VIROLOGY 209, 550–560 (1995)

Identification of S71-Related Human Endogenous Retroviral Sequences with Full-Length *pol* Genes

MANUELA HALTMEIER,* WOLFGANG SEIFARTH,* JÜRGEN BLUSCH,† VOLKER ERFLE,‡ RÜDIGER HEHLMANN,* and CHRISTINE LEIB-MÖSCH*,‡,[1]

*III Medizinische Klinik, Klinikum Mannheim der Universität Heidelberg, D-68305 Mannheim, Federal Republic of Germany; †Institut für Säugetiergenetik and ‡Institut für Molekulare Virologie, GSF-Forschungszentrum für Umwelt und Gesundheit, D-85758 Oberschleissheim, Federal Republic of Germany

Received August 5, 1994; accepted March 24, 1995

The human genome contains sequences related to the simian sarcoma-associated virus SSAV. One of these endogenous retroviral elements, S71, is truncated in the *pol* gene and carries an insertion of a solitary HERV-K LTR. Using a PCR approach we have now identified further S71-related retroviral elements that lack the HERV-K LTR insertion and contain a full-length retroviral reverse transcriptase. Two of these sequences, pCRTK1 and pCRTK6, were cloned and further characterized. Clones pCRTK1 and pCRTK6 showed between 85 and 90% nucleotide homology to each other and to S71 within the "tether" region of the *pol* gene, indicating that pCRTK1 and pCRTK6 clearly belong to the S71 subgroup of C-type-related human endogenous retroviral elements. Some point mutations inactivating the reverse transcriptase are located at the same positions in pCRTK1 and pCRTK6. Therefore, we assume that these S71-related elements were dispersed in the human genome by reintegration as defective proviruses, probably using enzymes for retrotransposition provided *in trans* by other retrotransposons or by cellular genes. Examination of the presence of S71-related elements in apes and Old World monkeys revealed that the deletion of reverse transcriptase sequences in S71 has occurred in the lineage of primates prior to the insertion of the HERV-K LTR.

INTRODUCTION

The flow of information from DNA to RNA to protein was an early dogma of biology. This dogma was challenged 1970 by the discovery of an enzyme able to convert RNA into DNA: reverse transcriptase (Temin and Mizutani, 1970; Baltimore, 1970). Since this enzyme was found to be associated with retroviruses and to be essential for the replication of these RNA viruses, reverse transcription was generally believed to occur exclusively as part of the retrovirus life cycle. This view has since then been repudiated by the discovery that not only the eukaryotic genome but also bacteria contain sequences related to retroviral reverse transcriptase and endonuclease/integrase genes (Lampson *et al.*, 1989; for review see Varmus, 1989; Inouye and Inouye, 1991). In vertebrates, about 10% of the genome is thought to have emerged by reintegration of products of reverse transcription (Weiner *et al.*, 1986). The information generated in this manner includes sequences which seem to have employed cellular mechanisms for passive retroposition as well as retroelements containing reverse transcriptase- and/or integrase-related sequences possibly initiating their own retrotransposition (Temin, 1989; Hull and Will, 1989). Thus, reverse transcription is a basic mechanism that has contributed substantially to the architecture of the eukaryotic genome (Finnegan, 1989). Among the retroelements which in themselves may have the capacity to transpose are nonviral elements such as LINES (long interspersed nuclear sequences) (Hutchison *et al.*,1989) as well as retrovirus-like elements with structural analogies to infectious retroviruses.

In humans about 1% of the genome consists of such endogenous retrovirus-related sequences. Most of them resemble mammalian C-type or B-type retroviruses (for review see: Callahan, 1988; Brack-Werner *et al.*, 1989b; Larssen *et al.*, 1989; Leib-Mösch *et al.*, 1990; Wilkinson *et al.*, 1994). Generally, these elements do not contain full-length retroviral genomes, but are partially deleted, or their genes are interrupted by point mutations or frameshifts. To date, no replication competent human endogenous retrovirus (HERV) was detected. Nevertheless, a great number of HERVs is actively transcribed (reviewed in Leib-Mösch *et al.*, 1992; Wilkinson *et al.*, 1994). Most transcripts are observed in placenta and embryonic tissue, but many HERVs are also transcribed in human neoplastic cells and normal tissues. Furthermore, antigens immunologically related to retroviral structural proteins were discovered in human sera and

The nucleotide sequence data reported in this article have been deposited with the GenBank database under Accession Nos. U12969 and U12970.

[1] To whom correspondence and reprint requests should be addressed at GSF-Forschungszentrum für Umwelt und Gesundheit, Institut für Molekulare Virologie, Ingolstädter Landstr. 1, Postfach 1129, D-85758 Oberschleissheim, Federal Republic of Germany.

0042-6822/95 $6.00

tissues (Larsson *et al.*, 1989). Some HERVs are single-copy elements, but the majority represents sequence families with up to 1000 members, pointing out the enormous reservoir of retroviral genes in the human genome. Recently, a B-type-related HERV-family (HERV-K) was associated with the formation of retrovirus-like particles (Löwer *et al.*, 1993; Boller *et al.*, 1993). Since these particles have not been found to be productively infectious, they might be pseudotypes generated by complementation of several expressed HERV-K elements containing a defective retroviral genome.

We have shown previously that S71 is a truncated retroviral element homologous to simian sarcoma-associated virus (SSAV). S71 consists of a complete *gag* gene, a partly deleted *pol* gene lacking the coding region for reverse transcriptase and an LTR-like structure at the 3′ terminus (Leib-Mösch *et al.*, 1986; Brack-Werner *et al.*, 1989a; Werner *et al.*, 1990). Further studies of the genomic organization of S71 revealed that *gag* and *pol* homologous sequences are separated by a 920-bp insertion of a solitary HERV-K LTR element which exists in several thousand copies in the human genome (Leib-Mösch *et al.*, 1993). The endogenous element S71 is found only once per haploid genome and was localized on chromosome 18q21 (Brack-Werner *et al.*, 1989a). However, we have estimated the human genome to contain at least 25–35 SSAV-related sequences in addition to S71 by low-stringency hybridization with a SSAV *pol* probe (Leib-Mösch *et al.*, 1986). In this report we demonstrate that about 15–20 of these sequences are related to S71 *pol* and *gag* sequences and that at least some of these elements contain *pol* genes with full-length retroviral reverse transcriptase.

MATERIALS AND METHODS

DNA

Genomic DNA used for PCR analysis was extracted from human lymphocytes by standard procedures (Sambrook *et al.*, 1989). PCR controls were performed with genomic DNA of spinach, kindly provided by Dr. M. Salomon (Botanisches Institut, München). Sources of primate genomic DNA were: orangutan, liver tissue; and marmoset, blood (Zoological garden Helabrunn, München); African green monkey, vero kidney cell line (ATCC CCL 81); rhesus monkey, LLC-MK_{2D} kidney cell line (ATCC CCL 7.1). Chimpanzee genomic DNA was kindly provided by Dr. J. Heeney (Primate Center TNO, Rijswijk, The Netherlands).

PCR analysis

PCR analyses were carried out using primers derived from S71-specific regions flanking the HERV-K LTR insertion (Fig. 2). Two 5′ primers (1, 2) originated from *gag* p10 sequences of S71, two 3′ primers (3, 4) were derived from the tether region of S71 *pol*, and one 3′ primer (5) was designed using a highly conserved domain of reverse transcriptase (Shih *et al.*, 1989). Sequences of primers are shown in Table 1. All of them were produced by an Applied Biosystems oligonucleotide synthesizer (0.2 μmol scale) and purified with OPC columns according to manufacturer's instructions (Applied Biosystems, Weiterstadt, Germany). PCR conditions per 100 μl reaction volume were 10 m*M* Tris, pH 8.3, 50 m*M* KCl, 1.5 m*M* $MgCl_2$, 0.001% gelatine, 200 μ*M* each dNTP, 2.5 Units AmpliTaq polymerase, 0.5 μ*M* each primer. PCRs were performed with 1 μg human genomic DNA. PCR cycle parameters were: 2 min 94°C; 2 min 45°C (50°C, 55°C); 3 min 72°C, 1×; 1 min 94°C; 2 min 45°C (50°C, 55°C); 3 min 72°C, 29×; 7 min 72°C, 1×. To check for possible contaminations, each set of reactions also included two negative controls with 1 μg spinach DNA and without DNA, respectively. These controls gave uniformly negative results.

TABLE 1

Primers Used for Amplification of S71-Related *pol* Sequences

Primer	Sequence	Position[a]
1: *gag* specific	5′ ATCACAGAACACTAGCCAGG 3′	1831–1849
2: *gag* conserved	5′ CAAGCAGATAGGACACTGG 3′	1924–1942
3: Tether	5′ GACAAATAGATGACTGGG 3′	3130–3113
4: Tether	5′ GAAGTCTGGATCAGGGAGC 3′	3409–3392
5: Reverse transcriptase[b]	5′ CTTGGGGTAGTACTTTCC 3′	

[a] Positions of S71-specific primers 1–4 are numbered according to Werner *et al.* (1990).

[b] Primer 5 was derived from a highly conserved region within the reverse transcriptase of retroviruses (Shih *et al.*, 1989).

Southern blot analysis

A 1/10 vol of the PCR mixtures was electrophoresed in 1.5–2% agarose gels and transferred to Zeta-Probe membranes (Bio-Rad, München) by the vacu-blot procedure (Vacu-Gene XL, Pharmacia/LKB, Freiburg, Germany). The membranes were hybridized sequentially to probes specific for HERV-K LTR, S71 *pol*/tether and reverse transcriptase of AKV (Fig. 2). Prehybridization was carried out for 2 hr at 65°C in 6X SSC (1X SSC = 0.15 *M* NaCl, 0.015 *M* sodium citrate), 0.05 *M* NaH_2PO_4, pH 6.5, 0.5% SDS, 0.7% powdered milk. Hybridization was performed for at least 16 hr at 65°C in 6X SSC, 0.02 *M* NaH_2PO_4, 0.5% SDS, and 0.7% powdered milk. The final wash was done at 55°C in 0.5X SSC, 0.1% SDS. Between hybridizations the membranes were stripped using boiling 0.1X SSC and 0.5% SDS. Probes were labeled by the nick translation method (Rigby *et al.*, 1977) and 1×10^6 cpm ^{32}P-labeled DNA/ml solution was used for each hybridization.

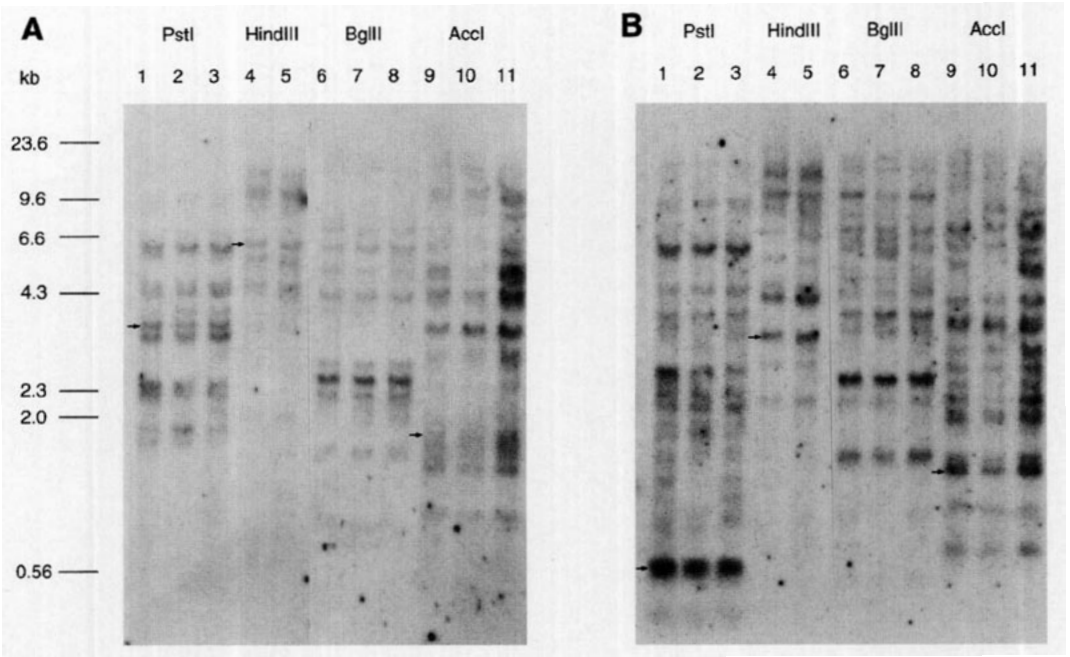

FIG. 1. Southern analysis of S71-related endogenous retroviral elements in human genomic DNA. Genomic DNA from 3 unrelated individuals was digested with *PstI* (lanes 1–3), *Hin*dIII (lanes 4 and 5), *Bgl*II (lanes 5–8), and *Acc*I (lanes 9–11) and sequentially hybridized to a S71 *gag*-specific probe (A) and to a S71 *pol*-specific probe (B) under low stringency conditions. Arrows indicate bands specific for the original S71 element.

Sequence analysis

Amplification products were extracted from the agarose gel by the freeze–squeeze method (Tautz and Renz, 1982) and subcloned into the *Sma*I site of pUC19 vector. For direct sequencing fragments were purified with the Qiaquick gel extraction kit (Quiagen, Hilden, Germany). Sequencing was performed using the dideoxy-chain termination method (Sanger *et al.*, 1977). The method was modified for use with double-stranded templates and AmpliTaq polymerase on the ABI sequencer 373A (Applied Biosystems, Weiterstadt, Germany). Labeling of the reaction products resulted from fluorescence-labeled primers or terminators (Prober *et al.*, 1987). Sequence data were collected and evaluated by the ABI Sequencer 373A and Macintosh IIcx (ABI analysis program for ABI Sequencer 373A). Sequence comparison and alignments were done with the aid of the software package GeneWorks (IntelliGenetics, Inc.) and the IntelliGenetics-suite for the VAX station 3200 (VMS 5.1).

RESULTS

Identification of S71-related *pol* sequences

To address the question whether the human genome contains endogenous retroviral sequences related to S71 we carried out Southern blot analyses (Fig. 1). Human genomic DNA was digested with different restriction endonucleases and the Southern blot was probed sequentially with a fragment comprising the whole S71 *gag* gene and a S71 *pol*-specific probe (Fig. 2), respectively. Hybridizations were carried out under low-stringency conditions. Both S71 *pol* and S71 *gag* probes yielded complex signal patterns. However, the hybridization with the *pol*-specific probe showed bands of higher intensity than the pattern obtained with S71 *gag*. This observation can be explained by the fact that the *pol* regions of retroviruses contain highly conserved domains (McClure *et al.*, 1988; Xiong and Eickbusch, 1990) and that many different endogenous retroviral elements with related *pol* genes are dispersed throughout the human genome. We estimated S71-related endogenous elements to encompass about 15–20 copies in the haploid genome on the basis of number and intensity of the signals compared with the single-copy S71-specific hybridization signals.

We then investigated whether any S71 homologous sequences identified by Southern blotting contain full-length *pol* genes and lack the HERV-K LTR insertion. For this purpose we used the PCR strategy outlined in Fig. 2. Primer combinations for PCR consisted of one of two different 5′ primers derived from the S71 *gag*/p10 region, combined with one of two 3′ primers from the S71 *pol*/tether domain, or with a 3′ primer from a highly conserved part of the C-type retroviral reverse transcriptase. The PCR amplification products of all possible primer combinations were analyzed by agarose gel electrophoresis and sequential hybridizations of the appropriate Southern blots (Fig. 3) with probes specific for S71 HERV-K LTR and S71 *pol*/tether region (Fig. 2A). Furthermore, a 1-kb fragment containing the full-length reverse transcriptase of the murine leukemia virus AKV was used to

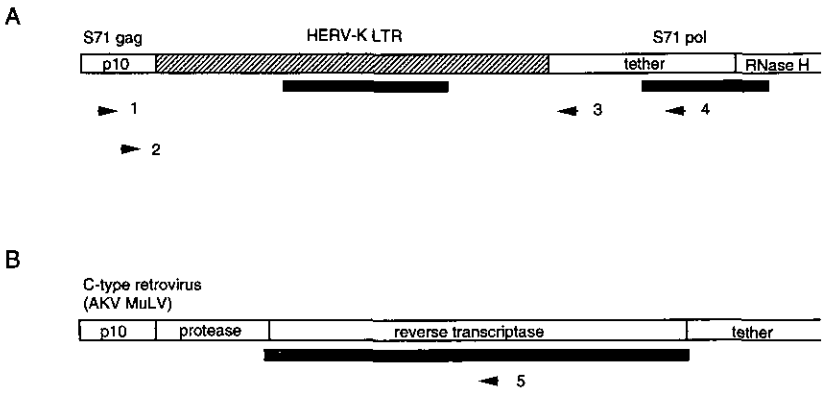

FIG. 2. PCR strategy for identification of S71-related reverse transcriptase coding sequences. The numbered arrows indicate the locations of primers listed in Table 1. Hybridization probes are represented by black bars.

detect reverse transcriptase-specific amplification products (Fig. 2B). The ethidium bromide-stained agarose gel showed a few additional bands beside the S71-specific fragments, above all with primer combination 2/4 (Fig. 3A). Hybridization with HERV-K LTR showed signals corresponding to the molecular weights expected for S71 products (Fig. 3B). In contrast, the S71 *pol*-specific probe yielded in addition to the S71 products bands of 1.9,

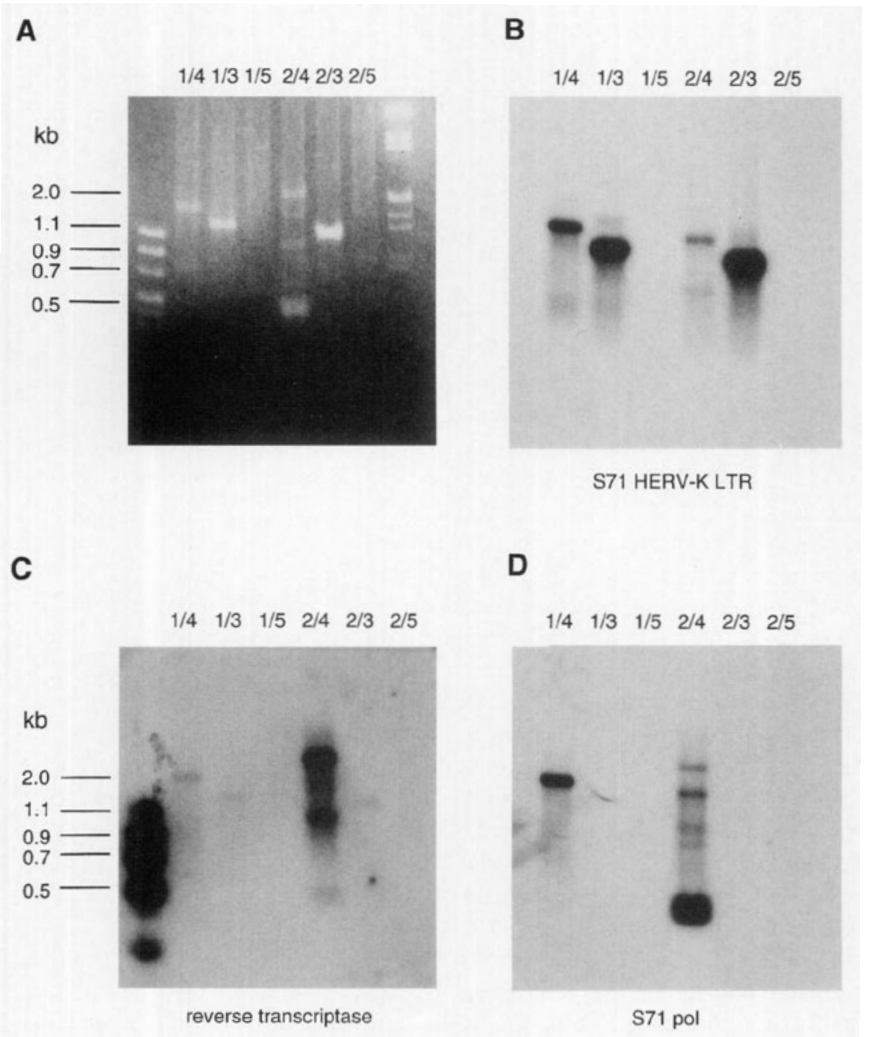

FIG. 3. PCR analysis of S71 *pol*-related sequences. (A) Ethidium bromide-stained agarose gel of PCR products. Amplified S71-related RT sequences were identified by hybridization with (B) S71 HERV-K LTR, (C) AKV reverse transcriptase, and (D) S71 *pol*.

0.7–0.9, and 0.4–0.5 kb (Fig. 3D). Two of these bands obtained with primer combination 2/4 also hybridized with the reverse transcriptase probe (Fig. 3C). The upper fragment corresponded to a length of 1.9 kb which is sufficient to contain the predicted *pol* region missing in S71. This 1.9-kb band was extracted from an agarose gel and subcloned into the pUC19 vector. Two of the subclones (pCRTK1 and pCRTK6) hybridized strongly with the AKV reverse transcriptase probe (data not shown) and were subsequently sequenced.

Sequence analysis of amplified S71-related *pol* sequences

A nucleotide sequence comparison of pCRTK1 with pCRTK6 revealed that these clones show 86% nucleotide identity. In order to assess the relation of pCRTK1/6 to S71, both subclones were compared with the S71 sequence in a DNA dot matrix analysis. Figure 4A shows the comparison of the DNA sequences of S71 and pCRTK6. The abscissa corresponds to the whole insert of pCRTK6, whereas the ordinate comprises the S71 segment between the PCR primer 2 within the p10 sequence and primer 4 within the tether region that were used for the amplification reaction (Fig. 2). This illustration indicates regions with >60 % similarity as diagonals. The first 60 bp (diagonal a) of the pCRTK6 sequence are homologous to S71 sequences preceding the HERV-K LTR insertion which is indicated by the large interruption between a and b. (Fig. 4A). Diagonal b shows a homologous domain corresponding to a 100-bp stretch in S71. The third domain with an obvious relationship between S71 and pCRTK6 covers the tether region (diagonal c). Between diagonals b and c is a large deletion within the S71 sequence that comprises the reverse transcriptase domain of the *pol* gene. This figure shows that the isolated PCR clone lacks the HERV-K LTR insertion (shifting of the diagonals a and b) and instead contains 1500 additional basepairs preceeding the tether region of S71 *pol* (shifting of the diagonals b and c). The comparison of clone pCRTK1 and S71 produced similar results (data not shown).

For identification of the additional sequences found in the PCR clones, nucleotide and amino acid sequences of pCRTK1 and pCRTK6 were compared with the *pol* gene of AKV. Figure 4B shows the resulting dot matrices of DNA and deduced amino acid sequences, respectively. The abscissae correspond to the *pol* gene of AKV; the ordinates represent the PCR clone pCRTK6. The dot matrices show more or less uninterrupted diagonals over the complete length of the compared sequences, illustrating the colinearity and high degree of similarity of both sequences. The relationship of the sequences is more clearly revealed in the dot matrix based on the amino acid data, where the background signals are less and the diagonal is more continuous, indicating a higher

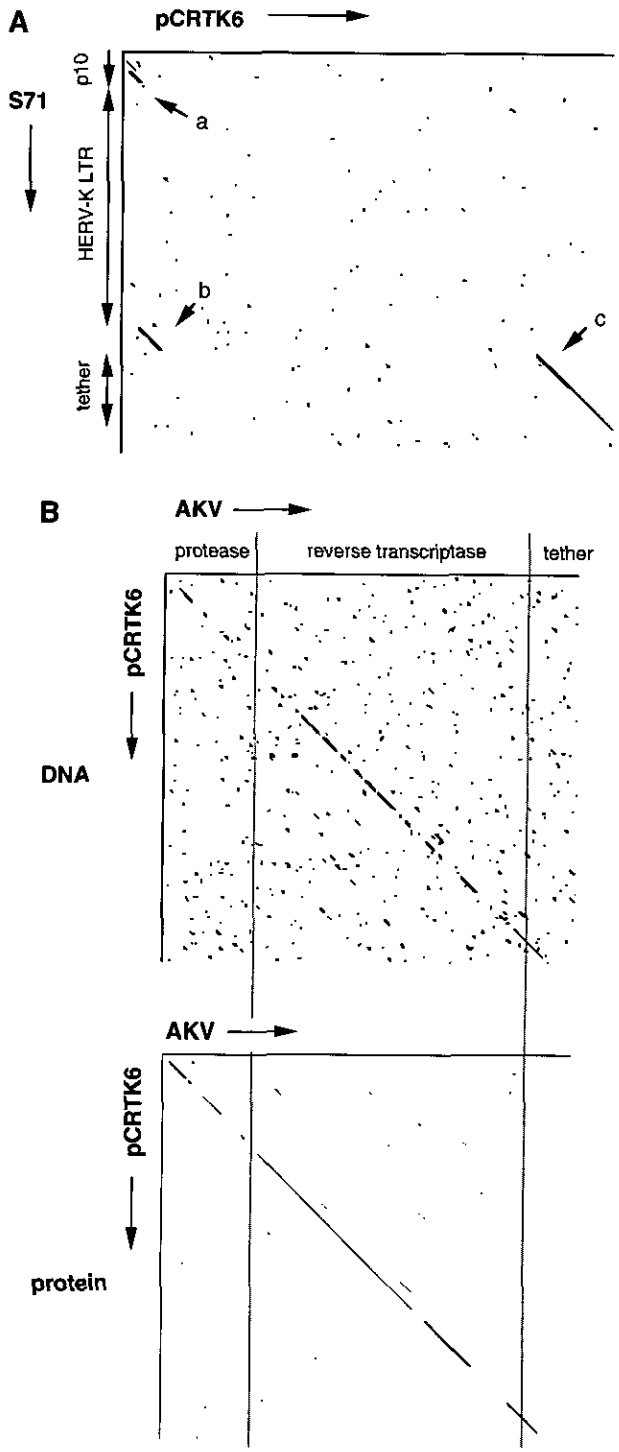

FIG. 4. Dot matrix comparison of (A) pCRTK6 and S71, nucleotide position 1900–3500 (Werner *et al.*, 1990) and (B) pCRTK6 and AKV *pol* sequences, nucleotide position 2250–4080 (Etzerodt *et al.*, 1984). Dots signify a minimal match of 60% in a window size of 20 residues.

degree of conservation on the amino acid level than on the nucleotide level. The most distinct homology is located in the reverse transcriptase segment, yielding up to 70% identity on protein level. The overall amino acid identity amounts to 46%. Similar results were obtained

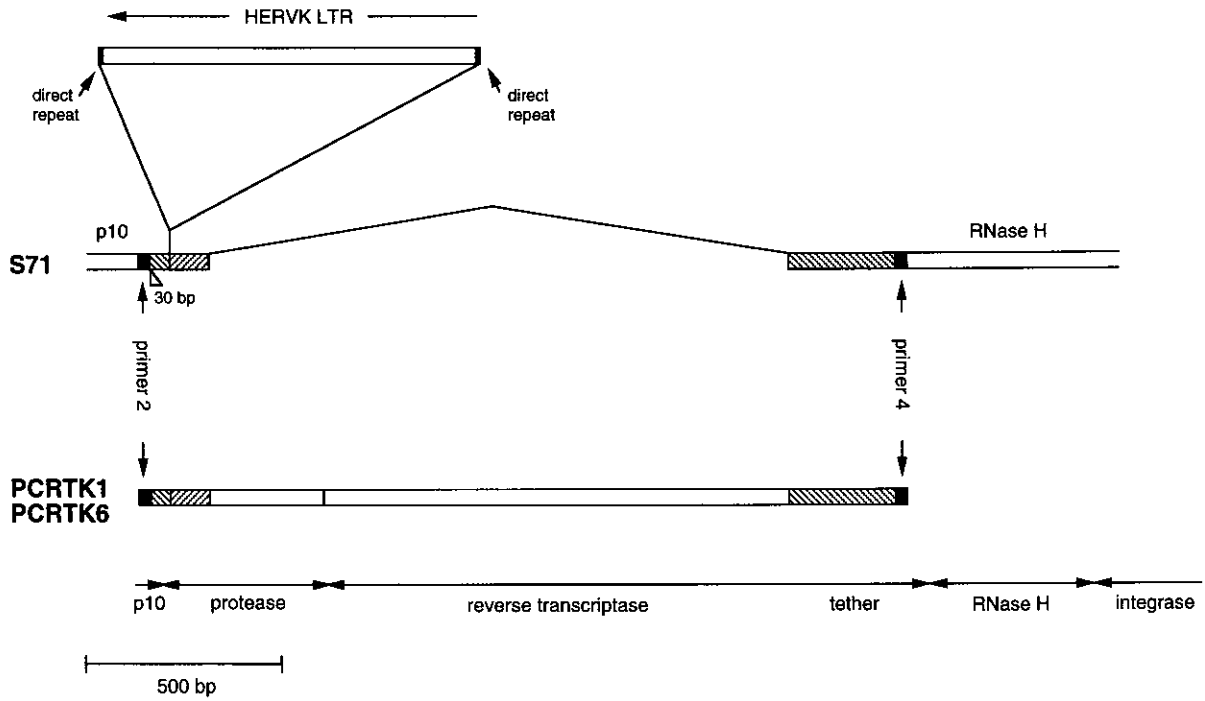

FIG. 5. Structure of S71 and pCRTK1/pCRTK6 *pol* region. The striped boxes indicate the regions homologous in S71 and TK1/6. The position of primers is shown by stippled boxes. S71 carries an additional 30-bp deletion between p10 and the protease coding region compared with pCRTK1 and pCRTK6.

for clone pCRTK1 (data not shown). Hence, pCRTK1 and pCRTK6 contain complete protease and reverse transcriptase domains that are missing in S71. These results are summarized in Fig. 5 showing the structure of S71 and pCRTK1/6 derived from the dot matrices. The first 160 bp of pCRTK1/6 are strongly related to S71 and correspond to the 3′ part of *gag*/p10 and the 5′ domain of the retroviral protease. In S71 the homologous region is interrupted by an HERV-K LTR insertion (Leib-Mösch *et al.*, 1993). The remainder of the protease and the complete reverse transcriptase that is deleted in S71 is present in both PCR clones. The following sequences constituting the tether region are again homologous in S71 and pCRTK1/6. In contrast to S71, both S71-related sequences isolated by PCR contain full-length RT and protease domains. The coding regions of the protease as well as of the reverse transcriptase genes are disrupted by termination codons and frame shifts; therefore, functional proteins are not expected to be produced. Some of the point mutations leading to inactivation of the open reading frames are identical with respect to nucleotide and position in pCRTK1 and in pCRTK6, indicating that these elements may have been integrated into the human genome after inactivation of these genes (Fig. 6).

Phylogenetic analysis of S71-related *pol* genes

We have previously shown that S71 represents a separated phylogenetic subgroup among C-type-related HERVs (Werner *et al.*, 1990). The S71 element seems to be more related to infectious murine and primate proviruses than to other human endogenous retroviral elements. To determine the phylogenetic position of pCRTK1 and pCRTK6, a 300-bp-long nucleotide sequence from the tether region was used for a multiple alignment with the corresponding sequences of other HERVs, such as S71, ERV4-1, and ERV9 as well as the mammalian endogenous and exogenous retroviruses SSAV, GaLV, BaEV, FeLV, and AKV. The resulting phylogenetic tree indicates that pCRTK1 and pCRTK6 clearly belong to the S71 subgroup (Fig.7). ERV4−1, another C-type-retrovirus-related human endogenous element, shows about the same phylogenetic distance to the S71 group as to the mammalian C-type retroviruses AKV, FeLV, BaEV, GaLV, and SSAV. Of all analyzed sequences, the element with the greatest distance to the S71 elements is the human endogenous retrovirus ERV9.

Previously performed Southern analysis showed that S71-related sequences are present in apes and Old World monkeys (Leib-Mösch *et al.*, 1992). In order to investigate the spreading of S71-related elements in primates and to trace back the insertion and deletion events that truncated the S71 element, PCR analyses with genomic DNA from orangutan, chimpanzee, African green monkey, rhesus monkey, and marmoset were performed. We used the same primer combinations (1/4, 1/3, 2/4, and 2/3) that led to the identification of S71-related reverse transcriptase sequences in human DNA. The amplification products were separated by agarose gel electropho-

A

```
                3156
AKV     CAATTAACCTGGACCAGACTCCCACAGGGTTTCAAAAACAGTCCCACCCTGTTTGATGAG
        Q  L  T  W  T  R  L  P  Q  G  F  K  N  S  P  T  L  F  D  E

pCRTK1  CAATTAACCTGGACTCAATTACCCCAGGGTTTTAAAAACTCCCCTACCCTTTTTGGGGAA
        Q  L  T  W  T  Q  L  P  Q  G  F  K  N  S  P  T  L  F  G  E

pCRTK6  CAATTGACCTGGACCCGGTTACCTCAAGGTTTTAAAAATTCCGCCACCCTTTTTGGGGAA
        Q  L  T  W  T  R  L  P  Q  G  F  K  N  S  A  T  L  F  G  E

                                                              3272
AKV     GGACTACACAGAGACCTAGCAGACTTCCGGATCCAGCACCCAGACTTGATCCTGCTACAG
        G  L  H  R  D  L  A  D  F  R  I  Q  H  P  D  L  I  L  L  Q

pCRTK1  GCCCTCCAACAGGATCTTATACCATTCTGAGCCAGTAACCCTCACTGCACTCTTCTCCAG
        A  L  Q  Q  D  L  I  P  F  *  A  S  N  P  H  C  T  L  L  Q

pCRTK6  GCCCTCCAACAAGATCTTCTACCATTCTGAGCCAGTCCCCTTAACTGTAACTCTTCTTCA
        A  L  Q  Q  D  L  L  P  F  *  A  S  P  L  N  C  N  S  S  S
```

B

```
                3702
AKV     AAGCCCTTTGAACTCTTTGTCGACGAGAAGCAGGGCTACGCCAAAGGCGTCCTAACGCAA
        K  P  F  E  L  F  V  D  E  K  Q  G  Y  A  K  G  V  L  T  Q

pCRTK1  AAACCATTTCACCTTTTTGTTCATGAAAGCCAGGGATCACTAAAGGGGTACTCACTTAGA
        K  P  F  H  L  F  V  H  E  S  Q  G  S  L  K  G  Y  S  L  R

pCRTK6  AAACCATTCCACCTTTTTTTCATGAAAGCCAAGGAGTCGCTAAAGGGGTACTCACTCAAA
        K  P  F  H  L  F  F  M  K  A  K  E  S  L  K  G  Y  S  L  K

                                                              3818
AKV     AAACTGGGACCTTGGCGTCGGCCCGTGGCCTACCTGTCCAAAAAGCTAGATCCAGTGGCA
        K  L  G  P  W  R  R  P  V  A  Y  L  S  K  K  L  D  P  V  A

pCRTK1  CTT-TGGGACCATGGCAATGCCCAGTGGCCTACTTGTCTAAGAGACTAGACCCTGAGGCC
        L     G  P  W  Q  C  P  V  A  Y  L  S  K  R  L  D  P  E  A

pCRTK6  CTT-TGCGGCCATGTCGACGCCCAGTGGCCTGTTTATCTAAAAGACTGGATCCTGTTGCC
        L     R  P  C  R  R  P  V  A  C  L  S  K  R  L  D  P  V  A
```

FIG. 6. Comparison of reverse transcriptase sequences of pCRTK1 and pCRTK6 with the corresponding AKV sequences (Etzerodt *et al.*, 1984). The aligned stretches show point mutations creating a termination codon (A) and a frameshift (B) that are conserved in pCRTK1 and pCRTK6.

resis and hybridized with pCRTK1 and HERV-K LTR probes. With the exception of marmoset DNA that did not yield any hybridizing bands (data not shown), fragments hybridizing with pCRTK1 could be amplified from all primate DNAs with at least one or two primer pairs (Figs. 8A–8D). Chimpanzee and African green monkey DNA yielded the expected bands of approximately 1.9 kb (Fig. 8A) and 1.7 kb (Fig. 8C) corresponding to a full-length retroviral reverse transcriptase with primer combinations 1/3 and 2/3, respectively. In case of orangutan DNA, only short fragments of 0.6 and 0.5 kb could be detected with primer pairs 1/4 and 2/4 (Fig. 8B). These fragments correspond in length to a truncated S71 *pol* gene with the deletion of reverse transcriptase, but without the insertion of HERV-K LTR. In rhesus monkey DNA, both types, the truncated and the nontruncated *pol* sequences, represented by bands of 1.9 and 0.6 kb, respectively, were amplified with primer combination 1/4 (Fig. 8D). Fragments hybridizing with HERV-K LTR could only be identified for chimpanzee and orangutan DNA (Figs. 8E and 8F), indicating that the deletion of reverse transcriptase sequences has occurred in the lineage of primates prior to the integration of the HERV-K LTR.

To confirm the results obtaind by Southern analysis, short fragments hybridizing with S71 *pol* were isolated from agarose gels. Sequence analysis verified that the deletion site of reverse transcriptase sequences is identical in short *pol* fragments amplified from human DNA (S71DEL-HU1 and S71DEL-HU2) and orangutan DNA (S71DEL-OU), compared with the original S71 element (Fig. 9). The human *pol* fragments show an overall nucleotide identity of about 80% with S71. The orangutan-derived sequence is less related to S71 showing a sequence homology of 74% within the region located downstream of the deletion site. Furthermore, S71DEL-OU contains an additional small deletion of 25 bp that is located 16 bp upstream from the large deletion. There are only three nucleotide changes between the two human sequences that may reflect either individual variations or the presence of several closely related elements in the human genome carrying the same *pol* deletion.

DISCUSSION

The majority of known HERVs represent families of related sequences with up to 1000 members in the haploid genome. In this study we demonstrate that S71, a truncated HERV related to the SSAV, represents one copy of a family consisting of about 15–20 members. In S71, the 5′ domain of the *pol* gene, coding for retroviral protease and reverse transcriptase, is deleted and the *gag–pol* region is interrupted by an HERV-K LTR insertion. Using the PCR strategy we succeeded in isolating two different genome fragments of the S71 family (pCRTK1 and pCRTK6) that contain the missing *pol* domains and also lack the HERV-K LTR insertion. The two *pol* isolates are not identical but show a nucleotide sequence identity of 86%. None of the investigated *pol* regions (protease, reverse transcriptase) show long open reading frames but contain stop codons and frameshift mutations. Thus,

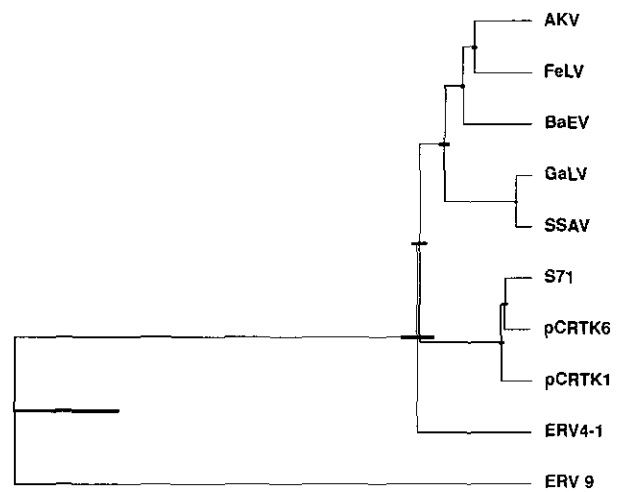

FIG. 7. Phylogenetic relationship of human endogenous retroviral elements and mammalian retroviruses based on sequence homology of *pol* tether region.

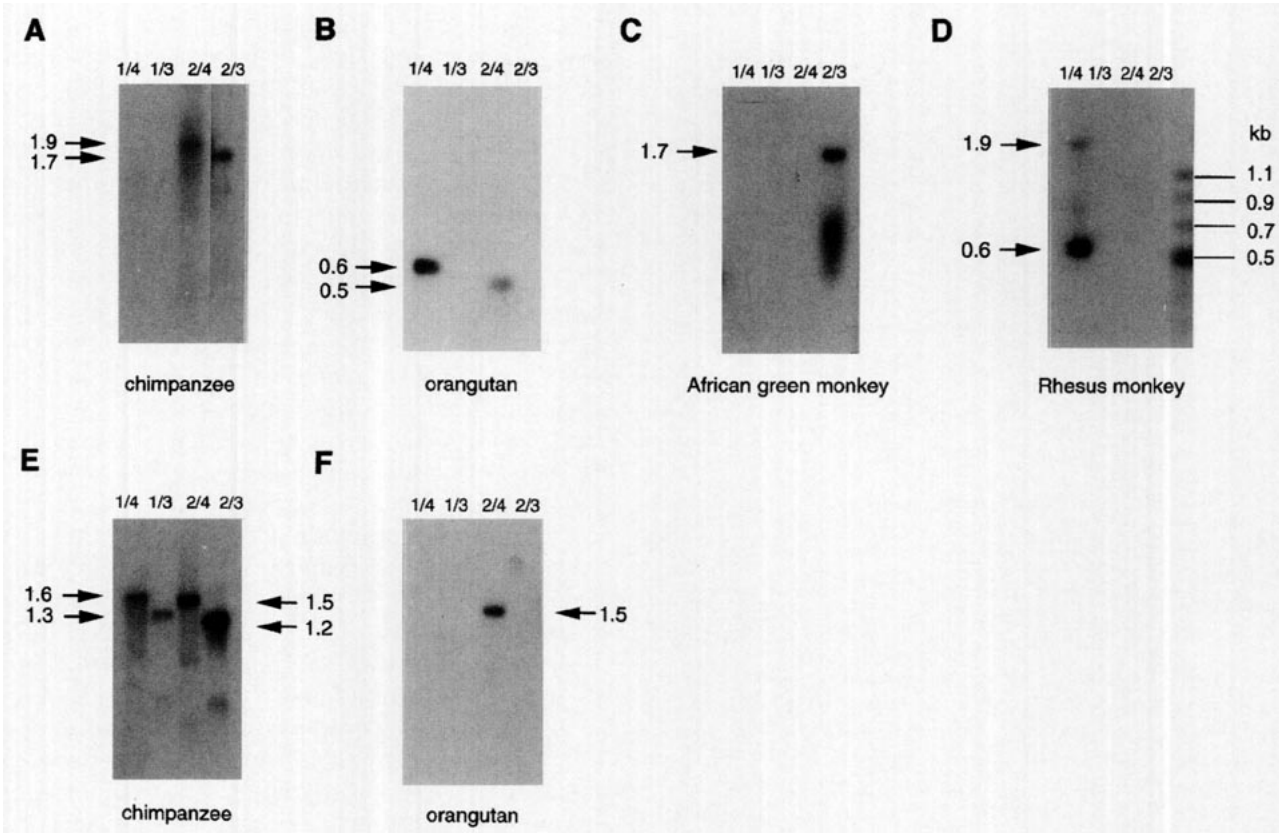

FIG. 8. PCR analysis of S71-related *pol* sequences in primates. (A) Chimpanzee, (B) orangutan, (C) African green monkey, and (D) rhesus monkey DNA hybridized with pCRTK1; (E) chimpanzee and (F) orangutan DNA hybridized with S71 HERV-K LTR.

functional proteins are not expected to be produced. The nucleotide as well as the deduced amino acid sequences of pCRTK1 and pCRTK6 were compared with the corresponding region of AKV, a reproduction competent endogenous retrovirus of the mouse. The resulting dot matrices show distinct but interrupted diagonals in all areas compared. The homology of the isolated *pol* genes to AKV is more clearly seen in the protein dot matrix which exhibits shorter interruptions and less background signals. This observation points to a higher conservation on protein level than on nucleotide level. Such differences in the relationship of nucleotide and protein sequences signify a former functional activity of the gene product. Selection pressure toward function leads to a slower mutation rate on protein level than on nucleotide level because the wobble base of the codon can mutate without changing the sequence of the gene product.

Comparison of the structure of pCRTK1 and pCRTK6 with the S71 element (Fig. 5) clarifies some details how the S71 mutations arose. Figure 5 shows that the HERV-K LTR insertion is located between S71 *gag* p10 and the 5′ part of the protease gene, whereas the *pol* deletion of the S71 element covers the 3′ domain of the retroviral protease and the complete reverse transcriptase. Thus, the HERV-K LTR insertion and the *pol* deletion are separated by about 100 bp of S71 protease sequences. Obviously, both mutations were independent events and the *pol* sequences were not lost as a consequence of the HERV-K LTR insertion. This suggestion is supported by Southern blot analyses of PCR products (Fig. 3) which show fragments of about 500 bp that hybridize strongly with the S71 *pol* probe but give no signals with the HERV-K LTR probe. These observations confirm the assumption that some S71-related elements in the human genome carry a partial *pol* deletion but are not interrupted by the HERV-K LTR insertion.

There is also evidence about the temporal succession of these events. S71-related deleted *pol* genes without HERV-K LTR insertions are found not only in humans but also in orangutan, African green monkey, and rhesus monkey. In contrast, S71 related sequences carrying a HERV-K LTR insertion could only be detected in orangutan DNA. Therefore, the deletion of S71 *pol* regions must have occurred before the divergence of cercopithecoidea and hominoidea, between 45 and 30 million years ago, whereas the HERV-K LTR integration probably happened later on during the evolution of the hominoidea lineage between 30 and 17 million years ago.

Like the S71 familiy, most examined HERV sequences arose early in primate evolution (Harada *et al.*, 1987; Fraser *et al.*, 1988; Johansen *et al.*, 1989; Mariani-Constatini *et al.*, 1989, Perl *et al.*, 1989; Kannan *et al.*, 1991;

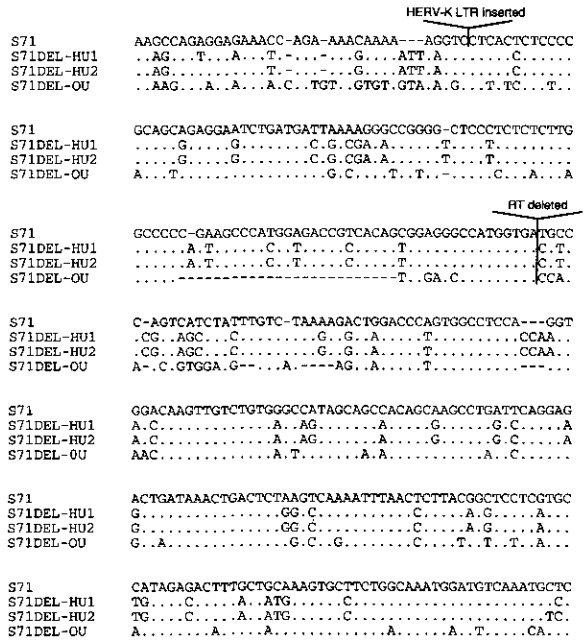

FIG. 9. Nucleotide sequence comparison of amplified short *pol* fragments with the corresponding regions of S71 *pol*. S71DEL-HU1 and S71DEL-HU2 were amplified from human genomic DNA of two unrelated individuals by semi-nested PCR using primers 2 and 4 (Table 1, Fig. 2) in the first PCR reaction. The second PCR reaction was carried out with primer 2 and a primer derived from a sequence highly conserved in S71, pCRTK1, and pCRTK6 tether region (5′ GGAGAGAAAGTTAAACGAGGCCAATC 3′, position 3365–3339 in S71). S71DEL-OU was isolated from orangutan DNA by PCR using primers 2 and 4.

La Mantia *et al.*, 1991; Shih *et al.*, 1991). The type I and type II subfamilies of RTVL-H elements, for instance, have also most significantly expanded after the split of Old World and New World monkeys but before the ape/Old World monkey divergence. In contrast, the type Ia RTVL-H elements form an ape-specific subfamily like the HERV-K LTR containing S71 elements. (Goodchild *et al.*,1993). Taken together the data available on the evolutionary history of HERVs, the most significant expansion of HERVs, and the majority of retrotransposition events seem to trace back to early primate evolution.

Retrotransposition leading to insertion mutagenesis has often been observed *in situ* in nonprimate mammals (Economou-Pachnis *et al.*, 1985; Katzir *et al.*, 1985; Leslie *et al.*, 1991; Steinmeyer *et al.*, 1991; Adachi *et al.*, 1993; Mitreiter *et al.*, 1994; for review see Favor and Morawetz, 1992). Some recent cases of insertion mutagenesis caused by retrotransposable elements have also been detected in the human genome (Kazazian *et al.*, 1988; Morse *et al.*, 1988; Miki *et al.*, 1992; Narita *et al.*, 1993; Holmes *et al.*, 1994). Two active LINE-1 elements, locus LRE1 (Dombroski *et al.*, 1991; Mathias *et al.*, 1991) and locus LRE2 (Holmes *et al.*, 1994), were identified recently. Locus LRE1 is the full-length direct precusor of a truncated LINE-1 element identified as *de novo* insertion in the factor VIII gene of an hemophilia A patient (Kazazian *et al.*, 1988). Another *de novo* insertion of a rearranged LINE-1 element originating from locus LRE2 was localized in the dystrophin gene leading to muscular dystrophy of the JH-1001 type (Holmes *et al.*, 1994). Both LRE1 and LRE2 contain two open reading frames encoding an intact reverse transcriptase and a p40 protein with unknown function.

The fact that the first two human LINE-1 elements known to have transposed recently code for functional reverse transcriptases supports the hypothesis that use of protein products *in cis* is preferred to *trans* retrotransposition of defective elements. The results of our sequence analyses of pCRTK1 and pCRTK6, however, suggests that defective retroviral elements may be retrotransposed with the help of proteins provided *in trans*, since point mutations inactivating the reverse transcriptase are located at the same position in both PCR clones. Furthermore, expansion of RTVL-H involved truncated elements with shared deletions within the *pol* genes (Goodchild *et al.*, 1993; Wilkinson *et al.*, 1994). There are several ways for duplicating defective proviruses: One possibility is the amplification on DNA level as has been observed for the human endogenous retrovirus 4–1 (Steele *et al.*, 1986). In this case, provirus flanking sequences have been found to be identical. Another possibility is the reintegration of the defective sequences making use of enzymes for retrotransposition delivered *in trans* by other retrotransposons or by cellular genes. Abundant expression of retroviral reverse transcriptase-related transcripts has been detected in many human tissues (Medstrand and Blomberg, 1993). Tchenio and Heidmann (1991) demonstrated that a defective retrovirus deleted for *gag*, *pol*, and *env* open reading frames can disperse in the human genome by intracellular transposition if *gag* and *pol* gene products are provided *in trans*. Synthesis of retroviral cDNA is even possible by two defective reverse transcriptase proteins that complement each other. Telesnitsky and Goff (1993) showed that retroviral DNA synthesis can be completed when DNA polymerase and RNaseH activities are provided by separate reverse transcriptase molecules, although much less efficiently. The integration of proviruses can also occur with reverse transcriptase and integrase delivered by different polyprotein precursors. Taken together, these data suggest that intact *pol* genes are not prerequisite for dispersion and consequently also for insertion mutagenesis of human endogenous retroviruses and retroviral elements.

ACKNOWLEDGMENTS

We thank W. Gimbel for introduction to and help with automatic sequencing and T. Werner for the synthesis of oligonucleotides. We also thank M. Jaenicke for expert technical assistance. We are grateful to R. Brack-Werner for critically reading the manuscript. This work was supported by the Commission of the European Union (Contract GENE-CT93–0019). M. Haltmeier was a recipient of a studentship from the BMBW (Hochschulsonderprogramm II).

REFERENCES

Adachi, M., Watanabe-Fukunaga, R., and Nagata, S. (1993). Aberrant transcription caused by the insertion of an early transposable element in an intron of the Fas antigen gene of *lpr* mice. *Proc. Natl. Acad. Sci. USA* **90,** 1756–1760.

Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumor viruses. *Nature* **226,** 1209–1211.

Boller, K., König, H., Sauter, M., Mueller-Lantzsch, N., Löwer, R., Löwer, J., and Kurth, R. (1993). Evidence that HERV-K is the endogenous retrovirus sequence that codes for the human teratocarcinoma-derived retrovirus HTDV. *Virology* **196,** 349–353.

Brack-Werner, R., Barton, D. E., Werner, T., Foellmer, B. E., Leib-Mösch, C., Francke, U., Erfle, V., and Hehlmann, R. (1989a). Human SSAV-related endogenous retroviral element: LTR-like sequence and chromosomal localization to 18q21. *Genomics* **4,** 68–75.

Brack-Werner, R., Leib-Mösch, C., Werner, T., Erfle, V., and Hehlmann, R. (1989b). Human endogenous retrovirus-like sequences. *Haematol. Blood Transfus.* **32,** 464–477.

Callahan, R. (1988). Two families of human endogenous retroviral genomes. *In* "Banbury Report 30: Eukaryotic Transposable Elements as Mutagenic Agents" (M. Lambert, J. McDonald, and I. Weinstein, Eds.), pp. 91–100, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Dombroski, B. A., Mathias, S. L., Nanthakumar, E., Scott, A. F., and Kazazian, H. H., Jr. (1991). Isolation of an active human transposable element. *Science* **254,** 1805–1808.

Economou-Pachnis, A., Lohse, M. A., Furano, A. V., and Tsichlis, P. N. (1985). Insertion of long interspersed repeated elements at the *Igh* (immunoglobulin heavy chain) and *Mlvi-2* (Moloney leukemia virus integration 2) loci of rats. *Proc. Natl. Acad. Sci. USA* **82,** 2857–2861.

Etzerodt, M., Mikkelsen, T., Pedersen, F. S., Kieldgaard, N. O., and Jørgensen, P. (1984). The nucleotide sequence of the AKV murine leukemia virus genome. *Virology* **134,** 196–207.

Favor, J., and Morawetz, C. (1992). Insertional mutations in mammals and mammalian cells. *Mutat . Res.* **284,** 53–74.

Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. *Trends Genet.* **5,** 103–107.

Fraser, C., Humphries, K., and Mager, D. L. (1988). Chromosomal distribution of the RTVL-H family of human endogenous retrovirus-like sequences. *Genomics* **2,** 280–287.

Goodchild, N. L., Wilkinson, D. A., and Mager, D. L. (1993). Recent evolutionary expansion of a subfamily of RTVL-H human endogenous retrovirus-like elements. *Virology* **196,** 778–788.

Harada, F., Tsukada, N., and Kato, N. (1987). Isolation of three kinds of human endogenous retrovirus-like sequences using $tRNA^{Pro}$ as a probe. *Nucleic Acids Res.* **15,** 9153–9162.

Holmes, S., Dombroski, B. A., Krebs, C. M., Boehm, C. D., and Kazazian, H. H. (1994). A new retrotransposable human L1 element from the *LRE2* locus on chromosome 1q produces a chimaeric insertion. *Nature Genet.* **7,** 143–148.

Hull, R., and Will, H. (1989). Molecular biology of viral and nonviral retroelements. *Trends Genet.* **5,** 357–359.

Hutchison, C. A., III, Hardies, S. C., Loeb, D. D., Shehee, W. R., and Edgell, M. H. (1989). LINEs and related retroposons: long interspersed repeated sequences in the eucaryotic genome. *In* "Mobile DNA" (D. E. Berg and M. M. Howe, Eds.), pp. 593–617, Am. Soc. Microbiol., Washington, DC.

Inouye, M., and Inouye, I. (1991). Retroelements in bacteria. *Trends Biochem.* **16,** 18–21.

Johansen, T., Holm, T., and Bjørklid, E., (1989). Members of the RTVL-H family of human endogenous retrovirus-like elements are expressed in placenta. *Gene* **79,** 259–267.

Kannan, P., Buettner, R., Pratt, D. R., and Tainsky, M. A. (1991). Identification of a retinoic acid-inducible endogenous retroviral transcript in the human teratocarcinoma-derived cell line PA-1. *J. Virol.* **65,** 6342–6348.

Katzir, N., Rechavi, G., Cohen, J. B., Unger, T., Simoni, F., Segal, S., Cohen, D., and Givol, D. (1985). "Retroposon" insertion into cellular oncogene *c-myc* in canine transmissible venereal tumor. *Proc. Natl. Acad. Sci. USA* **82,** 1054–1058.

Kazazian, H. H., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G., and Antonarakis, S. E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **332,** 164–166.

Lampson, B. C., Sun, J., Hsu, M.-Y., Vallejo-Ramirez, J., Inouye, S., and Inouye, M. (1989). Reverse transcriptase in a clinical strain of *Escherichia coli:* Production of branched RNA-linked msDNA. *Science* **243,** 1033–1038.

La Mantia, G., Maglione, D., Pengue, G., Di Cristofano, A., Simeone, A., Lanfrancone, L., and Lania, L. (1991). Identification and characterization of novel human endogenous retroviral sequences preferentially expressed in undifferentiated embrional carcinoma cells. *Nucleic Acids Res.* **19,** 1513–1520.

Larsson, E., Kato, N., and Cohen, M. (1989). Human endogenous proviruses. *Curr. Top. Microbiol. Immunol.* **148,** 115–132.

Leib-Mösch, C., Bachmann, M., Brack-Werner, R., Werner, T., Erfle, V., and Hehlmann, R. (1992). Expression and biological significance of human endogenous retroviral sequences. *Leukemia* **6,** 72S–75S.

Leib-Mösch, C., Brack, R., Werner, T., Erfle, V., and Hehlmann, R. (1986). Isolation of an SSAV-related endogenous sequence from human DNA. *Virology* **155,** 666–676.

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,** 5636s–5642s.

Leib-Mösch, C., Haltmeier, M., Werner, T., Geigl, E.-M., Brack-Werner, R., Francke, U., Erfle, V., and Hehlmann, R. (1993). Genomic distribution and transcription of solitary HERV-K LTRs. *Genomics* **18,** 261–269.

Leslie, K. B., Lee, F., and Schrader, J. W. (1991). Intracisternal A-type particle-mediated activations of cytokine genes in a murine myelomonocytic leukemia: Generation of functional cytokine mRNAs by retroviral splicing events. *Mol. Cell. Biol.* **11,** 5562–5570.

Löwer, R., Boller, K., Hasenmaier, B., Korbmacher, C., Mueller-Lantzsch, N., Löwer, J., and Kurth, R. (1993). Identification of human endogenous retroviruses with complex mRNA expression and particle formation. *Proc. Natl. Acad. Sci. USA* **90,** 4480–4484.

Mariani-Constantini, R., Horn, T. M., and Callahan, R. (1989). Ancestry of a human endogenous retrovirus family. *J. Virol.* **63,** 4982–4985.

Mathias, S. L., Scott, A. F., Kazazian, H. H., Boeke, J. D., and Gabriel, A. (1991). Reverse transcriptase encoded by a human transposable element. *Science* **254,** 1808–1810.

McClure, M. A., Johnson, M. S., Feng, D.-F., and Doolittle, R. F. (1988). Sequence comparison of retroviral proteins: Relative rates of changes and general phylogeny. *Proc. Natl. Acad. Sci. USA* **85,** 2469–2473.

Medstrand, P., and Blomberg, J. (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.

Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K. W., Vogelstein, B., and Nakamura, Y. (1992). Disruption of the *APC* gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* **52,** 643–645.

Mitreiter, K., Schmidt, J., Luz, A., Atkinson, M. J., Höfler, H., Erfle, V., and Strauss, P. G. (1994). Disruption of the murine p53 gene by insertion of an endogenous retrovirus-like element (ETn) in a cell line from radiation-induced osteosarcoma. *Virology* **200,** 837–841.

Morse, B., Rotherg, P. G., South, V. J., Spandorfer, J. M., and Astrin, S. M. (1988). Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. *Nature* **333,** 87–90.

Narita, N., Nishio, H., Kitoh, Y., Ishikawa, Y., Ishikawa, Y., Minami, R., Nakamura, H., and Matsubo, M. (1993). Insertion of a 5′ truncated L1 element into the 3′ end of exon 44 of the dystrophin gene resulted

in exon skipping of the exon during splicing in a case of Duchenne muscular dystrophy. *J. Clin. Invest.* **91,** 1862–1867.

Perl, A., Rosenblatt, J. D., Chen, I. S. Y., DiVincenzo, J. P., Bever, R., Poiesz, B. J., and Abraham, G. N. (1989). Detection and cloning of new HTLV-related endogenous sequences in man. *Nucleic Acids Res.* **17,** 6841–6854.

Prober, J. M., Trainor, G. L., Dam, R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., Cocuzza, A. J., Jensen, M. A., and Baumeister, K. (1987). A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* **238,** 336–341.

Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P. (1977). Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase, I. *J. Mol. Biol.* **113,** 237–251.

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.

Sanger, F., Nicklein, S., and Coulson, A. (1977). DNA sequencing with chain termination inhibitors. *Proc. Natl. Scad. Sci. USA* **74,** 5463–5467.

Shih, A., Coutavas, E. E., and Rush, M. G. (1991). Evolutionary implications of primate endogenous retroviruses. *Virology* **182,** 495–502.

Shih, A., Misra, R., and Rush, M. G. (1989). Detection of multiple, novel reverse transcriptase coding sequences in human nucleic acids: relation to primate retroviruses. *J. Virol.* **63,** 64–75.

Steele, P. E., Martin, M. A., Rabson, A. B., Bryan, T., and O'Brien, S. J. (1986). Amplification and chromosomal dispersion of human endogenous retroviral sequences. *J. Virol.* **59,** 545–550.

Steinmeyer, K., Klocke, R., Ortland, C., Gronemeier, M., Jockusch, H., Gründer, S., and Jentsch, T. J. (1991). Inactivation of muscle chloride channel by transposon insertion in myotonic mice. *Nature* **354,** 304–308.

Tautz, D., and Rentz, M. (1982). An optimized freeze squeeze method for the recovery of DNA-fragments from agarose gels. *Anal. Biochem.* **132,** 14–19.

Tchenio, T., and Heidmann, T. (1991). Defective retrovirus can disperse in the human genome by intracellular transposition. *J. Virol.* **65,** 2113–2118.

Telesnitsky, A., and Goff, S. P. (1993). Two defective forms of reverse transcriptase can complement to restore retroviral infectivity. *EMBO J.* **12,** 4433–4438.

Temin, H. M. (1989). Reverse transcriptases: Retrons in bacteria. *Nature* **339,** 254–255.

Temin, H. M., and Mizutani, S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226,** 1211–1213.

Varmus, H. E. (1989). Reverse transcription in bacteria. *Cell* **56,** 721–724.

Weiner, A. M., Deininger, P. L., and Efstratiadis, A. (1986). Nonviral retroposons: Genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* **55,** 631–661.

Werner, T., Brack-Werner, R., Leib-Mösch, C., Backhaus, H., Erfle, V., and Hehlmann, R. (1990). S71 is a phylogenetic distinct human endogenous retroviral element with structural and sequence homology to simian sarcoma virus (SSV). *Virology* **155,** 225–238.

Wilkinson, D. A., Mager, D. L., and Leong, J.-A. (1994). Endogenous human retroviruses. *In* "The Retroviridae" (J. Levy, Ed.), Vol. 3, pp. 465–535, Plenum, New York.

Xiong, Y., and Eickbush, T. H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9,** 3353–3362.