SHORT COMMUNICATION

HERV-K-T47D-Related Long Terminal Repeats Mediate Polyadenylation of Cellular Transcripts

Corinna Baust,*,[1] Wolfgang Seifarth,* Herbert Germaier,† Rüdiger Hehlmann,* and Christine Leib-Mösch*†

*Medical Clinic III, Faculty of Clinical Medicine Mannheim, University of Heidelberg, D-68305 Mannheim, Germany; and †GSF-National Research Center for Environment and Health, Institute of Molecular Virology, D-85758 Oberschleißheim, Germany

Received November 2, 1999; accepted February 28. 2000

The human genome harbors thousands of long terminal repeats (LTRs) that are derived from endogenous retroviruses and contain elements able to regulate the expression of neighboring cellular genes. We have investigated the ability of human endogenous retroviral (HERV)-K LTRs to provide transcriptional processing signals for nonviral sequences. Four chimeric cDNA clones isolated from a cDNA library derived from the human cell line T47D were found to be polyadenylated within an HERV-K-T47D-related LTR. Two transcripts containing an as yet unknown cellular sequence were probably derived from the same genomic locus but their 3′ ends were processed at different positions of the LTR. Structural analysis of the polyadenylation site suggests RNA stem-loop structures similar to the HTLV-1 Rex responsive element that bring the two remote AAUAAA and GU-rich elements into the spatial juxtaposition necessary for correct 3′ end processing. The cellular part of the third chimeric clone shows significant homology to an exon of the human tyrosine phosphatase 1 gene, although oriented in the antisense direction compared to the adjacent LTR. Furthermore, we found that the 3′ untranslated region of the human transmembrane tyrosine kinase gene *FLT4* is probably derived from a partial HERV-K-T47D LTR sequence. Taken together, our data suggest that LTRs of the HERV-K-T47D family display biological function by mediating polyadenylation of cellular sequences.

Human endogenous retroviral (HERV) elements are estimated to comprise up to 5% of human DNA (12). Originally derived from integrated exogenous retroviruses, they have spread throughout the human genome by retrotransposition. The vast majority of HERVs are defective; however, there still exist a number of elements that are actively transcribed or even give rise to retrovirus-like particles (for review see Refs. 16, 17, 19, 27). Several distinct HERV families have been identified, most of them consisting of full-length or truncated proviruses and a large number of solitary long terminal repeats (LTRs). HERV LTRs contain elements that can regulate the transcription and polyadenylation of neighboring cellular sequences. However, although many LTRs have been described to modulate transcriptional initiation (6, 8, 13, 18), up to now only LTRs of the HERV-H and the related HERV-F family have been found to process the 3′ ends of nonviral transcripts (10, 14, 20, 21).

To analyze further the influence of HERV LTRs on 3′ end processing of cellular genes, we focused on the human endogenous retroviral element HERV-K-T47D, a member of the HML-4 subgroup of the HERV-K superfamily (25). HERV-K-T47D was identified during the characterization of RNA sequences packaged in particles released by the human mammary carcinoma cell line T47D upon steroid treatment (24). We recently showed that the human genome contains several hundred solitary HERV-K-T47D-related LTRs (25), suggesting that this family represents a reservoir of potentially functional control elements with promoter, enhancer, or polyadenylation activity that could be involved in gene expression. In this study we screened a human oligo(dT) primed λ cDNA library with an HERV-K-T47D LTR probe and retrieved several chimeric cellular cDNA/LTR transcripts. We show here that the isolated HERV-K-T47D LTRs possess biological function, since they are able to mediate polyadenylation of cellular transcripts at two different viral polyadenylation sites.

To establish the cDNA library, T47D cells (American Type Culture Collection) were cultivated and treated

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. AF148948–AF148950.

[1] To whom correspondence should be addressed at III. Medizinische Klinik, Klinikum Mannheim, Universität Heidelberg, Wiesbadener Strasse 7-11, D-68305 Mannheim, Germany. Telephone: (+49) 621 383-4103. Fax: (+49) 621 383-4201. E-mail: Corinna.Baust@urz.uni-heidelberg.de.

0888-7543/00 $35.00

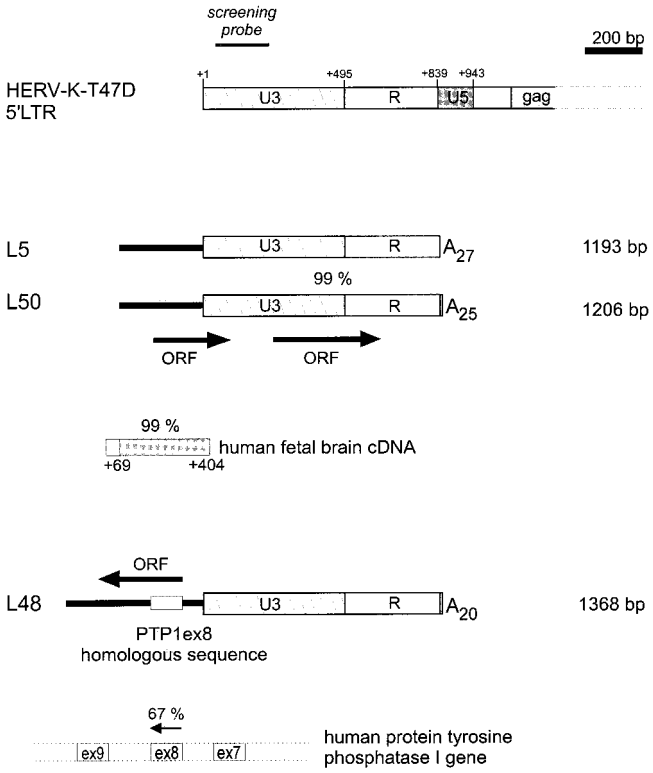

FIG. 1. Chimeric cellular cDNA/HERV-K-T47D-LTR clones. Clones are drawn to scale. U3, R, and U5 LTR regions as well as ORFs are indicated. Apart from the different poly(A) addition sites, L5 and L50 display 99% nucleotide identity. Mismatches are positioned at nt 249 and 731 of U3/R (numbering refers to the related HERV-K-T47D 5′ LTR). The upstream cellular sequence is highly homologous to a region of a cDNA isolated from human fetal brain. Two mismatches are located at nt 93 and 257 of the fetal brain cDNA. The cellular sequence of L48 contains a domain (shaded box) displaying 77% amino acid identity and 67% nucleotide identity to exon 8 of the protein tyrosine phosphatase 1C gene (PTP1ex8), but in the antisense orientation. The homologous regions are shaded, with the degree of homology being given as a percentage number. The probe used for screening of the cDNA library is indicated at the top. It was generated by PCR amplification of the LTR-U3 region of the HERV-K-T47D provirus (25), using the forward primer GCGGGATCCGAGGCAAGAGACTGAAGGCAC (nt 25–47 of the HERV-K-T47D 5′ LTR) and the reverse primer ACTTCTCACCAATGTCCCTTCAGC (nt 232–254 of the HERV-K-T47D 5′ LTR). Hybridization with the ^{32}P-labeled probe was performed under low-stringency conditions [30% formamide, 5× SSC (1× SSC is 0.15 M NaCl plus 0.0015 M sodium citrate), 50 mM NaH_2PO_4, pH 7.0, 5× Denhardt's solution, 0.1% SDS, 1 mM EDTA, and 100 μg of denatured sheared herring sperm DNA per milliliter]. Washing was carried out in 2× SSC and 0.1% SDS for 90 min at 55°C.

with steroids as described previously (25). Total RNA was isolated by the guanidine isothiocyanate method (3), followed by poly(A)$^+$ mRNA selection using oligo(dT) chromatography (22). The library was constructed with the ZAP cDNA synthesis kit (Stratagene Cloning Systems, La Jolla, CA) as recommended by the supplier's protocols. Phages were transferred onto nitrocellulose membranes and hybridized with a ^{32}P- labeled 229-bp probe obtained by PCR amplification from the U3 region of the HERV-K-T47D 5′ LTR (Fig. 1). Upon screening of approximately one million plaques, we obtained about 100 hybridization signals. We isolated several hybridizing cDNA clones by amplification of the inserts using the forward primer CGCTCGAAATTAACCCTCACTAAAGGG and the reverse primer CTCACTATAGGGCGAATTCGGTACCGGGCC. PCR products were inserted into the *Eco*RI site of the vector pBS (SK+) (Stratagene Cloning Systems). Resulting clones were characterized by restriction enzyme analysis, and DNA sequences were determined using the T7 sequencing kit (Amersham Pharmacia Biotech, Uppsala, Sweden). For sequence analysis, we used the Nucleic Acid and Protein Sequence Analysis Program GeneWorks (IntelliGenetics Inc., Mountain View, CA). The structure of three of the isolated clones is illustrated schematically in Fig. 1. They are derived from chimeric transcripts, each containing nonviral sequences upstream of an LTR U3/R segment. The U3/R regions display 79% (L5), 81% (L50), and 83% (L48) nucleotide identity with the corresponding regions of the HERV-K-T47D 5′ LTR and were found to be polyadenylated. The complete sequences of the cDNA/LTR chimeras can be retrieved from the EMBL Data Library (Accession Nos. AF148948–AF148950).

The chimeric cDNA/LTR clones L5 and L50 differ by 2 nucleotides only, suggesting that they are derived from the same genomic locus. The mismatches may represent natural polymorphisms or result from generating the clones with reverse transcriptase and *Taq* polymerase, both of which are known to introduce errors into the amplified product (26). The nonviral portion upstream of U3/R and the first 14 nucleotides of the HERV U3 region display 99% identity with nt 69 to 404 of a 404-bp cDNA with an as yet unknown function that was isolated from human fetal brain by T. Fujiwara (unpublished results; Accession No. D60591). Further screening of the expressed sequence tag database revealed three additional cDNA clones (unpublished results; Accession Nos. AI085192, AI825942, and AI692688) derived from different human cells or tissues (senescent fibroblasts, adenocarcinoma, and lung, respectively) showing 94 to 100% nucleotide identity with clone L50. These clones comprise the complete nonviral part of the human fetal brain-derived sequence and extend 62 to 70 nt into the HERV-K-T47D LTR sequence. Unlike clones L5 and L50, the cDNAs do not contain the original polyadenylation site since they end at an A-rich stretch located in the U3 region of the LTR that probably served as a binding site for the oligo(dT) primer during cDNA synthesis.

Despite their almost identical nucleotide sequences, both L5 and L50 cDNAs are polyadenylated at different positions, indicating that they are derived from two separate transcripts. The transcript L5 is polyadenylated 69 nt downstream of a unique canonical AAUAAA sequence (3′ of the dinucleotide UA), whereas L50 uses a distinct poly(A) addition site (CA) located 84 nt downstream of the AAUAAA hexamer (Fig. 2A). The large distance between the putative polyadenylation signal and the two poly(A) addition sites is unusual since it normally comprises only about 20 nt (2). In general, to

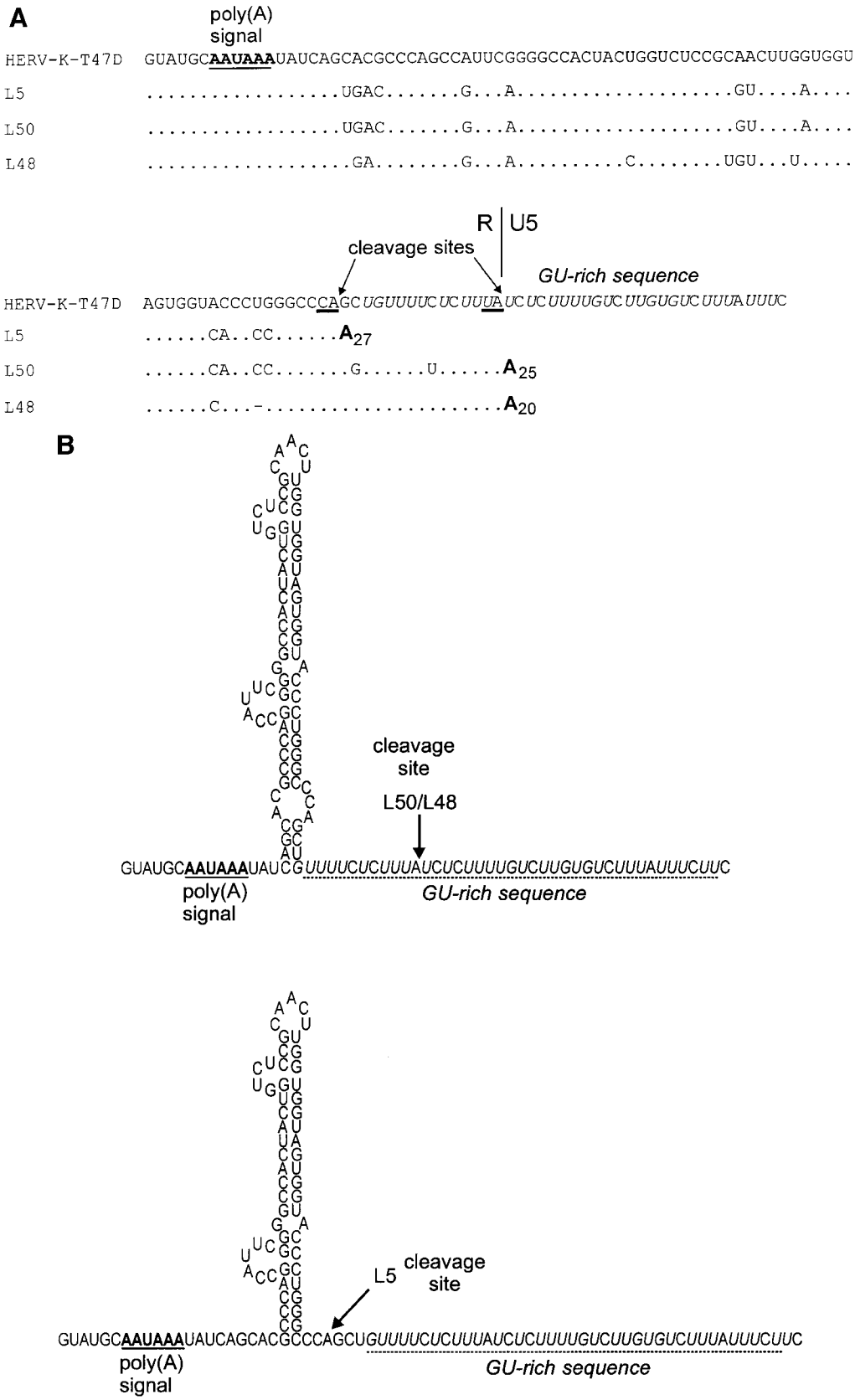

FIG. 2. Polyadenylation of chimeric cDNA/LTR transcripts. (**A**) Putative poly(A) elements located at the R/U5 boundary of the HERV-K T47D LTR. Identical nucleotides are depicted as dots, and missing nucleotides are indicated by dashes. The two potential poly(A) addition sites are underlined. (**B**) Secondary structures of the polyadenylation site. The putative poly(A) signal is underlined, and the U-rich sequences located further downstream are emphasized by italics.

ensure correct 3′ end processing, the binding of the cleavage/polyadenylation specificity factor CPSF to the AAUAAA element must be stabilized by cooperative interaction with the cleavage stimulating factor CstF, which recognizes a U- or GU-rich element located 3′ of the polyadenylation site (for review, see Ref. 4). This interaction is a prerequisite for the formation of a stable processing complex, which, after recruitment of cleavage factors CF I and CF II and poly(A) polymerase, ultimately leads to cleavage and subsequent polyadenylation. Changing the spacing between the AAUAAA element and the U- or GU-rich element effectively prevents correct 3′ end processing since the factors CPSF and CstF can no longer interact. In the case of HTLV-1, the large distance (274 nt) between the poly(A) signal and the 3′ processing site is reduced by the secondary structure of the Rex responsive element, thus spatially juxtaposing the widely separated AAUAAA hexamer and GU-rich elements (1). Here, we analyzed the secondary structure of the HERV-K-T47D LTR using the program mfold 3.0 (23) and found that, similar to the Rex responsive element, the sequence between the putative polyadenylation signal and the U-rich sequence located immediately downstream of the two cleavage sites adopts a stem-loop structure (Fig. 2B). This might serve to bring the poly(A) elements into proper alignment and thus allow correct 3′ end processing. Furthermore, the two possible stem-loop structures shown in Fig. 2B could explain the alternative use of two different polyadenylation sites in clones L5 and L50. Similar observations have also been made for HERV-K10 LTRs (Roswitha Löwer, Langen, Germany, 1999), whereas with HERV-H LTRs the distance between the polyadenylation signal and the poly(A) addition site lies within the normal range (10, 20) and therefore does not necessitate the formation of a hairpin structure.

The U3/R sequence located at the 3′ end of the chimeric clone L48 displays 88% nucleotide identity with the LTR of clones L5 and L50. As in the case of L50, the L48 transcript is polyadenylated 84 nt downstream of the AAUAAA hexamer. The nonviral part of L48 upstream of the LTR contains a continuous open reading frame (ORF) that is oriented in the antisense direction compared to the adjacent LTR (nt 420 to 145 of the L48 cDNA) (Fig. 1). The first 31 amino acids encoded by this ORF are 77% identical with the amino acid sequence encoded by exon 8 of the human protein tyrosine phosphatase 1 gene (*PTP1*).The high level of amino acid homology and the fact that the extension of the homologous region (named *PTP1ex8hom*, nt 418 to 325 of the L48 cDNA with 67% nucleotide identity to *PTP1* exon 8) exactly matches the length of *PTP1* exon 8 suggest that clone L48 might contain portions of an as yet unknown human protein tyrosine phosphatase gene related to *PTP1*. PTP1 is a nontransmembrane human phosphatase that has been described to influence the interferon-α/β stimulated Jak/Stat pathway in macrophages (5) and to counteract the erythropoietin-mediated activation of tyrosine kinase JAK2 (15). Northern blot experiments performed under high-stringency conditions with the *PTP1ex8hom* sequence of clone L48 as hybridization probe revealed that an mRNA of about 4 kb in length is expressed in a similar amount in untreated as well as in steroid-treated T47D and in HeLa cells (data not shown). Since the processed human *PTP1* transcripts described in GenBank do not exceed 2277 nt in length, we suggest that the 4-kb mRNA might represent the transcript from which clone L48 was derived.

We analyzed the nonviral sequences located 5′ and 3′ of *PTP1ex8hom* (317 and 77 bp, respectively), but could not detect homologies to known cellular genes. In particular, comparison with the intron sequences between exon 7 and exon 9 of the *PTP1* gene did not reveal any similarities. It is possible that *PTP1ex8hom* represents a solitary exon, a relic of a *PTP1* homologous gene that has arisen by duplication of a *PTP1* ancestor and then lost its functionality by genomic rearrangements and mutations in the course of evolution. Southern blot analysis using *PTP1ex8hom* as a probe revealed that the human genome harbors at least five distinct loci containing *PTP1ex8hom*-related sequences (data not shown). Since *PTP1ex8hom* is oriented in the opposite direction to the adjacent polyadenylated LTR, the chimeric cDNA/LTR transcript contains sequences that are complementary to an assumed *PTP1* transcript. It is tempting to speculate that this transcript may act as an antisense RNA that regulates the expression of *PTP1* or a *PTP1*-related gene by RNA:RNA hybridization. Although direct evidence for gene regulation by naturally occurring antisense RNA has not yet been found, the expression pattern of some antisense RNAs in mammalian cells suggests an interaction with sense DNA that may affect expression of the gene (for review, see Ref. 7).

An unusual involvement of an HERV-K-T47D-related LTR in regulation of gene expression was observed in the case of the human transmembrane tyrosine kinase gene *FLT4* (described in Ref. 9). In addition to the three chimeric clones described above, screening of the cDNA library with the HERV-K-T47D LTR probe led to the isolation of a 1269-bp fragment (L42 in Fig. 3) comprising the 3′ portion of the *FLT4* coding region as well as the complete 3′ untranslated region. Sequence analysis revealed that the latter is composed of an LTR fragment with 84% identity to the U3-R region (nt 158 to 671) of the HERV-K-T47D 5′LTR (Fig. 3). However, the proviral LTR is oriented in the opposite direction compared to the *FLT4* gene, thus excluding regulation of 3′ end processing of *FLT4* by the poly(A) signal located in the LTR-R region. The partial LTR sequence was probably assimilated by the *FLT4* gene in the course of evolution, with the hexamer AUUAAA located 31 nt upstream of the *FLT4* polyadenylation site serving as a new poly(A) signal. Although this sequence is a variant of the classical AAUAAA poly(A) signal, it nevertheless has been

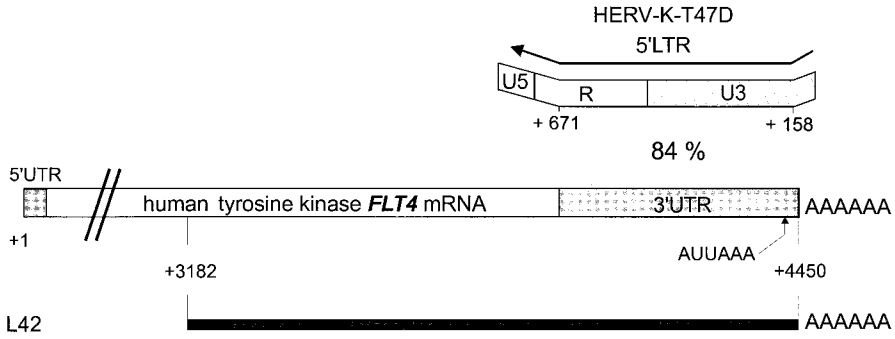

FIG. 3. Schematic comparison of the HERV-K-T47D 5′ LTR with the 3′ untranslated region of the human tyrosine kinase gene *FLT4*. About 160 bp of the 5′ and the 3′ ends of clone L42 (shown at the bottom) were sequenced, revealing 98 and 100% identity with *FLT4*, respectively. The entire clone comprises 1269 bp, corresponding to position 3182 to 4450 of the *FLT4* mRNA (as described in Ref. 9). An inner part of an HERV-K-T47D LTR present in the antisense orientation (nt 671–158) displays 84% nucleotide identity with the 3′ untranslated region (3′UTR) of the *FLT4* cDNA. A putative poly(A) signal AUUAAA located 31 nt upstream of the poly(A) addition site is provided by the LTR. The homology between the 3′UTR and the LTR sequence terminates 7 bp downstream of this putative signal, with the last 18 bp of the *FLT4* mRNA displaying no homology.

shown to be functional in several genes (28). A similar recruitment of a reverse-transcribed element for cellular gene polyadenylation has been described for the mouse thymidylate synthase gene (11). In this case, a truncated LINE 1 element has integrated in the opposite direction immediately downstream of the translational stop codon, thus providing a polyuridine stretch that leads to the activation of a cryptic polyadenylation site.

In conclusion, this study shows that HERV-K-T47D-related LTRs are able to provide functional signals for 3′ end processing of cellular mRNAs. So far, polyadenylation of cellular transcripts within LTRs has been reported only in association with the HERV-H and HERV-F families. Mager (20) isolated two chimeric clones, each consisting of a cellular sequence and an HERV-H LTR that is polyadenylated at the typical viral poly(A) site. Goodchild *et al.* (10) described a third HERV-H LTR that contributes a polyadenylation signal to the transcript of a novel gene (termed *PLT*) found in normal human placenta. By screening the expressed sequence tag database, Mager *et al.* recently identified two genes, termed HHLA2 and HHLA3, that also utilize an HERV-H LTR as a polyadenylation signal (21). Finally, an expressed sequence tag that ends in the 5′LTR of HERV-Fb has been identified (14). This element is located immediately downstream of the Krüppel-related zinc finger gene ZNF195, suggesting that at least a fraction of the ZNF195 mRNAs might use the HERV-Fb 5′LTR as an alternative poly(A) site.

The frequency of solitary HERV LTRs in the human genome and hence the potentially innumerable signals sufficient for 3′ RNA processing of cellular sequences strongly indicate that the number of LTRs involved in regulating gene expression is much higher than the few known cases described above. Moreover, the presence of potential RNA stem-loop structures that serve to bring poly(A) elements into correct juxtaposition could not only extend the repertoire of alternative transcripts, but also possibly contribute some added level of control over 3′end processing.

ACKNOWLEDGMENTS

We thank Dr. A. Arthur-Goettig for critically reading the manuscript. This work was supported by the Forschungsfonds of the Faculty of Clinical Medicine Mannheim, University of Heidelberg.

REFERENCES

1. Ahmed, Y. F., Gilmartin, G. M., Hanly, S. M., Nevins, J. R., and Greene, W. C. (1991). The HTLV-1 Rex response element mediates a novel form of mRNA polyadenylation. *Cell* **64:** 727–737.
2. Birnstiel, M. L., Busslinger, M., and Strub, K. (1985). Transcription termination and 3′ processing: The end is in site! *Cell* **41:** 349–359.
3. Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.* **162:** 156–159.
4. Colgan, D. F., and Manley, J. L. (1997). Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* **11:** 2755–2766.
5. David, M., Chen, H. E., Gielz, S., Larner, A. C., and Neel, B. G. (1995). Differential regulation of the alpha/beta interferon-stimulated Jak/Stat pathway by the SH2 domain-containing tyrosine phosphatase SHPTP1. *Mol. Cell. Biol.* **15:** 7050–7058.
6. Di Cristofano, A., Strazullo, M., Longo, L., and La Mantia, G. (1995). Characterization and genomic mapping of the ZNF80 locus: Expression of this zinc-finger gene is driven by a solitary LTR of ERV9 endogenous retroviral family. *Nucleic Acids Res.* **23:** 2823–2830.
7. Dolnick, B. J. (1997). Naturally occurring antisense RNA. *Pharmacol. Ther. Vol.* **75:** 179–184.
8. Feuchter-Murthy, A. E., Freeman, J. D., and Mager, D. L. (1993). Splicing of a human endogenous retrovirus to a novel phospholipase A2 related gene. *Nucleic Acids Res.* **21:** 135–143.
9. Galland, F., Karamysheva, A., Pebusque, M. J., Borg, J. P., Rottapel, R., Dubreuil, P., Rosnet, O., and Birnbaum, D. (1993). The FLT4 gene encodes a transmembrane tyrosine kinase related to the vascular endothelial growth factor receptor. *Oncogene* **8**(5): 1233–1240.
10. Goodchild, N. L., Wilkinson D. A., and Mager, D. (1992). A human endogenous long terminal repeat provides a polyadenylation signal to a novel, alternatively spliced transcript in normal placenta. *Gene* **121:** 278–294.
11. Harendza, C. J., and Johnson, L. F. (1990). Polyadenylation signal of the mouse thymidylate synthase gene was created by insertion of an L1 repetitive element downstream of the open reading frame. *Proc. Natl. Acad. Sci. USA* **87:** 2531–2535.

12. Jurka, J. (1998). Repeats in genomic DNA: Mining and meaning. *Curr. Opin. Struct. Biol.* **8:** 333–337.
13. Kato, N., Shimotohno, K., Van Leeuwen, D., and Cohen, M. (1990). Human proviral mRNAs down regulated in choriocarcinoma encode a zinc finger protein related to Kruppel. *Mol. Cell. Biol.* **10:** 4401–4405.
14. Kjellman, C., Sjogren, H. O., Salford, L. G., and Widegren, B. (1999). HERV-F (XA34) is a full-length human endogenous retrovirus expressed in placental and fetal tissues. *Gene* **239:** 99–107.
15. Klingmuller, U., Lorenz, U., Cantley, L. C., Neel, B. G., and Lodish, H. F. (1995). Specific recruitment of SH-PTP1 to the erythropoietin receptor causes inactivation of JAK2 and termination of proliferative signals. *Cell* **80:** 729–738.
16. Leib-Mösch, C., Brack-Werner, R., Werner, T., Bachmann, M., Faff, O., Erfle, V., and Hehlmann, R. (1990). Endogenous retroviral elements in human DNA. *Cancer Res.* **50**(Suppl. 1)**:** 5636–5642.
17. Leib-Mösch, C., and Seifarth, W. (1996). Evolution and biological significance of human retroelements. *Virus Genes* **11:** 133–145.
18. Liu, A. Y., and Abraham, B. A. (1991). Subtractive cloning of a hybrid human endogenous retrovirus and calbindin gene in the prostate cell line PC3. *Cancer Res.* **51:** 4107–4110.
19. Löwer, R., Löwer, J., and Kurth, R. (1996). The viruses in all of us: Characteristics and biological significance of human endogenous retrovirus sequences. *Proc. Natl. Acad. Sci. USA* **93:** 5177–5184.
20. Mager, D. L. (1989). Polyadenylation function and sequence variability of the long terminal repeats of the human endogenous retrovirus-like family HERV-H. *Virology* **173:** 591–599.
21. Mager, D. L., Hunter, D. G., Schertzer M., and Freeman, J. D. (1999). Endogenous retroviruses provide the primary polyadenylation signal for two new human genes (HHLA2 and HHLA3). *Genomics* **59:** 255–263.
22. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
23. Santa Lucia, J., Jr. (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA* **95:** 1460–1465.
24. Seifarth, W., Baust, C., Skladny, H., Krieg-Schneider, F., Murr, A., Blusch, J., Hehlmann, R., and Leib-Mösch, C. (1998). Proviral structure, chromosomal distribution and expression of HERV-K(T47D) a novel member of the HERV-K superfamily isolated from T47D cells and particles. *J. Virol.* **72:** 8384–8391.
25. Seifarth, W., Skladny, H., Krieg-Schneider, F., Reichert, A., Hehlmann, R., and Leib-Mösch, C. (1995). Retrovirus-like particles released from the human breast cancer cell line T47-D display B- and C-type related endogenous retroviral sequences. *J. Virol.* **69:** 6408–6416.
26. Tindall, K. R., and Kunke., T. A. (1988). Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry* **27:** 6008–6013.
27. Wilkinson, D. A., Mager, D. L., and Leong, J. C. (1994). Endogenous human retroviruses. *In* "The *Retroviridae*" (J. A. Levy, Ed.), Vol. 3, pp. 465–535, Plenum, New York.
28. Wilusz, J., Pettine, S. M., and Shenk, T. (1989). Functional analysis of point mutations in the AAUAAA motif of the SV40 late polyadenylation signal. *Nucleic Acids Res.* **17:** 67–73.