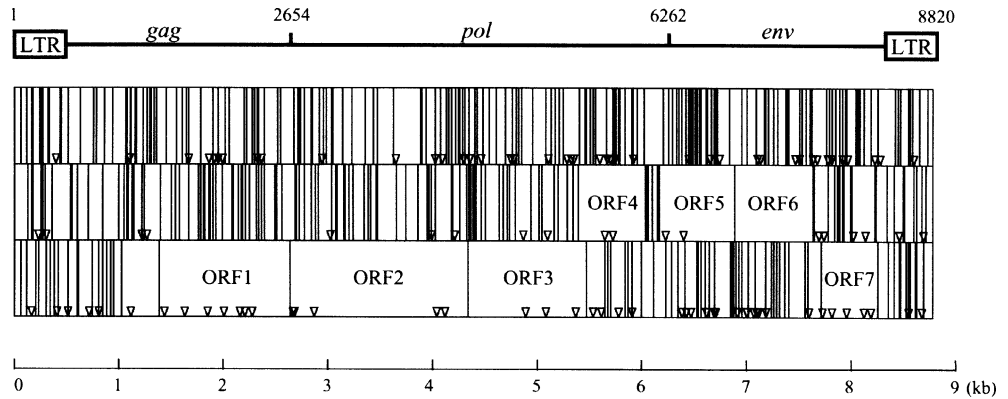
analyses predicted the presence of only a single active locus among members of the *HERV-E* family, at least in the placenta.

Isolation and structural analysis of the active *HERV-E* gene

From sequence comparison of RT-PCR and genomic PCR clones, we prepared an anti-sense oligonucleotide probe (oligoprobe) that differed from the others in two or more nucleotides. Using this probe, we isolated cosmid clones that carried the active *HERV-E* gene (*ERVE1*) and determined its primary structure and flanking sequences (GenBank, AB062274). A full-length (8820 bp) proviral sequence was demarcated by 4-bp direct repeats. The overall

homology with λ 4-1 was 93%. While the poly(A) signal was present in both long terminal repeats (LTRs), the 5' LTR (495 bp) was devoid of either a TATA box or a CAAT box sequence. The many termination codons probably rendered the provirus defective for replication. However, several open reading frames (ORFs) were present (Fig. 2). Among them, the long ORFs in the *gag* and *pol* regions coded for 417- and 572-residue proteins that shared 33% identity and 49% similarity (or positivity) with a feline leukemia viral Gag polyprotein (Swissprot, P26803) and 53% identity (and 66% similarity) with the Friend murine leukemia viral Pol polyprotein (P26809), respectively. ORF3 and ORF4 corresponded to sequences that code for the RNase H domain and the Integrase C-terminal portion of the Pol polyprotein, respectively. The *env* region was divided into three ORFs (ORF5–ORF7) by termination codons. ORF6 and ORF7

Fig. 2. Primary structure of *ERVE1*. The overall structure of the gene is schematically drawn with a line and boxes indicating long terminal repeats (*LTRs*). From comparison with the λ 4-1 sequence, boundaries with *gag-pol* and *pol-env* are assigned and indicated by vertical lines (top). Open reading frames (*ORFs*) in three frameworks are shown below the gene. They are arbitrarily numbered in order from 5' to 3' (*ORF1-ORF7*). Methionine codons are indicated by triangles



represented parts of the Env polyprotein homologous to those of various kinds of retroviruses. Putative proteins translated from ORF5 shared no significant similarity with any proteins in the database. However, close inspection revealed the presence of short stretches of amino acids that showed 57% similarity over 40 amino acid residues of CD1C (P29017) and 44% similarity over 90 amino acid residues deduced from a mouse expressed sequence tag (EST) (BF159844). The homologous portion of CD1C corresponded to the class I MHC α 3 domain that interacts with CD8. The mouse EST seemed to represent a part of mouse FK506 binding protein 4 (p59) mRNA (X70887), which is an immunophilin functioning as a molecular chaperone.

The *ERVE1* gene was partially or fully contained in two database sequences (one complete and one draft sequence: AC002094 and AC061975, respectively). The gene was inserted in the opposite direction in the spacer region between the *PPY2* (pancreatic polypeptide 2) and *PYY2* (seminalplasmin) genes that were mapped to 17q11 (Couzens et al. 2000).

RNA expression analyses

The *ERVE1* gene-specific antisense oligoprobe detected a 3.3-kb major signal associated with a 3.9-kb faint signal in the pancreas and thyroid. The expression of these transcripts was uncertain in the placenta, but instead a very faint signal with a smaller size (1.1 kb) was recognized (Fig. 3a). The PCR-amplified *env*-fragment probe demonstrated essentially the same pattern. Thus, the Northern blot hybridizations confirmed that the expression from the *ERVE1* locus was tissue specific and suggested the presence of spliced forms of transcripts. To examine whether the transcripts derived from a single locus, the *env* region of a pancreatic Marathon-Ready cDNA (Clontech) was amplified with the same primers used in the RT-PCR amplification, cloned, and sequenced. Among 30 clones analyzed, 18 showed a perfect match with the *env* sequence of the *ERVE1* locus, and 11 had only one (7 clones) or two (4 clones) mismatches at random sites. Thus, all but one clone seemed to derive from the single *ERVE1* locus. Taken

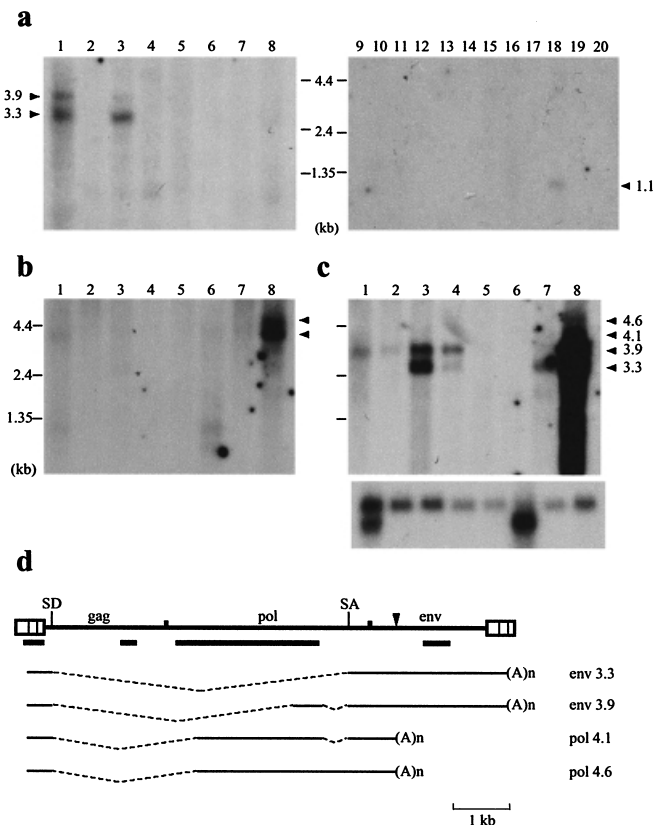


Fig. 3a-d. *HERV-E* expression in normal human tissues. Northern blots (Clontech) were hybridized with an antisense oligonucleotide probe in the *env* region (a), or with a DNA fragment from the *pol* region (b) or the LTR (c). The lower panel in (c) displays a blot probed with β -actin. Approximate sizes are indicated beside the arrowheads. Probes employed are shown as thick lines below the gene in (d). Based on the combined data from the Northern blot and database (expressed sequence tag) analyses, putative structures of transcripts are illustrated in (d). A cryptic poly(A) signal is indicated with a vertical arrowhead. RNAs were from pancreas (lane 1), adrenal medulla (lane 2), thyroid (lane 3), adrenal cortex (lane 4), testis (lane 5), thymus (lanes 6 and 13), small intestine (lanes 7 and 17), stomach (lane 8), brain (lane 9), heart (lane 10), skeletal muscle (lane 11), colon (lane 12), spleen (lane 14), kidney (lane 15), liver (lane 16), placenta (lane 18), lung (lane 19), and peripheral blood leukocytes (lane 20) in a, and heart (lane 1), brain (lane 2), placenta (lane 3), lung (lane 4), liver (lane 5), skeletal muscle (lane 6), kidney (lane 7), and pancreas (lane 8) in b and c. RNAs are all from adult tissues

together, we concluded that the *env*-containing transcripts detected in the pancreas (and possibly in the thyroid) were mostly and exclusively expressed from the *ERVE1* locus.

The *pol*-fragment probe also detected similar doublet signals with approximate sizes of 4.1 and 4.6 kb in the pancreas (Fig. 3b). Besides the two signals, a very faint band 3.9 kb in size seemed to be present in the pancreas, but the very weak intensity made the presence of this transcript uncertain. The *gag* probe did not detect any signals at all (data not shown).

An LTR probe revealed four signals that were consistent in size with those detected with *env* and *pol* probes in the pancreas (Fig. 3c). However, the signal intensities of the larger doublet were quite weak when compared with those of the smaller (*env*-containing) doublet. The *env*-containing signals were clearly observed in the placenta in this blot, and the 3.9-kb signal was faintly visible in the heart, brain, and lung. Although it is likely that the 3.3-kb transcript in the placenta is mostly expressed from the *ERVE1* locus, it is not known whether the *pol*-containing transcripts in the pancreas and the 3.9-kb transcript in other tissues were products from the same single locus.

A database search detected 16 ESTs from the *ERVE1* locus. Among them, 8 ESTs represented only a part of the 3' LTR sequence, 4 contained continuous *env*-3' LTR sequences, and 2 were parts of the *pol* region. One EST (AI417962) possessed a discontinuous *pol* and *env* sequence (5404–5421 and 5997–6449). This may be a splicing product. The other EST (AI741486) revealed an *env* sequence followed by a 20-nucleotide poly(A) sequence 12 bp downstream from a cryptic polyadenylation signal (6731–6736). This may explain the *pol*-containing transcripts that were not detected by *env* probes (oligoprobe, 7478–7495; *env*-fragment probe, 7217–7685). Taking into consideration all the above results, possible transcripts from the *ERVE1* locus are schematically drawn in Fig. 3d.

5' LTR sequence comparison

To search for sequences that might confer on the *ERVE1* locus tissue-specific enhancer and promoter activity and differentiate it from other *HERV-E* loci, we picked up 8 full-length and partial or truncated *HERV-E* loci containing at least the 5' LTR sequence from the GenBank database. Homologies of these 5' LTR sequences with that of *ERVE1* ranged from 67% to 95%. In addition, ten solo LTRs showing more than 90% homology were selected and aligned together with the eight 5' LTRs and the *ERVE1* 5' LTR (Fig. 4). A total of seven substitutions in the 5' LTR of *ERVE1* were found that differed from all the other LTRs. Among them, two resulted in the generation of three consensus sequences (CS) including the gamma interferon responsive element (IRE)-CS (234A; CATTATAT), the granulocyte macrophage-colony stimulating factor (GM-CSF)-CS (234A; CATTAA), and the eukaryotic transcription factor GCN4-CS (319G; GATGACTCTT). In addition, there was a G substitution at nucleotide 176 that gave rise to the alpha-INF.2 site (AARKGA) in the 5' LTRs of *ERVE1* and an AC017104 locus. It remains to be examined whether these sequences are critical to the tissue-specific expression of the *ERVE1* gene.

Discussion

We previously reported a strategy to identify active *HERV* genes (Sugimoto et al. 2001). In the present study, we have applied this strategy to the *HERV-E* family and have identified an active locus that is expressed in normal pancreas and thyroid tissues with enough transcripts for detection by Northern blot hybridization. The expression of *HERVs* is usually strictly suppressed in normal tissues, and has previ-

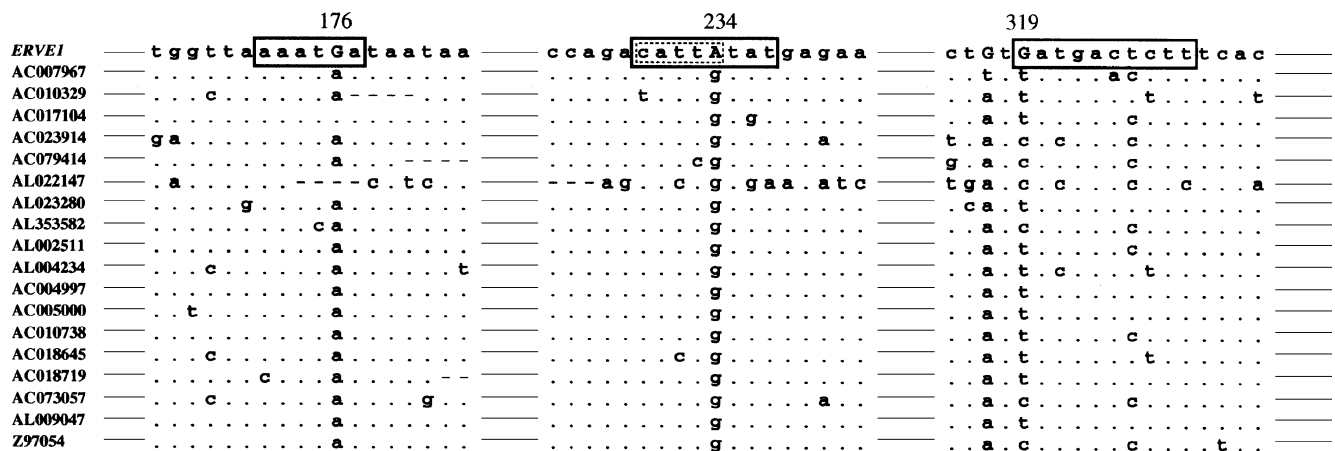


Fig. 4. Sequence comparison of 5' or solitary LTRs. The 5' LTR sequence of *ERVE1* was compared with eight 5' LTRs of distinct *HERV-E* loci, and with ten solitary LTRs that showed more than 90% homology with the *ERVE1* 5' LTR. Only portions of the 5' LTR are depicted in which base substitutions resulted in consensus sequences that were highly specific to *ERVE1*. Consensus sequences of gamma-

IRE (WKKANNY), GM-CSF (CATTW), GCN4 (RRTGAYTSWY), and alpha-INF.2 (AARKGA) are boxed and the *ERVE1*-specific substitutions are shown with capital letters. Identical sequences are depicted with dots, and deletions (or absences of corresponding nucleotides) are indicated with *hyphens*

ously been primarily observed in various tumors and tumor cell lines or predominantly in the placenta among normal tissues (Kato et al. 1987; Kjellman et al. 1999; Löwer et al. 1993, 1996). Therefore, this is, to the best of our knowledge, the first report of predominant expression of HERV in normal tissues other than placenta. It is highly likely that *HERV-E* transcripts (at least the *env*-containing ones) in the pancreas were exclusively derived from the *ERVE1* locus, and we suggest that this is the only active locus among *HERV-E* loci in the human genome in a normal state. Thus, these results together with the results of the sequencing analysis of RT-PCR clones from placental RNA, suggest that it is possible that mRNAs detected with DNA probes derived from λ 4-1, another distinct *HERV-E* locus, may have been products from the *ERVE1* locus (Rabson et al. 1983). While there are examples of *HERV-E* contributing to the expression of authentic genes as a tissue-specific promoter of the salivary amylase gene (Ting et al. 1992) and pleiotrophin (Schulte et al. 1996) and as an alternative promoter and/or enhancer of the *MIDI1* gene (AF041208), the apolipoprotein CI gene, and the endothelin B receptor gene (Medstrand et al. 2001), this is the first case in which the mRNA consists of the *HERV-E* locus itself.

The *ERVE1* locus essentially expressed two kinds of mRNAs in the pancreas: one containing *pol* and the other containing *env*. The *pol* ORF possessed a full RT domain that is highly homologous to those derived from a wide range of retroviruses such as murine leukemia virus, gibbon leukemia virus, and even human immunodeficiency virus. The *env*-containing mRNA was devoid of *pol* and *gag* sequences, and thus seemed to contain the entire *env* sequence. The *env* region was divided into three ORFs (ORF5–ORF7) by the presence of termination codons. Both ORF6 and ORF7 contained parts of the Env polyprotein, although their frameworks were different from each other. Interestingly, ORF5 revealed two portions that possessed similarities, although of limited length, with the immune-related protein CD1C and with FK506-binding proteins. CD1C is a member of the CD1 family, which belongs to the MHC superfamily and presents lipid antigens (Blumberg et al. 1995). The portion of CD1C showing a homology with ORF5 corresponds to the CD8 interaction site of the Class I MHC α 3 domain (Hansen et al. 2000). It is tempting to speculate that the ORF5 product competes with CD8 to interfere with the recognition of self-antigens by T cells. FK506-binding proteins are immunophilins and function in a wide range of cellular activities including in such events as folding, assembly, and trafficking of proteins, as well as mediate the immunosuppressant drug FK506 to inhibit T cell activation (Pratt and Toft 1997). Alternatively, it is conceivable that the ORF5 product might function as a molecular chaperone of proteins such as immunophilins synthesized and secreted in the pancreas and thyroid. While *ERVE1* is localized to 17q11, no autoimmune disease or inherited disease has been linked to this chromosomal region. This might be due to the functional redundancy of the gene.

More than 50 copies of *HERV-E* have been estimated to exist in the human genome (Tristem 2000). However, the

present study suggests that the only active locus in normal adult tissues may be the *ERVE1* locus. Although it is possible, of course, that some loci are active in a limited temporal or spatial manner, or that the PCR primers employed in this study resulted in a slightly biased amplification, this locus could be a good control to investigate the regulating mechanisms of retrotransposons. Although a quick inspection of *HERV-E* LTR sequences identified three consensus sites that were present only in the *ERVE1* LTR in the relevant places, it awaits further experiments to determine whether they actually confer tissue-specific promoter/enhancer activity on the locus. On the other hand, we have found the *PPY2* gene 8kb upstream of the *ERVE1* gene, oriented in the opposite direction (AC002094). It is possible that the *PPY2* enhancer affects the activity of *ERVE1*, but this does not explain the expression of *ERVE1* in the thyroid. In addition, *PPY2* is expressed in the liver and intestine as well as in the pancreas. In a 167-kb sequence of AC002094 that ends in the *pol* region of the *ERVE1* gene, more than ten genes have been identified, and, further, the *PPY2* gene is also present oriented in the opposite direction 2.5kb downstream of the *ERVE1* 3' end (AC061975). This fairly gene-rich chromosomal context may allow the *ERVE1* gene to evade suppressive mechanisms directed to retrotransposons.

The activity of retrotransposons is usually more strictly controlled in humans than in mice. Thus, we expect that humans have acquired some mechanisms to regulate them in addition to those common to both species, or have developed similar mechanisms that operate in different ways and to different extents. Our present study is a successful example of our strategy to identify active HERV genes, and is a first step in the elucidation of such mechanisms as well as the physiological and evolutionary roles of HERVs in the human genome.

Acknowledgment This study was supported by a Grant-in-Aid for Scientific Research in priority area (C).

References

- Best S, Le Tissier P, Towers G, Stoye JP (1996) Positional cloning of the mouse retrovirus restriction gene Fv1. *Nature* 382:826–829
- Blumberg RS, Gerdes D, Chott A, Porcelli SA, Balk SP (1995) Structure and function of the CD1 family of MHC-like cell surface proteins. *Immunol Rev* 147:5–29
- Boeke JD (1997) LINEs and Alus — the polyA connection. *Nat Genet* 16:6–7
- Conrad B, Weissmahr RN, Böni J, Arcari R, Schüpbach J, Mach B (1997) A human endogenous retroviral superantigen as candidate autoimmune gene in type I diabetes. *Cell* 90:303–313
- Couzens M, Liu M, Tuchler C, Kofler B, Nessler-Menardi C, Parker RM, Klocker H, Herzog H (2000) Peptide YY-2 (PPY2) and pancreatic polypeptide-2 (PPY2): species-specific evolution of novel members of the neuropeptide Y gene family. *Genomics* 64:318–323
- Essnault C, Maestre J, Heidmann T (2000) Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* 24:363–367
- Hansen T, Balendiran G, Solheim J, Ostrov D, Nathenson S (2000) Structural features of MHC class I molecules that might facilitate alternative pathways of presentation. *Immunol Today* 21:83–88

- Jinno Y, Yun K, Nishiwaki K, Kubota T, Ogawa O, Reeve AE, Niikawa N (1994) Mosaic and polymorphic imprinting of the WT1 gene in humans. *Nat Genet* 63:305–309
- Kato N, Pfeifer-Ohlsson S, Kato M, Larsson E, Rydnert J, Ohlsson R, Cohen M (1987) Tissue-specific expression of human provirus ERV3 mRNA in human placenta: two of the three ERV3 mRNAs contain human cellular sequences. *J Virol* 61:2182–2191
- Kazazian HH Jr (2000) L1 retrotransposons shape the mammalian genome. *Science* 289:1152–1153
- Kazazian HH Jr, Moran JV (1998) The impact of L1 retrotransposons on the human genome. *Nat Genet* 19:19–24
- Kazazian HH Jr, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE (1988) Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* 1988 332:164–166
- Kjellman C, Sjögren HO, Salford LG, Widegren B (1999) HERV-F (XA34) is a full-length human endogenous retrovirus expressed in placental and fetal tissues. *Gene* 239:99–107
- Löwer R, Boller K, Hasenmaier B, Korbmacher C, Müller-Lantsch N, Löwer J, Kurth R (1993) Identification of human endogenous retroviruses with complex mRNA expression and particle formation. *Proc Natl Acad Sci U S A* 90:4480–4484
- Löwer R, Löwer J, Kurth R (1996) The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proc Natl Acad Sci U S A* 93:5177–5184
- Medstrand P, Landry JR, Mager DL (2001) Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. *J Biol Chem* 276:1896–1903
- Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S, Keith JC Jr, McCoy JM (2000) Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403:785–789
- Miki Y (1998) Retrotransposal integration of mobile genetic elements in human diseases. *J Hum Genet* 43:77–84
- Miki Y, Katagiri T, Kasumi F, Yoshimoto T, Nakamura Y (1996) Mutation analysis in the BRCA2 gene in primary breast cancers. *Nat Genet* 13:245–247
- Moran JV, DeBerardinis RJ, Kazazian HH Jr (1999) Exon shuffling by L1 retrotransposition. *Science* 283:1530–1534
- Nakamura TM, Cech TR (1998) Reversing time: origin of telomerase. *Cell* 92:587–590
- Pratt WB, Toft DO (1997) Steroid receptor interactions with heat shock protein and immunophilin chaperones. *Endocr Rev* 18:306–360
- Rabson AB, Steele PE, Garon CF, Martin MA (1983) mRNA transcripts related to full-length endogenous retroviral DNA in human cells. *Nature* 306:604–607
- Repaske R, Steele PE, O'Neill RR, Rabson AB, Martin MA (1985) Nucleotide sequence of a full-length human endogenous retroviral segment. *J Virol* 54:764–772
- Schulte AM, Lai S, Kurtz A, Czubayko F, Riegel AT, Wellstein A (1996) Human trophoblast and choriocarcinoma expression of the growth factor pleiotrophin attributable to germ-line insertion of an endogenous retrovirus. *Proc Natl Acad Sci U S A* 93:14759–14764
- Sugimoto J, Matsuura N, Kinjo Y, Takasu N, Oda T, Jinno Y (2001) Transcriptionally active HERV-K genes: identification, isolation and chromosomal mapping. *Genomics* 72:137–144
- Ting CN, Rosenberg MP, Snow CM, Samuelson LC, Meisler MH (1992) Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev* 6:1457–1465
- Tristem M (2000) Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. *J Virol* 74:3715–3730
- Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, Collins FS (1991) A de novo Alu insertion results in neurofibromatosis type 1. *Nature* 353:864–866