JOURNAL OF VIROLOGY, Jan. 1995, p. 141–149 Vol. 69, No. 1
0022-538X/95/$04.00+0


# Identification of a Rev-Related Protein by Analysis of Spliced Transcripts of the Human Endogenous Retroviruses HTDV/HERV-K

ROSWITHA LÖWER, RALF R. TÖNJES, CHRISTINE KORBMACHER, REINHARD KURTH, AND JOHANNES LÖWER*

*Paul Ehrlich Institut, D-63225 Langen, Germany*

Received 20 June 1994/Accepted 5 October 1994

**The human endogenous retrovirus family HTDV/HERV-K codes for the viral particles observed in teratocarcinoma cell lines. Two types of proviral genomes exist; these differ in the presence or absence of a stretch of 292 nucleotides. This sequence comprises the amino-terminal part of the *env* gene, the putative signal peptide, which overlaps in part with the carboxy terminus of the *pol* gene. Type 2 genomes containing this sequence presumably more closely reflect the structure of the infectious, replication-competent retrovirus ancestors of the HERV-K family than do type 1 genomes that lack the sequence. In human teratocarcinoma cell lines, both variants are expressed. Type 1 genomes, in which *pol* and *env* genes are fused, are deficient in splicing. Type 2 transcripts are spliced to subgenomic *env* mRNA and smaller messages. A doubly spliced transcript encodes a short open reading frame, preliminarily designated cORF (R. Löwer, K. Boller, B. Hasenmeier, C. Korbmacher, N. Mueller-Lantzsch, J. Löwer, and R. Kurth, Proc. Natl. Acad. Sci. USA 90:4480–4484). The genomic organization of cORF resembles that of nonprimate lentivirus *rev* genes: the first exon comprises nearly the entire signal peptide of *env*, and the second exon is derived from a different reading frame in the 3′ part of the genome. A nucleolar localization signal, which is also a putative RNA binding domain, as well as a sequence with similarities to the Rev effector domain consensus sequence is present in the first exon. Secondary structure analysis reveals similarities to basic helix-loop-helix proteins. cORF is a small protein with an apparent molecular mass of 14 kDa which accumulates in the nucleolus as has been described for Rev proteins.**

The human genome harbors a variety of endogenous retroviral sequences ranging from single-copy genes to families with several thousand members. Most of these elements have been detected by cross hybridization with conserved *pol* probes derived from the polymerase genes of exogenous animal retroviruses. These *pol* gene homologies have been used as a criterion to classify the elements into human endogenous retrovirus (HERV) class I, related to mammalian type C retroviruses, and class II, with homology to type A, B, and D families as well as to the avian type C retroviruses (for a review see reference 48). HERVs have been tentatively named according to the tRNAs they may use as primers for reverse transcription. The suffix K, for instance, is used when a lysine tRNA can bind to the primer binding site (17). Most HERV elements described so far are highly defective. Although they may be transcribed occasionally, there are very few instances in which these transcripts seem to be translated into proteins (48).

In 1980, we reported the existence (16) of the retrovirus particles designated HTDV (human teratocarcinoma-derived particles) in human teratocarcinoma cell lines, and subsequently we reported their characteristics (19, 21). Recently, we were able to demonstrate that the human endogenous retrovirus family HERV-K codes for these particles (3, 20, 22). HERV-K genomic sequences were first detected in 1986 (33) by using a fragment from the *pol* region of Syrian hamster intracisternal type A particles as a probe in Southern blot analyses. Thus, these sequences can be classified as class II elements. The 9.2-kb genomic full-length proviral clone HERV-K10 has been sequenced and found to be defective (34). Several other reports suggest the existence of at least six different sequence families in the human genome with varying degrees of homology to HERV-K10 and mouse mammary tumor virus (MMTV) *pol* genes, as well as to each other (4, 6, 8, 11, 27).

At present, most of these endogenous retroviruses are only defined by short sequences of their *pol* genes without any information about their putative primer binding sites. Unrelated exogenous retrovirus families often possess the same primer binding site: for example, MMTV and human immunodeficiency virus (HIV) both use as a primer a lysine tRNA having a UUU anticodon. A similar situation can be expected with endogenous retroviruses. The fact that a lysine tRNA primer with a CUU anticodon is described for HERV-K and a lysine tRNA primer with a UUU anticodon is described for the more diverse HERV proviral sequence NMWV 4 (25) already causes some confusion in nomenclature, although HERV-K(CUU) could still be distinguished from HERV-K(UUU). To prevent future ambiguities, we would like to propose the term HTDV/HERV-K instead of HERV-K(CUU) to designate the well-defined elements coding for the teratocarcinoma particles, at least until a more useful nomenclature can be found.

HTDV/HERV-K is a moderately repetitive family with approximately 30 full-length genomes as well as a few copies with vast deletions, for instance, in the *gag* genes. It entered the genome of Old World monkeys after their divergence from New World monkeys (43). In addition, there are several thousand solitary long terminal repeats (18), a feature that has

* Corresponding author. Mailing address: Paul-Ehrlich-Institut, Paul-Ehrlich-Str. 51-59, D-63225 Langen, Germany. Phone: (49) 6103 77-2000. Fax: (49) 6103 77-1252.

also been reported for several other endogenous retrovirus families. These long terminal repeats may be created by homologous recombination between the 5′ and 3′ long terminal repeats, followed by excision of the proviral sequences in between. High copy numbers may be explained not only by infection but also by retrotransposition. There is no evidence yet that the observed virus particles are infectious or that retrotransposition still occurs. These possibilities, however, have to be investigated more thoroughly.

Unlike all other known human endogenous retrovirus sequences, HERV-K genomes possess open reading frames possibly encoding functional Gag, Pol, and Env proteins. We have been able to show that in teratocarcinoma cell lines proviruses with intact *gag* genes are expressed, leading to the production of viral core proteins and virus particles (15, 20).

Expression in *Escherichia coli* of a fusion protein including the *gag* and protease region resulted in the presumably autoproteolytic cleavage of the fusion protein (28, 43). We have been able to demonstrate the formation of HERV-K *gag* precursors as well as their cleavage products in teratocarcinoma cell lines (15), suggesting the existence of a functional viral protease.

The expression pattern of HERV-K in teratocarcinoma cell lines, consisting of full-length as well as singly and multiply spliced subgenomic transcripts, is similar to that seen with complex regulated exogenous retroviruses (see Fig. 1) (20). In lentiviruses and human T-cell lymphotropic virus-related viruses, small multiply spliced mRNAs code for regulatory proteins involved in transcription, translation, and transport of viral RNAs (10). Although HERV-K is not related to these viruses in terms of sequence homology, one of its spliced mRNAs encodes a predicted protein sequence containing a stretch of basic amino acids with similarities to the RNA binding domains of Rex, Rev, and Tat proteins. We had designated this putative protein cORF (central open reading frame). In addition, we were able to demonstrate the existence of a putative subgenomic *env* mRNA in teratocarcinoma cells (20). In this communication we report a detailed analysis of *env* and cORF gene expression.

## MATERIALS AND METHODS

**Cells.** Cells were grown as described previously (22). HeLa (human cervical carcinoma) and HepG2 (liver carcinoma) cell lines were obtained from the American Type Culture Collection. The MRC-5 (human fibroblast), Tera 2, and GH (human teratocarcinoma) cell lines were described earlier (21).

**Primers, probes, and cDNA clones.** The following primers were synthesized with an Applied Biosystems oligonucleotide synthesizer by Martin Selbert at the Paul Ehrlich Institute: P1 (integrase region), CAACAGGAATTCTAATAATTCCC; P3 (*env* region), CAATCTTGTAGAATTCTTTTGCC; and P2 (U5 region), GAGGCTGGCGGGATCCTC (internal *Eco*RI and *Bam*HI restriction sites are underlined). Probes were excised from pcK30*env* (see below) with restriction endonucleases (Gibco BRL) and gel purified prior to labeling reactions performed with an Amersham Megaprime kit. S1 is an *Eco*RI-*Bam*HI fragment (nucleotides [nt] 7562 to 7752), and S2 is an *Xba*I-*Pvu*II fragment (nt 6640 to 6734). pcK30*env* and pcK32*cORF* are cDNA clones isolated from an oligo(dT)-primed GH cDNA library cloned directionally into λZAPII (Stratagene). pcK30*env* was identified with a PCR-generated probe comprising most of the SU region (20), and pcK32*cORF* was identified by an *Sst*II-*Pvu*II subfragment of pcK30*cORF* including the N terminus of the *env* gene. The cDNA clones were characterized by restriction enzyme analysis and sequenced by the dideoxy chain termination method. pcK30*env* is derived from a full-length transcript and comprises the 3′ half of HTDV/HERV-K from the integrase region (nt 6015) to the poly(A) tail. pcK32*cORF* is a cDNA clone which contains the entire cORF transcript. As we have not yet determined the exact transcription start site, we do not know whether the 5′ end of pcK32*cORF* represents the authentic cap site or whether cDNA synthesis had stopped at a premature site due to secondary RNA structure.

**RT-PCR.** mRNA was prepared as described previously (22). Reverse transcription PCR (RT-PCR) was performed with poly(A)-selected RNA from GH cells by using a Perkin-Elmer Cetus kit according to the manufacturer's instructions. Reverse transcription was carried out for 10 min at 37°C and then for 30 min at 42°C. For amplification the following cycle scheme was used (Biometra thermo cycler): 1.5 min at 94°C, 1 min at 55°C, and 1.5 min at 72°C for 40 cycles and 10 min at 72°C.

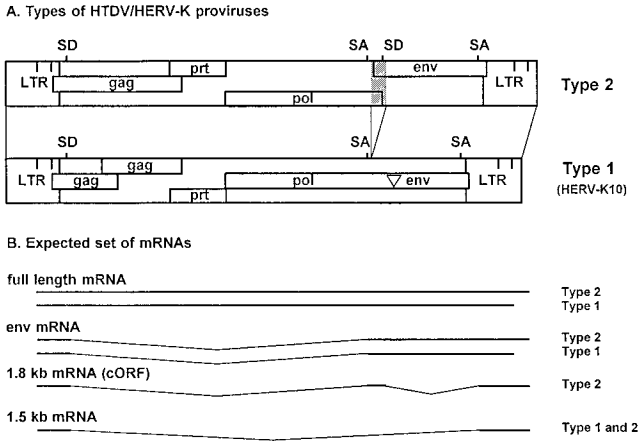


FIG. 1. In the human genome, two types of full-length HTDV/HERV-K proviruses which differ by the presence (type 2) or absence (type 1) of a stretch of 292 nt (shadowed) at the *pol*/*env* boundary exist (A). A type 1 clone (HERV-K10) has been completely sequenced (34). Its peculiarities (a split *gag* gene and a stop codon [open triangle] in *env*) have been depicted, though they may not be present in all type 1 proviruses. Two sets of splice donor sites (SD) and SAs have been identified by analysis of the 1.8- and 1.5-kb transcripts present in oligo(dT)-selected mRNA from particle-producing teratocarcinoma cell lines (20). A set of mRNAs expected on the basis of the presence of these splice sites in the two provirus types has been compiled (B).

**Northern (RNA) and Southern blots.** PCR products were run on 1.2% agarose gels, and then they were blotted on nylon membranes (Amersham) by employing the Posiblot pressure blot/UV Stratalinker system (Stratagene) according to standard procedures similar to the procedure described previously for Northern blots (22). Hybridization was performed with 500,000 cpm of labeled probe per ml. Hybridization conditions were adjusted to a stringency of 25°C below melting temperature, and washing was performed to reach a stringency of 10 to 15°C below melting temperature (melting temperature was calculated according to the equation described in reference 47).

**cORF antiserum.** For prokaryotic expression of cORF the pMAL-c2 system was used according to the manufacturer's instructions (New England Biolabs). The coding region of cORF was cloned in frame into the *Eco*RI cloning site downstream of the *malE* gene, giving rise to a bacterial maltose-binding protein–cORF fusion protein. The fusion protein was purified by affinity chromatography on amylose resin and subsequently by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Excised maltose-binding protein–cORF bands were eluted and used for immunization of rabbits with Freund's incomplete adjuvant. Animals received boosters twice before antiserum was drawn.

**Radioimmunoprecipitation assays and immunofluorescence.** Radioimmunoprecipitation assays were performed as described previously (15). Briefly, cells were labeled with 400 μCi of [$^{35}$S]methionine and [$^{35}$S]cysteine (Amersham in vitro cell labeling mix) in Dulbecco modified Eagle medium lacking these amino acids, lysed, and precipitated with rabbit anti-cORF antibody or preimmune serum. Precipitates were washed and analyzed by SDS–15% PAGE. Immunofluorescence studies were performed as described previously (3). Preparations were examined with a laser scan microscope (MCR; Bio-Rad; attached to a Zeiss Axiovert fluorescence microscope).

**Nucleotide sequence accession numbers.** The EMBL accession numbers for pcK30*env* and pcK32*cORF* are X82272 and X82271, respectively.

## RESULTS

**Analysis of the *pol*/*env* boundary.** The sequenced full-length genomic clone HERV-K10 represents a defective provirus (34). Besides displaying a split *gag* gene, it also lacks a stop codon between the *pol* gene and the putative *env* gene and it has an early stop codon in *env* (type 1 in Fig. 1A). The finding that other partially sequenced provirus clones (e.g., genomic clone K8) had an extra stretch of 290 nt at the boundary of *pol* and *env* prompted Ono et al. to simply insert this stretch into the type 1 sequence of HERV-K10 at nt 6500 and to publish the composed sequence HERV-K10+ (34). Such a provirus

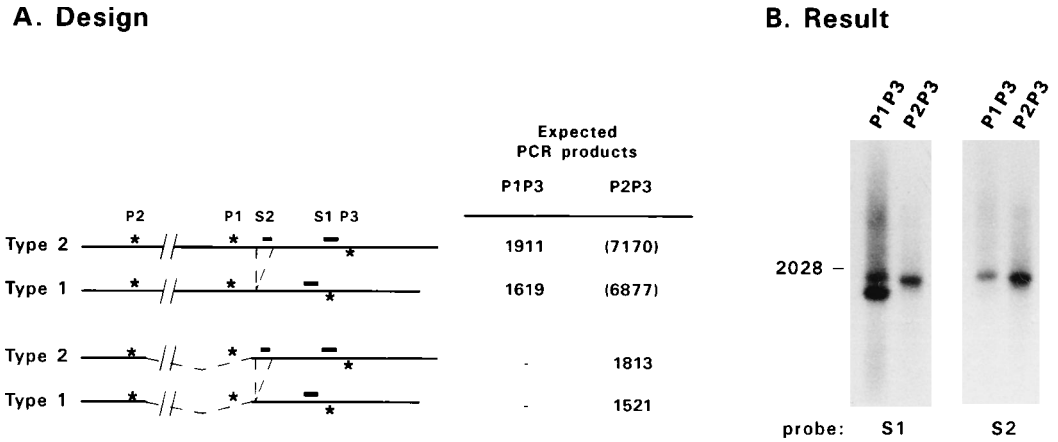


FIG. 2. RT-PCRs were designed to discriminate between full-length and subgenomic *env* transcripts present in oligo(dT)-selected mRNA from the teratocarcinoma cell line GH. (A) Primer 1 (P1) is located in the integrase region of the *pol* gene, primer 2 (P2) is located in the 5′ U5 region, and primer 3 (P3) is located in the surface part of the *env* gene. The primer pair P1-P3 will only amplify cDNA derived from full-length transcripts, because the *gag*/*pol* intron is spliced out in *env* mRNAs. Subgenomic *env* transcripts are targets for the primer pair P2-P3. Primers are indicated by asterisks. The lengths of the expected RT-PCR products (in nucleotides) are listed. Products in parentheses were too long to be detectable by PCR. (B) RT-PCRs were analyzed by Southern blotting with probes specific for type 1 and type 2 proviruses. Probe S1 was a short fragment from the surface region which hybridized to both types. Probe S2 was specific for type 2 genomes. Probes are depicted as bars in panel A. The exact localization of primers and probes is described in Materials and Methods.

(type 2-like) would now provide a termination codon for the *pol* gene and a separate reading frame for the *env* gene. We will compare the HERV-K10+ sequence with the expressed HTDV/HERV-K sequences, in which three important nucleotide insertions have been detected previously (20), the first, at nt 1749, leading to an intact *gag* gene and the other two, in the extra stretch (then 292 nt in length) (type 2 in Fig. 1A), also resulting in a stop codon for *pol* and an open reading frame for *env* (discussed below; see also Fig. 3). Numbering will include these insertions.

We have already shown that HTDV/HERV-K type 1 is expressed in teratocarcinoma cell lines by analyzing the *pol*/*env* region in an RT-PCR (20). To determine whether type 2 genomes are also expressed in particle-producing cell lines, to elucidate their exact genomic organization at the *pol*/*env* boundary, and to analyze the expected subgenomic *env* mRNAs (Fig. 1B), we performed RT-PCRs with mRNA from GH cells. The primer pair P1 (integrase region of the *pol* gene) and P3 (*env* region) binds to full-length transcripts (Fig. 2A). Products generated by RT-PCR comprise 1,911 nt when the transcripts are derived from type 2 proviral genomes and 1,619 nt when the transcripts are derived from type 1 proviral genomes. RT-PCR with the primer pair P2 (U5 region) and P3 will efficiently amplify only sequences from subgenomic *env* transcripts, the predicted lengths being 1,813 and 1,521 nt.

Specificity of the PCR product was confirmed by Southern blot analysis. As can be seen in Fig. 2B, the predicted 1.6- and 1.9-kb fragments hybridized to the S1 probe specific for the outer membrane *env* region, indicating that both types of proviral structures were present in the full-length mRNA populations. Under semiquantitative conditions, there was an excess of type 1 full-length transcripts compared with type 2 full-length transcripts in GH cells. Similarly, probe S2 only hybridized to the PCR product derived from type 2 genomes, as this sequence is not present in a type 1 provirus. RT-PCR control reactions omitting reverse transcriptase were negative (data not shown).

Amplification from subgenomic *env* transcripts with primers P2 and P3 generated the 1.8-kb fragment indicative of a type 2 provirus, which was detected by the S1 probe as well as the S2 probe. The 1.5-kb fragment predicted for type 1 genomes was missing, strongly suggesting that type 1 full-length transcripts are not spliced to subgenomic *env* mRNAs. Therefore, we conclude that although both types of HTDV/HERV-K genomes are primarily transcribed, only type 2 proviruses harboring the 292 nt may be targets for posttranscriptional modifications such as splicing.

**Definition of the splice acceptor site (SA) for the subgenomic *env* mRNA and description of the *env* protein leader sequence.** For further analysis, PCR products were cloned and sequenced. Sequences derived from type 1 full-length transcripts have already been described (20). They closely resemble genomic clones HERV-K10 and K18 (33, 34). PCR products derived from the subgenomic *env* transcripts turned out to have the same sequence as the cDNA clone pcK30*env* isolated from a GH cDNA library. pcK30*env* encompasses the 3′ part of a viral genome (nt 6015 to 9381), including the extra 292 nt at the *pol*/*env* boundary. Thus, we would like to refer to this cDNA clone as a true type 2 prototype, while HERV-K10+ mostly displays the sequence of a defective type 1 provirus.

Sequence alignment reveals that in subgenomic *env* transcripts the same splice donor site (nt 1072) and SA (nt 6434) are used as are used for the doubly spliced cORF message (20). Comparison of type 1 and type 2 sequences (K10 versus pcK30; see Fig. 3) shows that there is a remarkable accumulation of nucleotide exchanges in the splice acceptor region (nine mutations in the 100 nt from nt 6410 to 6510) which is not seen in other gene segments (zero to a maximum of four mutations per 100 nt in the *env* region).

The abrogation of a stop codon at nt 6496 and two insertions at nt 6614 and 6689 result in remarkable alterations in the 5′ region of the *env* gene encoded by pcK30*env*, yielding an Env protein with a signal peptide (SP) much longer than that predicted for the composed sequence HERVK10+ (34). Downstream from the SP there are only a few point mutations leading to amino acid changes.

SPs of retroviruses are identified by the existence of a hydrophobic core region at the amino terminus of the Env protein, preceded by an amino-terminal stretch (n) of charged amino acids and followed by the signal peptidase cleavage site,

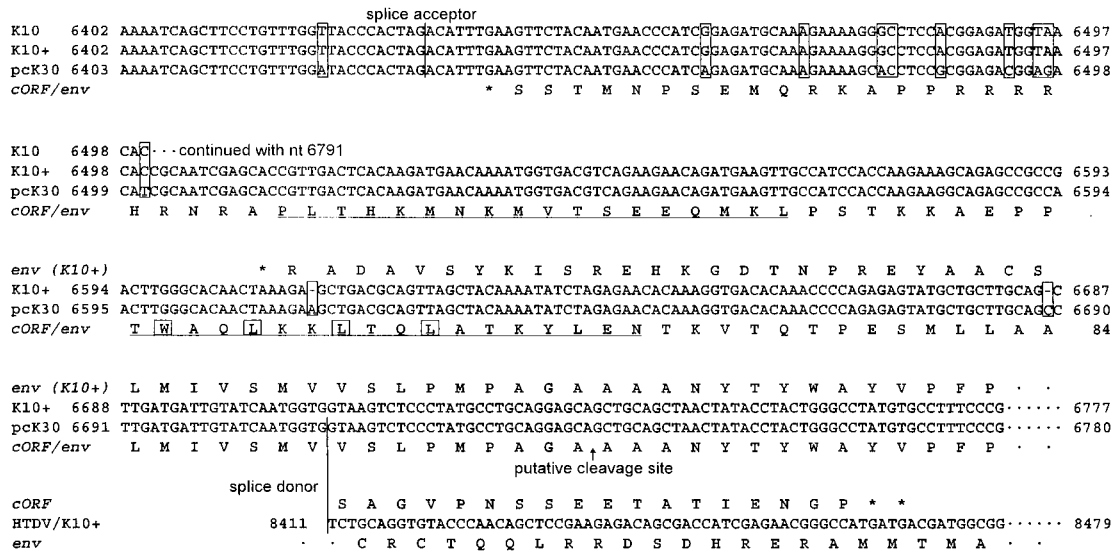


FIG. 3. Nucleotide and amino acid sequence comparison of the HTDV/HERV-K type 2 clone pcK30 with the type 1 clone K10 (34) and the composed sequence K10+ (34) at the *pol*/*env* boundary and in the region of the second exon of cORF. *env* and cORF mRNAs were spliced from the same proviral full-length transcripts. Two additional nucleotide insertions (the first after nt 6613 and the second after nt 6688, numbering according to the HERV-K10+ sequence [34]) led to a start of the *env* gene different from that predicted for HERV-K10+. Nucleotide exchanges and insertions are boxed. Asterisks are used to denote stop codons. Putative amphipathic helices (underlined) were identified by the HIBIO Prosis protein analysis system (Hitachi). The WxxLxxLxxL motif similar to the Rev effector domain consensus sequence is boxed.

which is supposed to fulfill the "−3,−1" rule established by von Heijne (44, 46). In many retrovirus families the n domain varies extremely in length and charge (35). The putative HTDV/HERV-K type 2 SP has a long, highly positively charged amino-terminal *env* region. In this respect it resembles the SP identified for the MMTV Env protein (13). This observation is complemented by a comparison of the hydrophobicity profiles of HTDV/HERV-K and MMTV Env proteins, which are very similar also in the surface and transmembrane domains (see Fig. 4). The first signal peptidase cleavage site after the hydrophobic core of HTDV/HERV-K Env SP that follows the −3,−1 rule is at position 96.

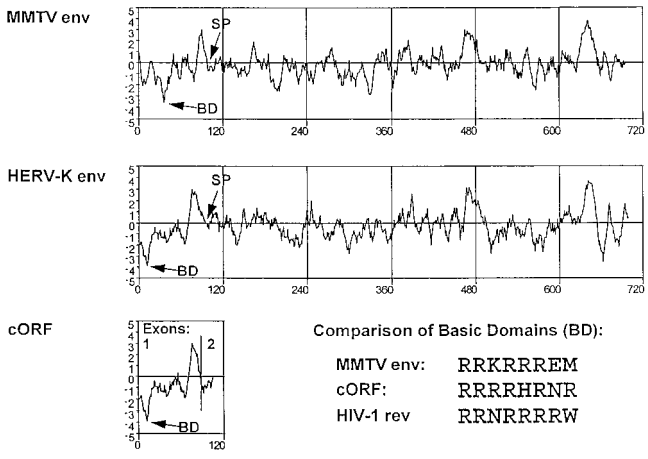


FIG. 4. Comparison of the hydrophobicity profiles of cORF, HTDV/HERV-K *env*, and MMTV *env* genes with the HIBIO Prosis system. The position of the SP cleavage site (SP) for HTDV/HERV-K was estimated by using the −3,−1 rule (44, 46), and that for MMTV was determined by Henderson et al. (13). In addition, the sequences of the basic domains (BD) present in the SP of HTDV/HERV-K and MMTV are compared with that of the BD of HIV type 1 (HIV-1) *rev*.

**cORF is part of the amino-terminal region of the *env* gene and is the predominant splice product.** In human immunodeficiency viruses (HIV), the first exons of the *tat* and *rev* genes are located in the region between the *pol* and *env* genes but in different reading frames (for a review see reference 37). In the case of HTDV/HERV-K, sequencing revealed that for the formation of cORF and *env* RNAs the same splice donor and acceptor sites are used to excise the *gag*/*pol* intron. To determine the location of cORF in relation to *env*, the nucleotide sequences as well as the putative amino acid sequences of the two genes were compared. cORF and *env* reading frames start with an identical codon. The first exon of cORF comprises the amino-terminal *env* sequence from amino acid (aa) 1 to 87, where it is spliced to the second exon in the 3′ part of the viral genome but into a reading frame different from that of the *env* gene (3′ SA [Fig. 1]; sequences are presented in Fig. 3). Sequencing of a cDNA clone (pcK32*cORF*) isolated from the GH cDNA library confirmed that the doubly spliced cORF mRNA was derived from a proviral genome similar to that of pcK30*env*. In both cDNA clones polyadenylation starts at nt 9381, 63 nt downstream of the 3′ end of the R region predicted by Ono (33).

In lentivirus gene expression, a complex regulatory mechanism governs the occurrence of spliced and unspliced mRNAs in infected cell lines. To determine the relative amount of spliced HTDV/HERV-K transcripts versus that of unspliced transcripts in teratocarcinoma cells, Northern blot analysis was performed with a probe derived from the surface region of the *env* gene (S1) and the amino-terminal SP-specific S2 probe, which also hybridizes to the cORF mRNA. Hybridization (Fig. 5) with the S1 probe illustrated that a major proportion of mRNA was unspliced, since the 3.3-kb subgenomic message was stained more faintly than the full-length transcripts. Although it is difficult to determine the exact sizes of full-length transcripts, we estimate that the bulk of the full-length RNA hybridizing to probe S1 was slightly smaller in size than the RNA band detected by the type 2-specific S2 probe, indicating

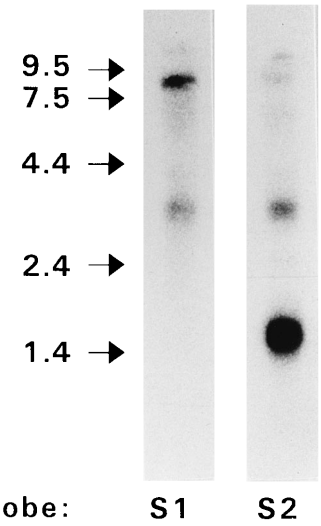


FIG. 5. Northern blot analysis of oligo(dT)-selected RNA from the particle-producing teratocarcinoma cell line GH with probes S1 and S2. Localization of the probes is indicated by bars in Fig. 2. mRNA content was normalized to equal amounts of *env* transcripts.

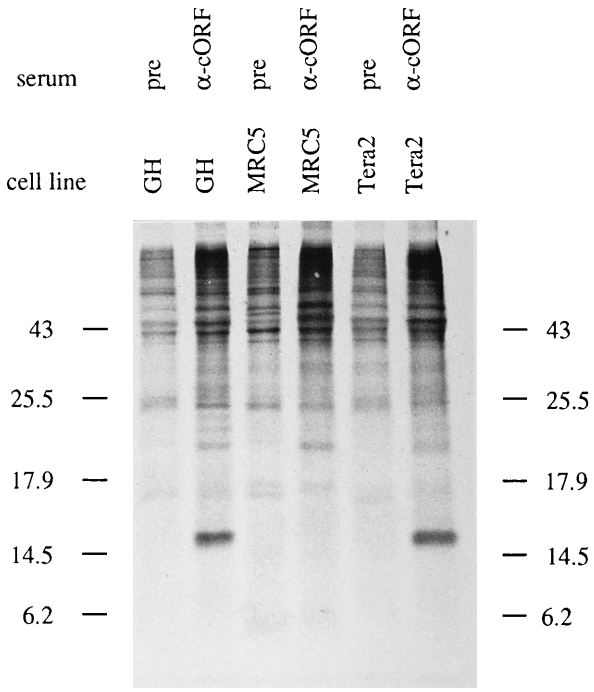


FIG. 6. Radioimmunoprecipitation assays with labeled cell lysates and pre-immune serum (pre) or rabbit anti-cORF antibodies (α-cORF). Cell lines used were GH and Tera 2 (both teratocarcinomas) and MRC-5 (normal fibroblasts).

that the quantitatively dominating transcripts seen with S1 were probably derived from splicing-deficient type 1 proviruses. This is in agreement with the observation that the PCR amplification performed with full-length transcripts (Fig. 2) indicated an excess of type 1 transcripts.

Hybridization with S2 quite obviously showed that the expression pattern of HERV-K type 2 was shifted extremely from the prevalence of full-length transcripts to the prevalence of spliced mRNA species, the most prominent band being the doubly spliced cORF message. The shortage of type 2 full-length mRNA was also visible in the imbalanced generation of RT-PCR products shown in Fig. 2. Two faint bands at approximately 4 and 12 kb (Fig. 5) may have been generated by defective proviral genomes, by skipping of the viral polyadenylation signal, or by initiation of transcripts from cellular promoters.

These results support the hypothesis that both types of genomes are expressed. Type 1 genomes are not spliced, and they accumulate as full-length transcripts, whereas type 2 transcripts are subject to extensive splicing events.

**cORF shares the Rev/Rex activation domain consensus sequence.** Mutational analysis of Rev and Rex proteins has identified two domains involved in their function (7, 23). (i) N-terminal Rev sequences include an arginine-rich nucleolar localization motif, which also serves as an RNA binding domain. It is flanked by a region involved in protein multimerization. (ii) A Rev effector domain consensus sequence has been determined to exist in the C-terminal region by detailed mutational analysis. It consists of a cluster of three to four leucine (L) or other hydrophobic amino acids with a critical spacing which is highly conserved among primate and even some nonprimate lentivirus families. The effector domains of simian lentivirus Rev proteins and those of Rex proteins can replace the HIV Rev domain, just as the effector domains of feline and equine immunodeficiency viruses can, although the latter exhibit a somewhat different structure (24).

HTDV/HERV-K cORF has sequence similarities not only with the arginine-rich motif involved in RNA binding and nucleolar translocation but also with the conserved Rev effector domain consensus sequence LxxLxxLxL (Fig. 3; see Discussion). We are currently investigating whether this domain can replace the Rev activation domain.

**The 14-kDa cORF protein is expressed in teratocarcinoma cell lines, and it accumulates in the nucleolus.** We had reported that the cORF message has the coding capacity for a putative 12-kDa protein with a basic domain (20). In in vitro translation assays (data not shown), a 14-kDa protein was produced. The somewhat slower migration rate may be a consequence of the highly positive charge of the protein, a conclusion also drawn for the bovine immunodeficiency virus Rev protein (32).

The coding sequences of cORF were cloned into the prokaryotic expression plasmid pMAL and expressed as a 36-kDa maltose-binding protein–cORF fusion protein in *E. coli*. Production of the fusion protein was induced with IPTG (isopropyl-β-D-thiogalactopyranoside), and purification from bacterial lysates was performed by affinity chromatography on amylose resin. This fusion protein was used to immunize rabbits.

Radioimmunoprecipitation assays with human cell lysates and the rabbit anti-cORF serum were employed to investigate the expression of cORF protein. Figure 6 exemplifies that the anti-cORF serum specifically precipitates a 14-kDa protein from teratocarcinoma cell lysates (GH and Tera 2) but not from normal human fibroblasts (MRC-5). Preliminary experiments suggest that in contrast to the HIV and bovine immunodeficiency virus Rev proteins, cORF seems not to be phosphorylated.

Immunofluorescence studies have so far revealed that cORF is expressed only in teratocarcinoma cells and not in other lines, for instance, normal human fibroblasts (MRC-5), HeLa cells, and HepG2. In the case of the particle-producing cell line GH (Fig. 7) cORF is present in nearly every cell, whereas in the case of Tera 2, a line in which particles are extremely rare,

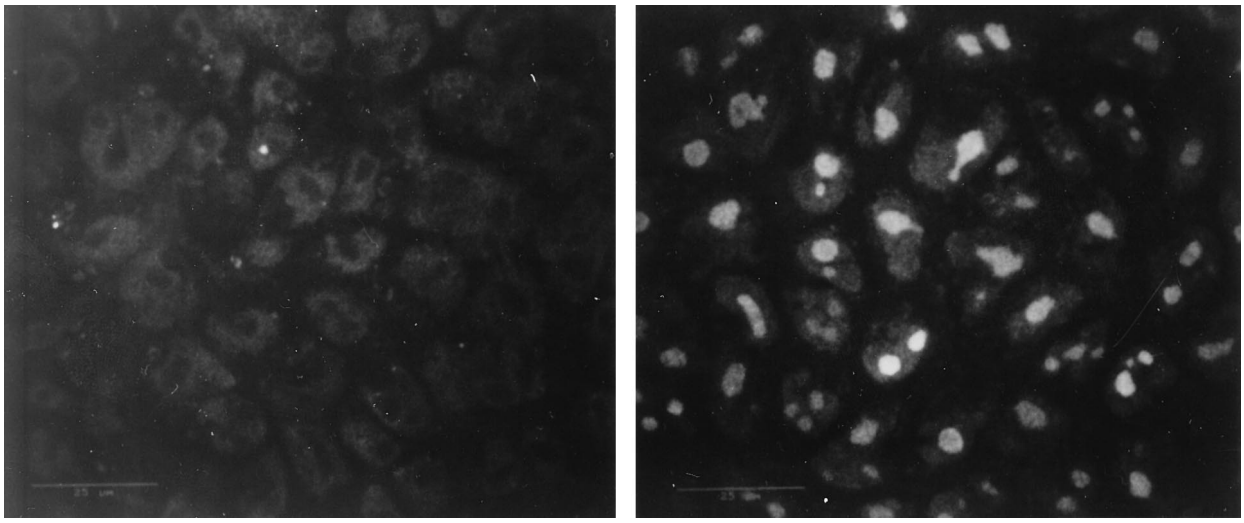

FIG. 7. Immunofluorescence with GH cells and preimmune serum or rabbit anti-cORF serum. Bars, 25 μm.

cORF is expressed in only a few cells (data not shown). cORF is a nuclear protein which resides in the nucleolus (Fig. 7), as was expected from the presence of the consensus nucleolar localization signal RRxR.

## DISCUSSION

**Infectivity of particles coded for by HTDV/HERV-K.** In 1986, Ono described two classes of genomic HERV-K sequences which differ in size by 290 nt and postulated that longer genomic clones like K8 and K22 are prototypes of this endogenous retrovirus family. Clone K10, which he sequenced, turned out to be defective because of a split *gag* gene, a vast deletion in the *pol*/*env* region, and an early stop codon in *env*. We have been able to demonstrate that in teratocarcinoma cell lines both variants, called HTDV/HERV-K types 1 and 2, are expressed as full-length transcripts. As we have never observed a split *gag* gene in mRNA preparations (20, 43), we conclude that most expressed proviral genomes, whether they are type 1 or type 2, possess an intact *gag* gene which is a prerequisite for particle formation. Accordingly, it has already been shown that, in teratocarcinoma cells, retrovirus particles which contain core proteins coded for by HTDV/HERV-K can be seen (3, 20).

Although a full-length provirus of type 2 has not yet been cloned and sequenced, the data presented earlier (20, 28) and in this paper hint at the possibility that type 2 proviral genomes that have open reading frames for all genes necessary for replication exist. However, replication-competent, infectious viruses have not yet been demonstrated to exist, not even in the particle-producing teratocarcinoma cell lines. There are several explanations for this.

(i) In our cocultivation studies (21), the permissive cell type may have been missed, as retroviruses sometimes have a very narrow host range. There are two virus groups related to HTDV/HERV-K by sequence homology and genome organization: MMTV (42) and Jaagsiekte sheep retrovirus (JSRV) (49). The host ranges of both viruses are also very limited, the main source of virus being naturally infected animals. An explanation for the difficulties in growing MMTV and JSRV in tissue culture may be their existence in an endogenous form as well as an exogenous form. Judging from these similarities between HTDV/HERV-K and MMTV as well as JSRV, we cannot completely exclude the possibility of an exogenous form of HTDV/HERV-K.

(ii) Electron microscopic investigations suggest that the particles observed in teratocarcinoma cells may not be infectious. The absence of an electron lucent space between the virus core and membrane and the very rare detection of particles with collapsed cores (15) indicate that there is at least a deficiency in virus maturation. Maturation can be impaired by mutations in the *gag*, protease, and *env* genes, which are all involved in virus assembly and formation of mature cores after the budding of virions. Studies with cloned individual viral genes suggest the existence of functional Gag and protease proteins (15, 28, 43). Characterization of polymerase and *env* gene products has turned out to be more difficult. Functional proteins may be rare: it is very difficult to detect reverse transcriptase activity in particle preparations (19), and it is impossible to unequivocally demonstrate the presence of retroviral *env* structures on the particles by electron microscopy (15).

(iii) The presence of transcripts with open reading frames per se is necessary for translation, but it may not be sufficient. Evidence that mRNA stability and ribosomal accessibility are regulated by positive and/or negative *cis*- and *trans*-acting elements is emerging. But even if we could prove that in principle all viral genes are present as functional proteins, there might be a situation of dominant negative complementation: a variety of nondefective and defective proteins may be expressed at the same time from different and sometimes deficient members of the multiple-copy-number family, their joint incorporation into particles probably resulting in a noninfectious phenotype. Further investigations of expressed and cloned full-length genomes should answer this question.

(iv) Another reason for noninfectivity could have been the lack of a packaging signal on the viral RNA. Together with Boyd (4) we have addressed this question, and we have been able to demonstrate that HTDV/HERV-K mRNA is indeed packaged into virus particles. In teratocarcinoma cell lines, type 2 proviruses are predominantly spliced, whereas the bulk of the full-length transcripts which can also serve as retrovirus genomes consist of type 1 proviruses defective in *env* splicing. Packaging of such genomes would result in noninfectious offspring after a first round of replication. We are currently

investigating which genome types are present in particle preparations.

**Splicing strongly depends on the presence of the cORF sequence.** An unexpected feature of type 1 genomes is their inability to be spliced. Like most eukaryotic genes, most retrovirus genes contain at least one intron, the *gag/pol* region, which has to be excised, for example, for *env* expression. But unlike most eukaryotic transcripts, which are spliced to completion, retroviral mRNA has to be spliced in a regulated manner, giving rise to coordinated quantities of spliced and unspliced products. This regulation is perhaps more adequately compared to the regulatory mechanisms involved in alternative mRNA splicing (26). Relative efficiencies of splice site choices may be influenced by the presence or absence and/or the concentration of *cis*- and *trans*-acting factors involved in the splicing procedure.

The presence of suboptimal splice sites is a prerequisite for the preservation of *gag/pol* introns. Recently, it was shown that in an HIV-derived construct a poor branchpoint region and a poor polypyrimidine tract are responsible for low splicing efficiency (41). This region upstream of the SA is highly conserved in all HIV provirus clones studied. Sequences in the exon downstream of the SA can also influence splicing (12, 14, 36). In the HTDV/HERV-K SA, the consensus sequence is not well conserved and the polypyrimidine tract is often disrupted. This observation is consistent with the requirement for suboptimal splice sites. However, the main difference between type 1 and type 2 proviruses is the presence or absence of the 292-bp stretch located in the exon following the SA, including an accumulation of nucleotide changes immediately downstream of the SA. These changes may well influence the acceptance of this splice site: in type 1 genomes, missing this sequence, splicing does not occur. Secondary structure analysis suggests that in type 2 genomes the SA may be exposed on a loop, while it is probably hidden in a stem in type 1 sequences (data not shown).

In the case of the Rous sarcoma virus system, balance of splicing has been shown to depend not only on the existence of suboptimal 3′ splice sites but also on *cis*-acting sequences in the *gag* gene which negatively regulate splicing (NRS) (1). Lentiviruses are regulated in an even more complex manner. Unspliced and singly spliced transcripts are sequestered in the nucleus because of *cis*-acting sequences which are present in the viral *gag*, *pol*, and *env* regions and act as multiple, independent elements of nuclear retention and instability (INS), probably in concert with cellular factors (10). The viral regulatory protein Rev can overcome this inhibitory effect by binding to the Rev-responsive element (30, 39). HTDV/HERV-K genomes show an obvious imbalance of spliced mRNAs versus unspliced mRNAs. While transcripts of type 1 genomes are not spliced at all, doubly spliced mRNA or messages lacking any intron are derived from type 2 genomes and accumulate in the cytoplasm; singly spliced and, even more so, unspliced transcripts of type 2 genomes are rare. This phenomenon may be explained by the absence of NRS or INS sequences or the presence of sequences augmenting splice site usage in type 2 genomes. It will be interesting to explore whether in HTDV/HERV-K type 2 the amount of full-length mRNA in the cytoplasm may be enhanced by exchanging such sequences with corresponding nucleotide stretches from type 1 genomes. Vice versa, splicing of type 1 transcripts may be facilitated by the replacement of negative regulating elements by type 2 sequences. *trans*-acting mechanisms depending on a Rev-related protein and its response element may also be considered. In human T-cell lymphotropic virus, Rex is directly involved in suppression of splicing by binding to the Rex-responsive element (RxRE) which overlaps the splice donor (7). Alterations in the RxRE can disturb the balance of gene expression. If a Rev/Rex function can be assigned to cORF, the identification of a responsive element will be of prime importance.

**cORF has similarities to *rev*-related genes.** Type 2 is the prototype of a putative functional human endogenous retrovirus with the expected organization of viral genes. As a rule, sequence homology between *env* genes of different retrovirus families is poor. HTDV/HERV-K and MMTV *env* genes have only 23% identity at the amino acid level, but there is a striking overall similarity in the *env* hydrophobicity profiles and in the structures of the SP (Fig. 4). In both virus groups a very large region with a high incidence of charged residues (the n region) precedes the hydrophobic core, which represents the essential signal sequence. Such characteristics have also been described for the SP of the avian retrovirus Rous sarcoma virus, nonprimate lentiviruses, spuma viruses, and JSRV (9, 35). In HTDV/HERV-K and MMTV SP, there is a prevalence of positively charged amino acids in the n region (HTDV/HERV-K SP net charge of +14). The highest degree of positive net charge seen in lentiviruses is +8 for ovine maedi-visna virus SP (9), while for eukaryotic SP a mean value of +1.7 has been calculated (45).

Functional analysis of the *env* leader sequence (9) demonstrated an unusual tolerance to nucleotide changes, especially in the n region, suggesting that this region is not involved in the translocation process of the nascent Env protein chain. Therefore, alternative functions might be associated with this sequence. In nonprimate lentiviruses like equine immunodeficiency virus (2), bovine immunodeficiency virus (31), and caprine arthritis-encephalitis virus (38), the first coding exon of *rev* coincides with part of the SP n region. A splice donor present 5′ of the SP hydrophobic core sequence joins the first exon to the second exon of *rev*, which is located in the transmembrane region of *env* but in a different reading frame. As the SP is cleaved already shortly after translocation, mature Env and Rev proteins do not share any sequences. Preliminary results with bovine immunodeficiency virus indicate that *env* and *rev* transcripts do not use the same initiator AUG start codon: *env* translation may initiate at a start codon near the core signal sequence (31).

Intriguingly, HTDV/HERV-K cORF exhibits some similarities to the structures of lentivirus *rev* and *rev*-related genes, although there are major differences in exon distribution. Like the first exon of nonprimate lentivirus *rev* genes, the first coding exon of cORF is in the same reading frame as the *env* gene, but in contrast to the exon of lentiviruses it comprises the entire n region of the SP. The splice donor is located in the hydrophobic core region and thus is significantly closer to the putative cleavage site of the signal peptidase. But, in lentivirus *rev* genes, the second exon contains the domains involved in function. In HTDV/HERV-K cORF transcripts, it is the first exon which codes for motifs with sequence similarities to these Rev domains. The second exon, located at the 3′ end of the *env* region in a different reading frame, is very short (18 aa).

Two important protein domains have been associated with Rev function. (i) The N-terminal arginine-rich basic motif is involved in RNA binding, nucleolar translocation, and multimerization. cORF also harbors an N-terminal arginine-rich basic sequence (Fig. 3). The fact that the cORF protein accumulates in the nucleolus (Fig. 7), exerting a subcellular distribution similar to that which has been described for Rev-related proteins (7), indicates a functional nucleolar localization signal. Currently, we are investigating whether we can assign this function to the basic domain and whether we can demonstrate

RNA binding and multimerization. Preliminary computer-based secondary structure analysis of the region which follows the arginine stretch (aa 22 to 66) in cORF has revealed the existence of a putative helix-loop-helix motif: helix 1 comprises aa 22 to 39, helix 2 comprises aa 49 to 66, and the two are separated by a 9-aa loop bordered by prolines. A similar structure has been described for DNA binding and dimerization (29). Thus, this cORF basic helix-loop-helix structure may be involved in protein-protein interactions such as multimerization on a DNA template or perhaps on an RNA template, for instance, in a double-stranded stem structure.

(ii) The second domain involved in Rev function is the C-terminal leucine-rich region, although the exact molecular mechanism is not yet known (7). This Rev effector domain consensus sequence, LxxLxxLxL, exerts some highly conserved features: the spacing of leucine residues, a frequent replacement of the first L by another hydrophobic residue, a prevalence of hydrophilic xx residues in the second Lxx repeat, and a threonine residue often following the third leucine. In cORF, there is a region (WAQLKKLTQL, Fig. 3) showing the characteristics of a Rev consensus sequence (W replacing the first L, KK as hydrophilic residues in the second Lxx repeat, a T following the third L), although there is a slight variation in spacing. It will be of great interest to investigate a putative Rev-like function for this sequence.

These similarities between the HTDV/HERV-K cORF and *rev* genes are puzzling because there is no obvious sequence homology between the virus families, either at the nucleotide level or in the organization of the *pol*/*env* boundary. Characteristic of lentiviruses is the existence of central open reading frames between *pol* and *env* genes coding for regulatory proteins. *pol* and *env* genes overlap in HTDV/HERV-K as well as in MMTV, JSRV, and Mason Pfizer monkey virus-related viruses (40). There are no indications of additional regulatory genes like *tat*, *rev*, *vpu*, *vif*, etc., although there is an extra open reading frame (X) in JSRV which overlaps *pol* (49). Nevertheless, these viruses must have *cis*- or *trans*-acting sequences which ensure expression of unspliced or alternatively spliced gene products. Recently such *cis*-acting sequences replacing Rev in function were detected in Mason Pfizer monkey virus (5). If cORF turns out to be an ancient form of modern Rev proteins, it will be an interesting system for studying convergent evolution of proteins with similar functions and for elucidating the evolution of retroviruses in general, because HTDV/HERV-K, aged approximately 30 million years, is fairly old compared with the lentiviruses, which appeared in recent historical times.

## ACKNOWLEDGMENTS

We thank H. Hasche, U. Held, and G. Braun for excellent technical assistance and I. Plumbaum for help in manuscript preparation. We are also grateful to K. Boller, K. Chichutek, J. Denner, and U. Fröhlich for stimulating discussions.

In part, this work was financially supported by a donation from the Heinz Kuthe de Mouson Foundation to R.K.

## REFERENCES

1. **Arrigo, S., and K. Beemon.** 1988. Regulation of Rous sarcoma virus RNA splicing and stability. Mol. Cell. Biol. **8:**4858–4867.
2. **Beisel, C. E., J. F. Edwards, L. L. Dunn, and N. R. Rice.** 1993. Analysis of multiple mRNAs from pathogenic equine infectious anemia virus (EIAV) in an acutely infected horse reveals a novel protein, Ttm, derived from the carboxy terminus of the EIAV transmembrane protein. J. Virol. **67:**832–842.
3. **Boller, K., H. König, M. Sauter, N. Mueller-Lantzsch, R. Löwer, J. Löwer, and R. Kurth.** 1993. Evidence that HERV-K is the endogenous retrovirus sequence that codes for the human teratocarcinoma-derived retrovirus HTDV. Virology **196:**349–353.
4. **Boyd, M., et al.** Unpublished results.
5. **Bray, M., S. Prasad, J. W. Dubay, E. Hunter, K. Jeang, D. Rekosh, and M. Hammerskjöld.** 1994. A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication REV-independent. Proc. Natl. Acad. Sci. USA **91:**1256–1260.
6. **Callahan, R., W. Drohan, S. Tronick, and J. Schlom.** 1982. Detection and cloning of human DNA sequences related to the mouse mammary tumor virus genome. Proc. Natl. Acad. Sci. USA **79:**5503–5507.
7. **Cullen, B. R.** 1992. Mechanism of action of regulatory proteins encoded by complex retroviruses. Microbiol. Rev. **56:**375–394.
8. **Deen, K. C., and R. W. Sweet.** 1986. Murine mammary tumor virus *pol*-related sequences in human DNA: characterization and sequence comparison with the complete murine mammary tumor virus *pol* gene. J. Virol. **57:**422–432.
9. **Ellerbrok, H., L. D'Auriol, C. Vaquero, and M. Sitbon.** 1992. Functional tolerance of the human immunodeficiency virus type 1 envelope signal peptide to mutations in the amino-terminal and hydrophobic regions. J. Virol. **66:**5114–5118.
10. **Felber, B. K., and G. N. Pavlakis.** 1993. Molecular biology of HIV-1: positive and negative regulatory elements important for virus expression. AIDS **7:**51–62.
11. **Franklin, G. C., S. C. Chretien, I. M. Hanson, H. Rochefort, F. E. B. May, and B. R. Westley.** 1988. Expression of human sequences related to those of mouse mammary tumor virus. J. Virol. **62:**1203–1210.
12. **Fu, X.-D., R. A. Katz, A. M. Skalka, and T. Maniatis.** 1991. The role of branchpoint and 3′-exon sequences in the control of balanced splicing of avian retrovirus RNA. Genes Dev. **5:**211–220.
13. **Henderson, L. E., R. Sowder, G. Smythers, and S. Oroszlan.** 1983. Terminal amino acid sequences and proteolytic cleavage sites of mouse mammary tumor virus *env* gene products. J. Virol. **48:**314–319.
14. **Katz, R. A., and A. M. Skalka.** 1990. Control of retroviral RNA splicing through maintenance of suboptimal processing signals. Mol. Cell. Biol. **10:**696–704.
15. **Korbmacher, C., et al.** Unpublished data.
16. **Kurth, R., R. Löwer, J. Löwer, R. Harzmann, R. Pfeiffer, C. G. Schmidt, J. Fogh, and H. Frank.** 1980. Oncornavirus synthesis in human teratocarcinoma cultures and an increased antiviral immune reactivity in corresponding patients, p. 835–846. *In* M. Essex, G. Todaro, and H. zur Hausen (ed.), Viruses in naturally occurring cancers. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
17. **Larsson, E., N. Kato, and M. Cohen.** 1989. Human endogenous proviruses. Curr. Top. Microbiol. Immunol. **148:**115–132.
18. **Leib-Mösch, C., M. Haltmeier, T. Werner, E.-M. Geigl, R. Brack-Werner, U. Francke, V. Erfle, and R. Hehlmann.** 1993. Genomic distribution and transcription of solitary HERV-K LTRs. Genomics **18:**261–269.
19. **Löwer, J., E. M. Wondrak, and R. Kurth.** 1987. Genome analysis and reverse transcriptase activity of human teratocarcinoma-derived retroviruses. J. Gen. Virol. **68:**2807–2815.
20. **Löwer, R., K. Boller, B. Hasenmaier, C. Korbmacher, N. Mueller-Lantzsch, J. Löwer, and R. Kurth.** 1993. Identification of human endogenous retroviruses with complex mRNA expression and particle formation. Proc. Natl. Acad. Sci. USA **90:**4480–4484.
21. **Löwer, R., J. Löwer, H. Frank, R. Harzmann, and R. Kurth.** 1984. Human teratocarcinomas cultured in vitro produce unique retrovirus-like viruses. J. Gen. Virol. **65:**887–898.
22. **Löwer, R., J. Löwer, C. Tondera-Koch, and R. Kurth.** 1993. A general method for the identification of transcribed retrovirus sequences (R-U5 PCR) reveals the expression of the human endogenous retrovirus loci HERV-H and HERV-K in teratocarcinoma cells. Virology **192:**501–511.
23. **Malim, M. H., D. F. McCarn, L. S. Tiley, and B. R. Cullen.** 1991. Mutational definition of the human immunodeficiency virus type 1 Rev activation domain. J. Virol. **65:**4248–4254.
24. **Mancuso, V. A., T. J. Hope, L. Zhu, D. Derse, T. Philips, and T. G. Parslow.** 1994. Posttranscriptional effector domains in the Rev proteins of feline immunodeficiency virus and equine infectious anemia virus. J. Virol. **68:**1998–2001.
25. **May, F. E. B., and B. R. Westley.** 1986. Structure of a human retroviral sequence related to mouse mammary tumor virus. J. Virol. **60:**743–749.
26. **McKeown, M.** 1992. Alternative mRNA splicing. Annu. Rev. Cell Biol. **8:**133–155.
27. **Medstrand, P., and J. Blomberg.** 1993. Characterization of novel reverse transcriptase encoding human endogenous retroviral sequences similar to type A and type B retroviruses: differential transcription in normal human tissues. J. Virol. **67:**6778–6787.
28. **Mueller-Lantzsch, N., M. Sauter, A. Weiskircher, K. Kramer, B. Best, M. Buck, and F. Grässer.** 1993. The human endogenous retroviral element K10 (HER-K10) encodes for a full length gag homologous 73 KD protein and a functional protease. AIDS Res. Hum. Retroviruses **9:**343–350.
29. **Murre, C., P. McCaw, and D. Baltimore.** 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell **56:**777–783.
30. **Nasioulas, G., A. Zolotukhin, C. Tabernero, L. Solomin, C. P. Cunningham, G. N. Pavlakis, and B. K. Felber.** 1994. Elements distinct from human

immunodeficiency virus type 1 splice sites are responsible for the Rev dependence of *env* mRNA. J. Virol. **68:**2986–2993.
31. **Oberste, M. S., J. D. Greenwood, and M. A. Gonda.** 1991. Analysis of the transcription pattern and mapping of the putative *rev* and *env* splice junctions of bovine immunodeficiency-like virus. J. Virol. **65:**3932–3937.
32. **Oberste, M. S., J. C. Williamson, J. D. Greenwood, K. Nagashima, T. D. Copeland, and M. A. Gonda.** 1993. Characterization of bovine immunodeficiency virus *rev* cDNAs and identification and subcellular localization of the Rev protein. J. Virol. **67:**6395–6405.
33. **Ono, M.** 1986. Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. J. Virol. **58:**937–944.
34. **Ono, M., T. Yasunaga, T. Miyata, and H. Ushikubo.** 1986. Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. J. Virol. **60:**589–598.
35. **Pancino, G., H. Ellerbrok, M. Sitbon, and P. Sonigo.** 1994. Conserved framework of envelope glycoproteins among lentiviruses. Curr. Top. Microbiol. Immunol. **188:**77–105.
36. **Reed, R., and T. Maniatis.** 1986. A role for exon sequences and splice-site proximity in splice-site selection. Cell **46:**681–690.
37. **Rosen, C. A., and G. N. Pavlakis.** 1990. Tat and rev: positive regulators of HIV gene expression. AIDS **4:**499–509.
38. **Saltarelli, M., G. Querat, D. A. M. Konings, R. Vigne, and J. E. Clements.** 1990. Nucleotide sequence and transcriptional analysis of molecular clones of CAEV which generate infectious virus. Virology **179:**347–364.
39. **Schwartz, S., M. Campbell, G. Nasioulas, J. Harrison, B. K. Felber, and G. N. Pavlakis.** 1992. Mutational inactivation of an inhibitory sequence in human immunodeficiency virus type 1 results in Rev-independent *gag* expression. J. Virol. **66:**7176–7182.
40. **Sonigo, P., C. Barker, E. Hunter, and S. Wain-Hobson.** 1986. Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. Cell **45:**375–385.
41. **Staffa, A., and A. Cochrane.** 1994. The *tat*/*rev* intron of human immunodeficiency virus type 1 is inefficiently spliced because of suboptimal signals in the 3′ splice site. J. Virol. **68:**3071–3079.
42. **Stoye, J., and J. Coffin.** 1985. Supplement endogenous retroviruses, p. 375–404. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
43. **Tönjes, R. R., et al.** Unpublished results.
44. **von Heijne, G.** 1984. How signal sequences maintain cleavage specificity. J. Mol. Biol. **173:**243–251.
45. **von Heijne, G.** 1985. Signal sequences. The limits of variation. J. Mol. Biol. **184:**99–105.
46. **von Heijne, G.** 1986. A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. **14:**4683–4690.
47. **Wahl, G. M., S. L. Berger, and A. R. Kimmel.** 1987. Molecular hybridization of immobilized nucleic acids: theoretical concepts and practical considerations. Methods Enzymol. **152:**399–407.
48. **Wilkinson, D. A., D. L. Mager, and J. A. C. Leong.** 1994. Endogenous human retroviruses, p. 465–535. *In* J. A. Levy (ed.), The Retroviridae, vol. 3. Plenum Press, New York.
49. **York, D. F., R. Vigne, D. W. Verwoerd, and G. Querat.** 1992. Nucleotide sequence of the Jaagsiekte retrovirus, an exogenous and endogenous type D and B retrovirus of sheep and goats. J. Virol. **66:**4930–4939.