VIROLOGY 234, 14–30 (1997)
ARTICLE NO. VY978590

Comparative Analyses of LTRs of the ERV-H Family of Primate-Specific Retrovirus-like Elements Isolated from Marmoset, African Green Monkey, and Man

Sølvi Anderssen, Eva Sjøttem, Gunbjørg Svineng, and Terje Johansen[1]

Department of Biochemistry, Institute of Medical Biology, University of Tromsø, 9037 Tromsø, Norway

Received January 16, 1997; returned to author for revision March 10, 1997; accepted April 28, 1997

We have isolated 8 different long terminal repeat (LTR) sequences of the ERV-H family of endogenous retrovirus-like elements from human chromosome 18, 9 from African green monkey, and 28 from marmoset. Human ERV-H LTRs have been divided into three types designated Type I, Type Ia, and Type II. Comparative analyses of the 45 isolated LTRs and 60 human ERV-H LTRs enabled a further subdivision into 13 subtypes. Type I elements were widely distributed in all three species. Their average evolutionary age (40 MYr), estimated by a consensus sequence approach, suggests that they first expanded in the genomes at the time New- and Old World monkeys diverged. The occurence of some very old Type I sequences indicate that ERV-H elements may have integrated even before prosimians and primates diverged. Type Ia and -II elements were found in both monkey species. Promoter active Type I and Type Ia LTRs were found while Type II LTRs were inactive. Promoter active Type I LTRs generally contained a functional GC/GT box immediately 3′ to the TATA box, providing strong binding of Sp1 family proteins, while the highly promoter active Type Ia element H6 contained synergistically acting Sp1 binding sites located in the U3 enhancer region.

INTRODUCTION

Endogenous retrovirus-related sequences (ERVs) transmitted as stable Mendelian genes have been detected in the genomes of most eukaryotic species. Human ERVs (HERVs) are estimated to make up about 0.6% of the genome and are widely distributed over all chromosomes. The origin of HERV elements is unclear, but the fact that they are more related to rodent viruses than to known human infectious retroviruses indicates that the elements have existed in the genome for a long time (for a recent review see Wilkinson *et al.,* 1994). Most HERV families have integrated into the primate lineage after the divergence of New World and Old World monkeys (Dangel *et al.,* 1994; Haltmeier *et al.,* 1995; Steinhuber *et al.,* 1995). However, some elements are found also in New World monkeys (Kröger and Horak, 1987; Cohen *et al.,* 1988; Perl *et al.,* 1989; Mager and Freeman, 1995; Simpson *et al.,* 1996; Widegren *et al.,* 1996) and may therefore have integrated more than 40 million years (MYr) ago. The distribution of HERV families within primates has primarily been determined by Southern hybridization using conserved regions of the *pol* or *env* genes as probes (Mariani-Costantini *et al.,* 1989; Kannan *et al.,* 1991; Leib-Mösch *et al.,* 1993; Goodchild *et al.,* 1993; Lebedev *et al.,* 1995; Steinhuber *et al.,* 1995; Yeh *et al.,* 1995) or by PCR analyses of such regions (Shih *et al.,* 1991; Medstrand *et al.,* 1992; Medstrand and Blomberg, 1993; Haltmeier *et al.,* 1995; Li *et al.,* 1995; Mager and Freeman, 1995). These studies have concluded that most HERV families are approximately 30–40 MYr old.

The human ERV-H family (formerly RTVL-H), characterized by the presence of a PBS complementary to $tRNA^{His}$, is estimated to comprise around 900 copies of a common internally deleted 5.8-kb form and 50–100 8.7-kb full-length elements in addition to about 1000 solitary long terminal repeats (LTRs) (Mager and Henthorn, 1984; Goodchild *et al.,* 1993; Hirose *et al.,* 1993; Wilkinson *et al.,* 1993). The HERV-H LTRs have been divided into three different classes, designated Type I, Type Ia, and Type II, according to the repeated structures found in their U3 regions (Mager, 1989; Goodchild *et al.,* 1993). Transient transfection assays have revealed that several of the HERV-H LTRs of Type I and Type Ia analyzed display promoter activity (Feuchter and Mager, 1990; Sjøttem *et al.,* 1996). The promoter activity of Type I LTRs was recently shown to be critically dependent on binding of the transcription factor Sp1 to an inverted GC/GT box located just 3′ to the TATA box (Sjøttem *et al.,* 1996). In addition, transcripts of cellular genes initiated in HERV-H LTRs have been reported (Liu and Abraham, 1991; Feuchter *et al.,* 1992; Feuchter-Murthy *et al.,* 1993). The U3 region is found to stimulate transcription from a heterologous promoter (Sjøttem *et al.,* 1996), and both HERV-H transcripts and adjacent cellular genes are found to be 3′ end processed directed by the polyadenylation signal in the R region (Johansen *et al.,* 1989; Mager, 1989; Goodchild *et al.,* 1992). Recombination events between LTRs may lead to rearrangement of the host genome (Mager and Goodchild, 1989). Thus, at least some HERV-H LTRs

[1] To whom correspondence and reprint requests should be addressed at Department of Biochemistry, Institute of Medical Biology, University of Tromsø, 9037 Tromsø, Norway. Fax: 47 776 45350. E-mail: terjej@fagmed.uit.no.

0042-6822/97 $25.00

seem to have evolved into a regulatory role for cellular genes, and their potential to induce genetic rearrangements may have been important during evolution. Furthermore, the considerable number, random chromosomal distribution, and presence of spliced HERV-H elements with intact LTRs strongly suggest that they have been amplified in the genome as viral retrotransposons (Goodchild *et al.,* 1995).

Mager and Freeman (1995) recently reported the presence of *pol* sequences with homology to HERV-H elements in the genomes of the New World monkeys marmoset and owl monkey. Based on Southern blot analysis with a *pol* probe they concluded that such elements are present in less than 50 copies in the New World monkeys compared to about 1000 copies in Old World monkeys and humans. A major amplification of the deleted subfamily from about 50 to 800–1000 copies occured before Old World monkeys and hominoids diverged (Mager and Freeman, 1995). Studying the evolutionary distribution of specific ERV-H LTRs, Goodchild *et al.* (1993) found that the Type I and Type II subfamilies arose early in primate evolution and expanded before the divergence of hominoids from Old World monkeys while the Type Ia subfamily of LTRs was found only in the great apes. However, ERV-H LTRs have not been detected in New World monkeys using Southern blot analyses.

In this work we have used PCR to isolate 8 different ERV-H LTRs from human chromosome 18, 28 from the common marmoset *Callithrix jacchus* as a representative of New World monkeys, and 9 from African green monkey (*Cercopithecus aethiops*) as a representative of Old World monkeys. We present data on the structural and evolutionary distribution of ERV-H LTRs and on their functionality assessed by assaying promoter activities of representative LTRs in human cell lines. The promoter activity of Type I LTRs was strongly correlated with the presence of a functional Sp1 binding site just 3′ to the TATA box, while the promoter active Type Ia LTR H6 contained synergistically acting Sp1 binding sites in its U3 enhancer region.

MATERIALS AND METHODS

PCR-amplification, subcloning, and sequencing

To isolate HERV-H LTRs from a specific human chromosome, PCR was performed on genomic DNA isolated from a hamster–human somatic cell hybrid (line 324, BIOS corp.) carrying only human chromosome 18. Two sets of primers were used. Primerset 1 consisted of 5LTRA (5′-TGTCAGC/GCCTCTGAGC/TCT/CAG/AGC-3′) and 3LTRA (5′-GAGCT/CACCAAACAGGCTTA/TG-3′), which would amplify all three kinds of LTRs since both primers were located within the LTR sequence. Primerset 2 contained 5LTRA and 3PBS (5′-CGAT/CCCGA/GGT/CC/TA/TCGGCACCAA-3′) and would amplify only 5′ LTRs of full-length or truncated HERV-H elements. The Taq enzyme from Gibco BRL was used. To isolate Old- and New World monkey HERV-H LTRs, PCRs were performed on genomic DNA from the African green monkey cell line Vero (ATCC CCL 81) and the marmoset cell line HVS-Silva 40 (ATCC CRL 1773), respectively. Primerset 2 and the Ultma Taq polymerase (Perkin–Elmer–Cetus) were used. PCR was performed in 50-μl reaction mixtures containing 100 ng of genomic DNA as the template and 1 μM of each primer in 10 m*M* Tris–HCl [pH 8.4], 50 m*M* KCl, 1.5 m*M* $MgCl_2$, 0.001% gelatine. An initial denaturation at 94° for 1 min was followed by 25 cycles with denaturation at 94° for 20 sec, annealing at 58° for 20 sec, elongation at 72° for 1 min, and a final extension at 72° for 1 min. PCR products were analyzed on 2% MetaPhor (FMC BioProducts, Rockland, ME) agarose gels. To check for possible contaminating DNA, each set of reactions also included a negative control, which had no added DNA. These controls were uniformly negative. The TA-Cloning System (Version 1.0, Invitrogen) was used to clone the PCR-products from chromosome 18 into the multiple cloning site of pCR1000. The PCR-products from African green monkey and HVS-Silva 40 genomic DNA were cloned into the *Sma*I site of pGEM-3Zf(+) (Promega). Subcloning and DNA sequencing were done according to standard procedures (Sambrook *et al.,* 1989). The human PCR-products were subcloned into M13mp18 and -mp19 for sequencing while the monkey LTRs were sequenced employing universal sequencing primers flanking the polylinker of pGEM-3Zf(+). To isolate Type II LTRs from marmoset PCR was performed on genomic DNA from the marmoset cell line using the PCR primers 5LTRA and 3FPBS (5′-CCCGGGTCTTCGGCACCAA-3′) as described above. The PCR product was cloned into the *Sma*I site of pUC18 (Pharmacia) and sequenced using the M13 forward and reverse primers.

Computer-assisted analyses of DNA sequences

Multiple sequence alignments were produced using the PILEUP program and manually refined with the LINEUP multiple sequence editor of the GCG Software package (Version 8.1, Genetics Computer Group, Madison, WI). Following optimalization of the alignments, LINEUP was used to calculate a consensus sequence for each LTR subtype. The individual sequences and the consensus was then used in the GCG program DISTANCES employing the Kimura two-parameter method to determine sequence divergence (Kimura, 1980). Gaps were excluded from the calculations. BLAST (Altschul *et al.,* 1990) and FASTA (Pearson and Lipman, 1988) programs were used to search the primate and EST divisions of GenBank. Curvature analyses of the LTRs were performed using the BEND program developed by Goodsell and Dickerson (1994).

Southern blot analyses

Ten micrograms of genomic DNA from Vero cells and HVS Silva 40 cells were digested with *Eco*RI or *Xba*I and

electrophoresed in a 0.7% agarose gel at 20 V for 18 hr. The DNA was subsequently transferred to a Nytran membrane (Schleicher and Schuell) by vacuum blotting and the membrane hybridized to a mix of random priming-labeled LTRs. For marmoset DNA the LTRs Silva18, −20, and −48 were used as probes while Vero1, −2, and −3 were used to probe the African green monkey blot. Hybridizations were performed under stringent conditions at 65° as described (Church and Gilbert, 1984). The membranes were washed 4 × 15 min at 65° with 40 m*M* sodium phosphate, pH 7.2, 1% sodium dodecyl sulfate (SDS), and once for 30 min at 65° with 20 m*M* sodium phopshate, pH 7.2, 0.5% SDS.

Plasmid constructions

The promoterless chloramphenicol acetyltransferase (CAT) reporter gene plasmid pBLCAT3 (Luckow and Schütz, 1987) was used to assay promoter activities of cloned HERV-H LTRs. The LTRs 18102 and 18103 were released from pCR1000 by digestion with *Eco*RI and *Spe*I. The *Eco*RI end was made blunt with T4 DNA polymerase and the LTR-fragments inserted into the *Hin*dIII (end-filled) and *Xba*I sites upstream of the promoterless CAT gene of pBLCAT3. LTRs 18106, 18107, and 18321 were cloned as *Eco*RI (end-filled)–*Hin*dIII fragments into *Sal*I (end-filled)–*Hin*dIII-digested pBLCAT3. 18316 was inserted as an *Eco*RI (end-filled)–*Spe*I fragment into *Xho*I (end-filled)–*Xba*I digested pBLCAT3. The LTRs Vero2, Vero3, Vero12, Silva15, and Silva18 were inserted as *Sac*I (end-filled)–*Sal*I fragments into *Bam*HI (end-filled)–*Sal*I-digested pBLCAT3. LTRs Vero13, Vero22, Silva8, Silva16, and Silva20 were inserted as *Eco*RI (end-filled)–*Sal*I fragments into *Hin*dIII (end-filled)–*Sal*I-digested pBLCAT3. The LTRs Silva12 and -43 were inserted as *Eco*RI (end-filled)–*Sal*I fragments into *Bam*HI (end-filled)–*Sal*I-digested pBLCAT3.

The LTR 18321-mII construct has been described previously (Sjøttem *et al.,* 1996), while the 18321-mI and 18321-mIII mutants were generated using the QuickChange SiteDirected Mutagenesis kit (Stratagene) changing the (5′-GGTTCCTGCCTTA-3′) sequence to (5′-GGTCCCCGCCTTA-3′). The mutations were verified by sequencing. The H6 LTR was a generous gift from Dixie Mager and was inserted as a *Stu*I–*Bam*HI fragment (Feuchter and Mager, 1990) into the *Sal*I(blunted)–*Bam*HI site of pBLCAT3. The mutated H6 LTR constructs mI, mII, and mIII were generated using the QuickChange SiteDirected Mutagenesis kit (Stratagene) changing the GC box at position 105 from (5′-TGCCCCGCCTTA-3′) to (5′-TGCGACGTCTTA-3′), introducing an *Aat*II site, the GT box at position 227 from (5′- ACCCCCGCCCCTG-3′) to (5′-ACCCCAGCTGCTG-3′) introducing a *Pvu*II site, and the GT box at position 302 from (5′-GGCCCCACCCCTA-3′) to (5′-GGCCCGACGTCTA-3′), introducing an *Aat*II site, respectively. The mutations were verified by restriction enzyme digestion.

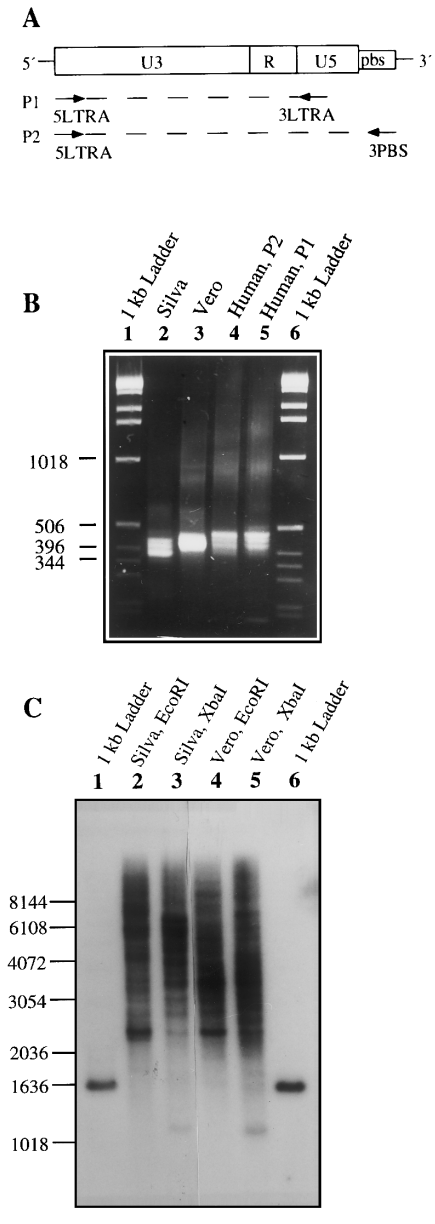


FIG. 1. ERV-H LTRs are present both in New- and Old World monkeys. (A) Schematic of ERV-H LTRs displaying the location of primers used for PCR. Primerset 1, consisting of 5LTRA and 3LTRA, would amplify both solo LTRs as well as 5′- and 3′ LTRs from ERV-H elements with interior sequences. Primerset 2, containing 5LTRA and 3PBS, would amplify only 5′ LTRs containing parts of ERV-H interior sequences including the $tRNA^{His}$ PBS at the 3′ end. (B) A high resolution agarose gel showing the ERV-H LTR PCR products obtained from marmoset (Silva), African green monkey (Vero), and human chromosome 18 (Human, P2 and Human, P1), respectively. Both primerset 1 (denoted P1) and primerset 2 (denoted P2) were used to amplify LTRs from human chromosome 18, while primerset 2 was used to amplify 5′ LTRs from marmoset and African green monkey. The numbers to the left refer to the lengths in basepairs (bp) of specific fragments of the 1-kb ladder used as a size marker. (C) Southern blot analyses confirmed the presence of ERV-H LTRs in marmoset (Silva) and African green monkey (Vero). Genomic DNA (10 μg) from marmoset (lanes 2 and 3) and African green monkey (lanes 4 and 5) were digested with *Eco*RI (lanes 2 and 4) or *Xba*I (lanes 3 and 5) and hybridized to a mix of LTR subtypes from the respective species. The sizes in bp of marker fragments from the 1-kb ladder are indicated to the left.

Cell culture and transient transfection assays

HeLa, JEG-3, and NTera2-D1 cells were cultured as described previously (Sjøttem *et al.,* 1996). COS-7 cells (ATCC CRL 1650) were grown in Dulbecco's modified Eagle's medium containing 10% fetal calf serum.

NTera2-D1-, HeLa-, and JEG3 cells were transfected by the calcium phosphate coprecipitation method with 8 μg reporter plasmid per transfection exactly as described by Sjøttem *et al.* (1996), except that JEG3 cells received a 90-sec glycerol shock. Preparation of cell extracts and CAT assays were carried out as described (Sjøttem *et al.,* 1996). All transfections were carried out in triplicate in at least three independent experiments using different DNA preparations.

Gel mobility shift assays and circular permutation analyses

The preparation of nuclear extracts and gel mobility shift assays (GMSA) were performed as recently described (Sjøttem *et al.,* 1996).

To produce circularly permuted DNA fragments containing the Type I repeat of LTRs 18321 and 18102, a 51-bp *Mse*I (end-filled) fragment (nucleotide positions 80 to 131 of LTR 18321 and positions 118 to 169 of LTR 18102) from the U3 region was first inserted into the *Sal*I (end-filled) site of pBend2 (Kim *et al.,* 1989). Circularly permuted fragments containing the LTR 18321 or the LTR 18102 Type I repeats were then created by digesting the constructed plasmids with *Mlu*I, *Nhe*I, *Xho*I, *Eco*RV, *Stu*I, *Rsa*I, or *Bam*HI, respectively. The ~180-bp fragments were separated on a 5% (29:1) polyacrylamide gel run at 5° and 230 V for 2–3 hr and visualized by ethidium bromide staining.

RESULTS

Isolation of ERV-H LTRs from human chromosome 18 and from the genomes of African green monkey and marmoset

In order to study the sequence diversity of HERV-H LTRs present on a single human chromosome we took advantage of the fact that these elements are absent from rodent genomes (Fraser *et al.,* 1988). Thus, DNA from a hamster–human somatic cell hybrid containing only human chromosome 18 was used in two sets of PCR reactions to obtain HERV-H sequences (Fig. 1A). Both PCR reactions contained the same upstream primer (5LTRA) complementary to the 5′ end of the LTRs. In one reaction a downstream primer (3PBS) complementary to the $tRNA^{His}$ PBS was used to specifically amplify 5′ LTRs, whereas in the other reaction a primer (3LTRA) annealing to the 5′ half of the U5 region was utilized to enable amplification of both 5′, 3′ and solitary LTR sequences present on chromosome 18. To enable direct sequence comparisons between human, Old- and New World monkey ERV-H LTRs primers 5LTRA and 3PBS were used to amplify 5′ LTRs from genomic DNA isolated from African green monkey (*Cercopithecus aethiops*) as a representative of Old World monkeys and from the common marmoset *Callithrix jacchus* as a representative of New World monkeys. High resolution agarose gel electrophoresis of the PCR products showed several bands ranging from 370 to 450 bp in length (Fig. 1B). The major band of marmoset LTRs (lane 2) corresponded to about 370 bp, while the major bands from human chromosome 18 were approximately 100 bp longer (lanes 4 and 5). The predominant size of African green monkey LTRs was intermediary compared to marmoset and human chromosome 18 LTRs, about 400 bp. This may indicate that different subfamilies of the ERV-H LTRs have amplified after the time of divergence of Old World monkeys from New World monkeys and also of humans from Old World monkeys. Alternatively, it may indicate that deletions and/or duplications of internal LTR sequences have occured frequently after the time of divergence. ERV-H elements are reported to be present in less than 50 copies in marmoset DNA based on Southern analysis with a cloned *pol* sequence (Mager and Freeman, 1995), while no LTRs have been detected in the marmoset using human ERV-H LTR probes (Goodchild *et al.,* 1993). We therefore performed Southern blot analyses of genomic DNA from African green monkey and marmoset using mixtures of cloned LTRs as probes for the relevant species to verify the presence of ERV-H LTRs in the marmoset genome (Fig. 1C).

The New World monkey marmoset contains almost exclusively Type I 5′ LTRs

Following cloning of ERV-H LTR PCR products into plasmid vectors, 11 clones from human chromosome 18, 10 African green monkey, and 30 marmoset clones were selected for sequencing based on size determination by high resolution agarose gel electrophoresis as well as

FIG. 2. Alignments of the ERV-H LTR sequences isolated from human chromosome 18, African green monkey, and marmoset. (A) Alignment of two ERV-H LTRs isolated from marmoset (Silva), nine from African green monkey (Vero), and eight from human chromosome 18. The locations of the Type I and Type II repeats, the unique regions I and II, the TATA box with the adjacent GC/GT box, and the polyadenylation signal are indicated above the sequences. The borders between the repeated sequences are denoted by brackets, while the extents of the U3, R, and U5 regions are indicated by arrowheads. The Vero13 and Silva8 LTRs are of Type Ia, while the LTRs shown above these are of Type I and those below of Type II. (B) Alignment of the 26 Type I marmoset ERV-H LTR sequences. The different regions are indicated as in (A). The pairwise sequence divergence of the marmoset Type I LTRs varied from 6 to 23% with an average of 14.5%. The human chromosome 18 LTRs have been assigned the GenBank Accession Nos. U95997–U96004. The accession numbers for the African green monkey and marmoset LTRs are U96005-U96013 and U96046-U96073, respectively.

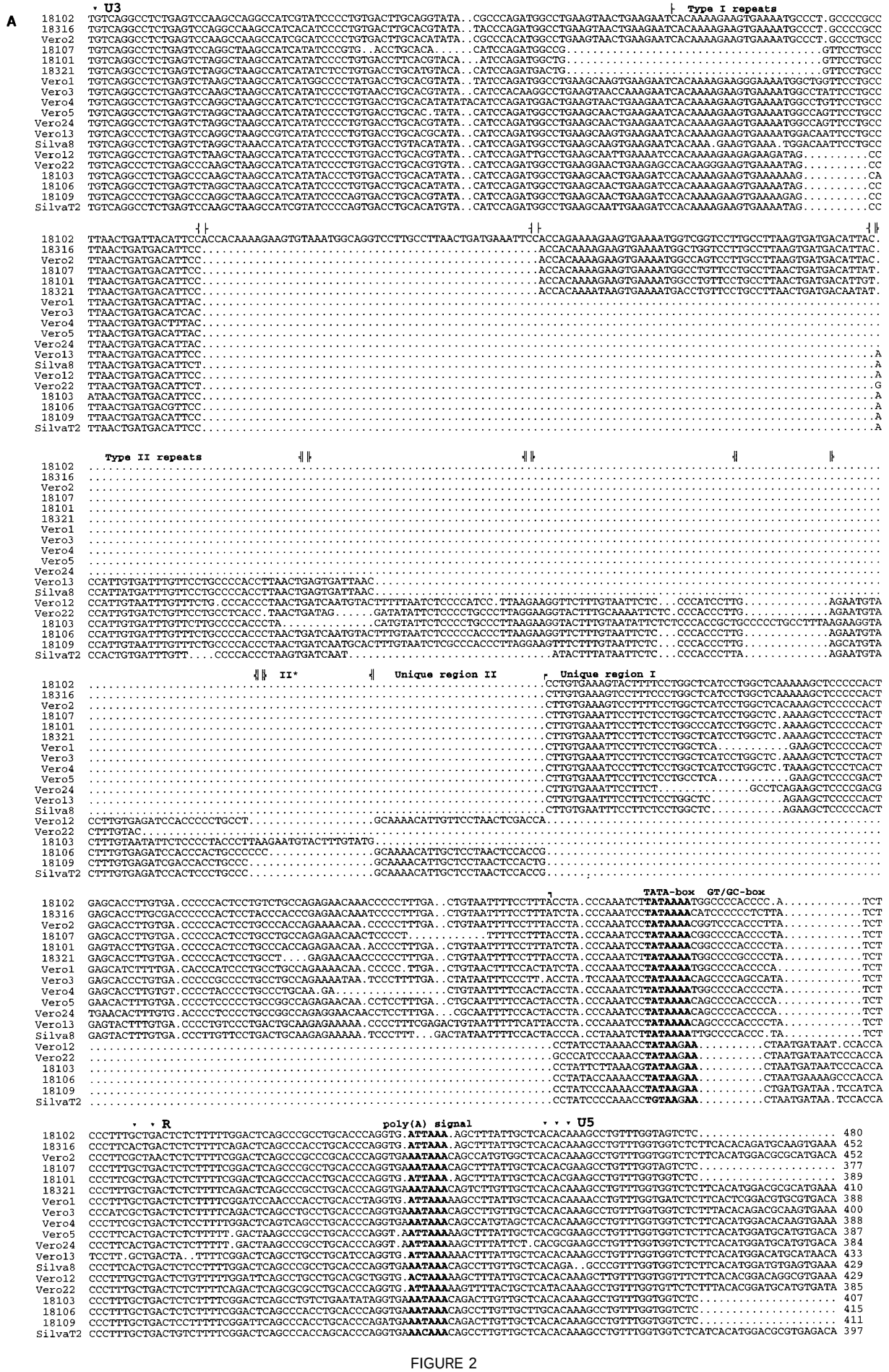

A
U3
Type I repeats
Type II repeats
II*
Unique region II
Unique region I
TATA-box
GT/GC-box
R
poly(A) signal
U5

FIGURE 2

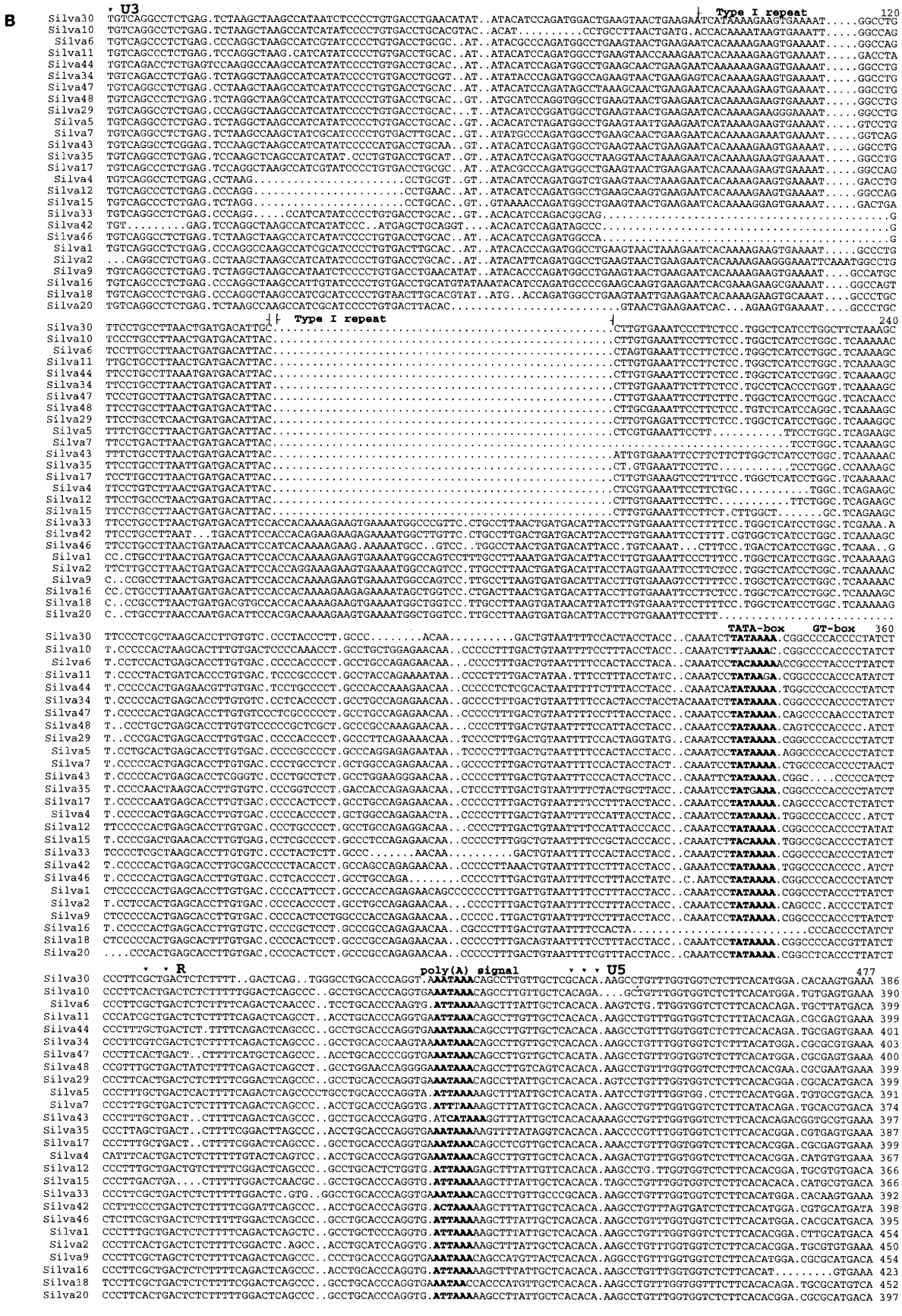


FIG. 2—*Continued*

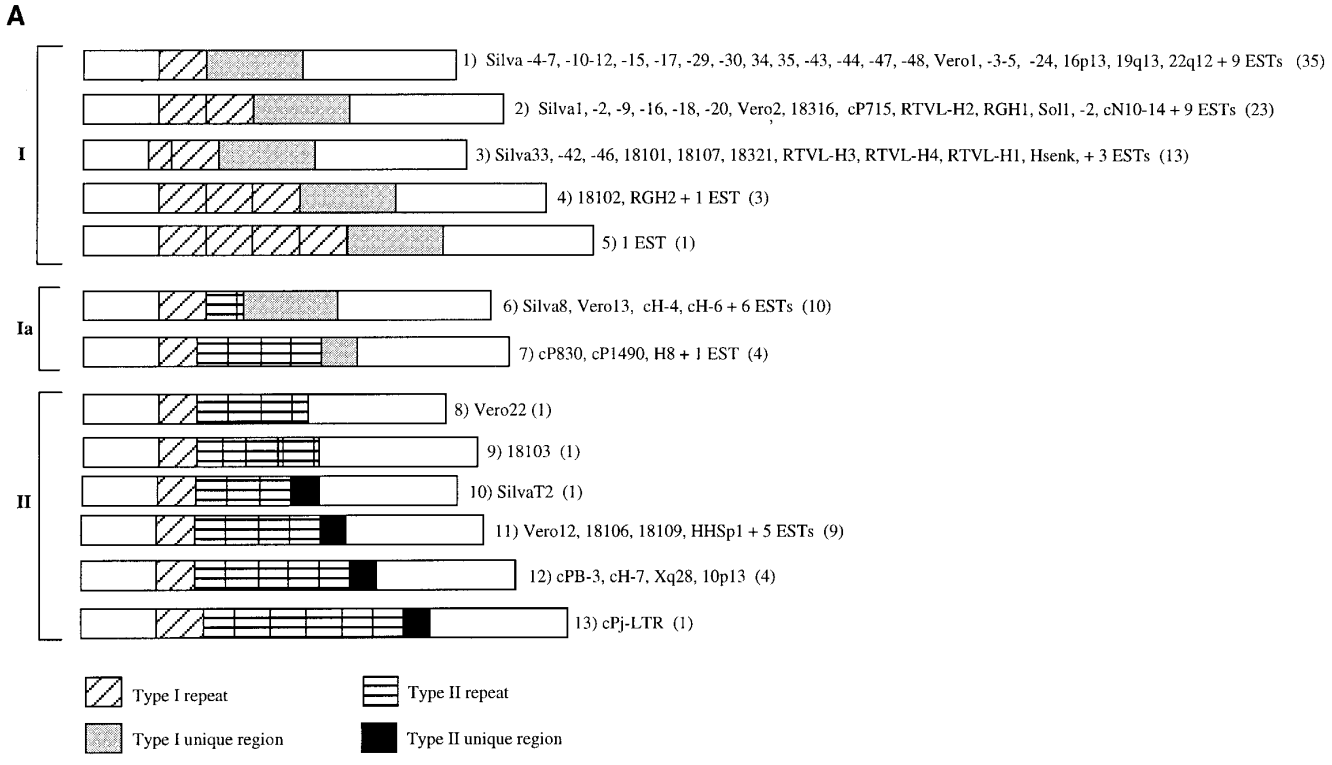
random selection. Of the sequenced clones, three duplicate clones were found among the human chromosome 18 LTRs, two among the marmoset LTRs, and one among the African green monkey LTRs. Database searches revealed that none of the eight different chromosome 18 LTR sequences were identical to previously reported sequences. Neither African green monkey nor marmoset ERV-H LTR sequences have been reported previously. A multialignment of the human chromosome 18 LTRs (18101, 18102, 18103, 18106, 18107, 18109, 18316, and 18321), the 9 different African green monkey LTRs (Vero1, Vero3, Vero4, Vero5, Vero12, Vero13, Vero22, Vero24) and two of the marmoset LTRs (Silva8 and SilvaT2) is shown in Fig. 2A, while the remaining 26 marmoset sequences are aligned in Fig. 2B. Three different subfamilies of human ERV-H LTRs, designated Type I, Type Ia, and Type II, have been reported (Mager, 1989; Goodchild *et al.,* 1993). These LTRs are very similar over the first 80 bp of the U3 region, and in the R and U5 regions. The remainder of the U3 region is for Type I sequences characterized by the presence of one or more Type I repeats, while Type II LTRs contain one copy of the Type I repeat in addition to several copies of a Type II repeat. Type Ia is a combination between Type I and Type II, and may have arisen by recombination between Type I and Type II LTRs (Goodchild *et al.,* 1993). Strikingly, 26 of the 28 marmoset sequences represent Type I LTRs. We found the LTRs to vary in length from 388 to 501 bp. This length variation is largely due to the number and types of repeats found in the U3 region. In Figs. 2A and 2B the LTRs are aligned according to their repeat structure, and the borders of the different repeats are indicated above the sequences. The Type I LTRs contain a 110- to 120-bp unique region upstream of the TATA box, while the Type II LTRs, except for 18103 and Vero22, contain a unique region of about 25 bp at this position (see also Fig. 3A). Both Type I and Type Ia LTRs contain a consensus TATA box in the 3′ part of the U3 region. The only exception is the marmoset LTR Silva16 which has a 25-bp deletion in this area. Type II LTRs contain an A to G transition at position 6 of the TATA box (Fig. 2A). In the Type I and Type Ia LTRs a GC-rich region of 10 to 12 bp is located just 3′ to the TATA box often creating a consensus GT or GC box. In the Type II LTRs an AT-rich region is succeeding the TATA box. Thus, the context of the basal promoter elements is different in Type I and Type II LTRs. Both Type I and Type II LTRs contain a putative initiator (Inr) element centered around the transcription initiation site (Feuchter and Mager, 1990), a polyadenylation signal in the R region, and one to three CA dinucleotides 15 to 20 bases further downstream functioning as a polyadenylation site(s) at the border between the R and U5 region (Johansen *et al.,* 1989; Mager, 1989; Wilkinson *et al.,* 1990; Goodchild *et al.,* 1992).

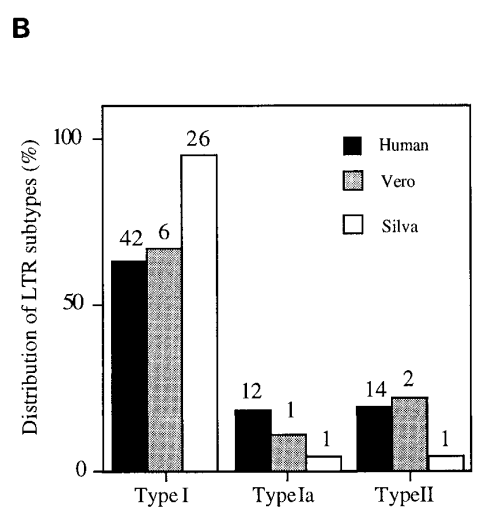
Intrinsic curvature within the Type I repeat

The Type I repeats display a modular nature with an A-rich 5′ stretch of about 20 bp, a central GC-rich segment of 13–14 basepairs, and a 3′ AT-rich region (Fig. 2). The 5′ A-rich stretch contains two oligo(A) tracts phased with the pitch of the DNA helix and is thus a canditate for inducing a bend or distortion of the DNA structure (Wu and Crothers, 1984; Koo *et al.,* 1986). In fact, we observed that 7 of 9 human LTRs tested showed anomalous migration in polyacrylamide gels, indicating a distorted DNA structure. This prompted us to perform a more detailed mobility study using polyacrylamide gels. Anomalous migration due to a distorted DNA structure decreases with increased temperature (Mizuno, 1987) and in the presence of the DNA binding drug distamycin (Radic *et al.,* 1987). Gel electrophoresis at 4°, 30°, and in the presence of distamycin demonstrated that all Type I LTRs tested showed anomalous migration, particularly subtype 3 LTRs. The single Type Ia and two Type II LTRs migrated as expected from their lengths with one exception. LTR 18109 showed a slightly retarded migration at low temperatures. The U3 region upstream of the TATA box showed a similar migration as the full-length fragments (data not shown). Thus, the distorted DNA stucture seems to be located within the repeated sequences in the U3 region. Computer analyses (Goodsell and Dickerson, 1994) predicted the location of a DNA bend in the A-rich stretch of the Type I repeat. Finally, circular permutation assays of the Type I repeat of 18321 and 18102 revealed that this repeat has an intrinsically bent or distorted DNA structure (Fig. 4). The Type I repeat of 18321 shows a significantly larger bend than the Type I repeat of 18102. Hence, differences in the specific sequences of the Type I repeats of individual LTRs may influence the degree of intrinsic curvature.

Distribution of the different LTR subtypes

Alignments of the marmoset, African green monkey, and human chromosome 18 LTRs with previously reported human LTRs (Mager and Henthorn, 1984; Mager and Freeman, 1987; Johansen *et al.,* 1989; Mager, 1989; Feuchter and Mager, 1990; Wilkinson *et al.,* 1990; Goodchild *et al.,* 1992, 1993, 1995; Hirose *et al.,* 1993) as well as 35 expressed sequence tags (ESTs), detected by searching GenBank, revealed that the three LTR types can be further divided into 13 different subtypes. A schematic compilation of these subtypes and the LTRs belonging to each of them is shown in Fig. 3A. We have classified the sequences into subtypes based on the number of Type I or Type II repeats and the presence or absence of the unique Type I or unique Type II regions. Diversity in the repeat structure may have orginated by replication slippage or slipped-strand mispairing (Levinson and Gutman, 1987). Such types of events occur in DNA regions containing contiguous short repeats up to 20–30 nucleotides long. Duplication or deletion of longer sequence stretches may have occured by unequal crossing over, homologous recombination, or gene conversion (Li and Graur, 1991). Most of the deletions/insertions





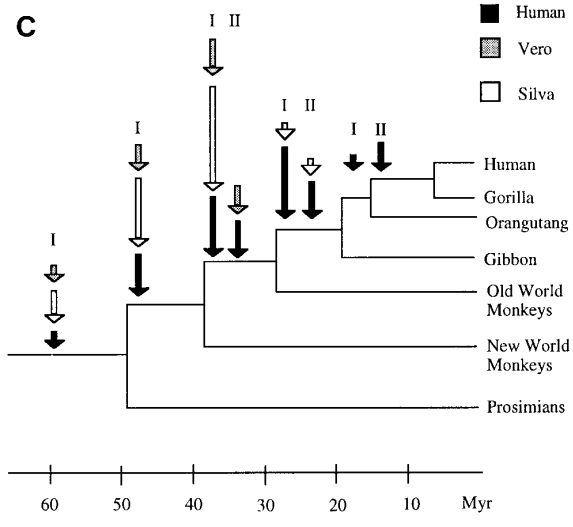


FIG. 3. The structure, distribution, and amplification of the different ERV-H LTR types. (A) A schematic representation of the sequence structure of the ERV-H LTR types and subtypes. A total of 106 LTR sequences were analyzed. The brackets to the left denote structures found within Type I, Type Ia, and Type II LTRs, while the different subtypes and the LTRs found to belong to each subtype are indicated to the right. Only the number of EST sequences within each group are indicated. The number of LTRs belonging to each subtype is shown in parenthesis to the right. (B) More than 60% of the LTR sequences belong to Type I. The column diagram displays the distribution of Type I, -Ia, and -II LTRs among the sequenced human, African green monkey (Vero), and marmoset (Silva) LTRs. The number of isolated elements within each group is indicated above the columns. (C) Diagram illustration of the possible integration/expansion times of the ERV-H elements during primate evolution based on the data presented in Table 1. The LTR types (I and II) are indicated by arrows of which lengths illustrate the number of elements of this type that may have integrated at the particular time period. The time scale is given below in MYr. The branchpoint times of the phylogeny are based on analyses of the molecular evolution of noncoding sequences of the β-globin gene cluster (Goodman *et al.,* 1994) and should be considered approximate.

probably result from single events. Hence, gaps appearing at identical positions in several sequences strongly suggest that these LTRs are derived from a common ancestor which encountered the particular event before the divergence of the sequences. The number of LTRs belonging to each type varies a lot (Fig. 3B). In the initial screening 26 of 27 sequenced marmoset LTRs belonged to Type I, 17 of these as the simplest "archetypical" subtype 1. One marmoset LTR was found to be of Type Ia, while no marmoset LTRs of Type II were isolated.

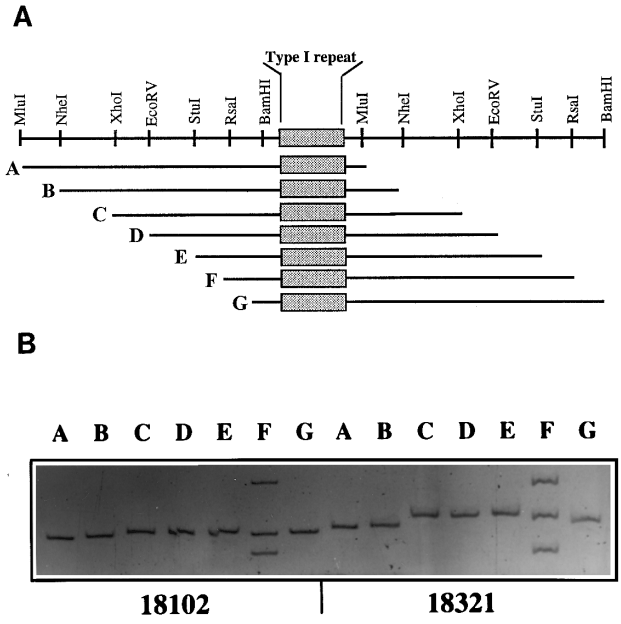


FIG. 4. The Type I repeat contains an intrinsic DNA curvature. (A) DNA fragments used for circular permutation analyses. Seven around 180-bp-long fragments denoted A to G were generated by restriction endonuclease cleavage as indicated. The Type I repeats (generated by *Mse*I digestion) inserted into the *Xba*I site of pBend2 is represented by a shaded rectangle. (B) The more retarded bands of the LTR 18321 show that the Type I repeat of LTR 18321 bend DNA more extensively than the Type I repeat of LTR 18102.

However, the presence of marmoset Type Ia LTRs indicated that Type II elements exist, or have existed, in New World monkeys since Type Ia LTRs are thought to have arisen by a recombination event between Type I and Type II elements (Goodchild *et al.,* 1993). Because several of the human Type II LTRs have been observed to contain a PBS complementary to $tRNA^{Phe}$ instead of $tRNA^{His}$ (Goodchild *et al.,* 1995), we performed PCR on genomic DNA from the marmoset with a new 3′ primer aligning to a $tRNA^{Phe}$ PBS. A faint band of about 500 bp was obtained. Sequencing revealed this to be a Type II LTR similar to subclass 11, but with three Type II repeats instead of four (Fig. 3A). Hence, this marmoset Type II element defined a new LTR subtype.

The faint PCR product obtained together with our failure to isolate Type II elements from marmoset using the $tRNA^{His}$ PBS primer suggest a significantly lower number of Type II elements in New World monkeys than in Old World monkeys and humans where the number of isolated Type II elements is about 20%. The frequency of marmoset Type I elements is about 95% while about 60% of the sequenced human and African green monkey LTRs are of Type I (Fig. 3B). From Southern analyses with cloned *pol* probes Mager and Freeman (1995) estimated the copy number of marmoset ERV-H elements to between 25 and 50. Our isolation of 28 different marmoset 5′ LTRs that are linked to internal ERV-H sequences suggests that we have characterized a majority of marmoset ERV-H LTRs linked to either internally deleted or full-length elements. Solitary LTRs, which most certainly also exist, would not be detected by our strategy. Taken together, these findings suggest that Type I elements have expanded to a copy number of about 50 before the divergence of New World monkeys and Old World monkeys, while Type II elements, in contrast, most probably expanded after the divergence. A second major expansion of Type I elements occured after the split of these primate lineages correlated with the amplification of the common deleted subfamily of ERV-H elements (Mager and Freeman, 1995).

Relative evolutionary age of ERV-H LTRs estimated from sequence divergence

In order to obtain relative estimates of the time frames in which the different ERV-H elements may have integrated into the cellular genomes we employed a consensus approach where consensus sequences for different LTR subtypes were calculated. The percentage divergence from the relevant consensus was then determined for each element (Table 1), employing the Kimura two-parameter method (Kimura, 1980). The approximate ages of the elements was calculated using a nucleotide mutation rate of 0.2% per MYr based on interspecies distances of noncoding sequences and time of primate evolutionary branchpoints (Goodman *et al.,* 1994, 1990) (Table 1). We found the evolutionary ages of the Type I elements to range from 18 to 67 MYr, with most elements between 30 and 50 MYr. The 26 marmoset Type I elements allowed the calculation of an average age of about 43 MYr (Fig. 3C and Table 1). Interestingly, a few Type I elements seem to be very old, between 51 and 67 MYr (Vero1, Silva7, -18, -15, and 19q13). Hence, the corresponding ERV-H elements may have integrated into the genomes even before the divergence between prosimians and New World monkeys, indicating that ERV-H sequences may be found in prosimians.

The average age of Type II elements (32 MYr) indicates amplification around the time when Old World monkeys diverged from New World monkeys (see Fig. 3C). Their frequency suggests that they have had a lower transposition activity than Type I elements.

Type Ia elements have been suggested to constitute the youngest subtype, having experienced a major expansion after the divergence between the orangutang and the gorilla lineage (Goodchild *et al.,* 1993). However, we isolated a Type Ia LTR both from marmoset (Silva8) and African green monkey (Vero13), indicating their presence in both New- and Old World monkeys (Fig. 3). Since the number of isolated sequences within each subgroup was too low to derive a reliable consensus sequence, their relative ages could not be calculated.

The greatest variability both in structure (Fig. 3A) and age (Table 1) is found in the human lineage, including the youngest (Xq28, 13 MYr) and the oldest (19q13, 67 MYr) elements. The chromosome 18 LTRs are clustered

TABLE 1

Sequence Divergence[a] and Estimates of Evolutionary Age[b] of ERV-H LTRs

Marmoset type I			Type I-1			Type I-2-4			Type II		
Element	Percentage div.	Age	Element	Percentage div.	Age	Element	Percentage div.	Age	Element	Percentage div.	Age
Silva29	6.0	30	22q12	4.1	21	18107	3.5	18	Xq28	2.2	11
Silva17	6.0	30	16p13	5.2	26	RTVL-H4	4.8	24	18106	3.0	15
Silva2	6.1	31	19q13	13.3	67	RTVL-H1	5.0	25	Hhsp1	4.2	21
Silva30	6.6	33	Vero4	6.0	30	cP715	5.1	26	10p13	4.5	23
Silva34	6.8	34	Vero24	7.2	36	18101	5.4	27	cPB-3	5.1	26
Silva46	7.0	35	Vero5	7.4	37	Hsenk	6.0	30	18109	5.3	27
Silva47	7.1	36	Vero3	7.8	39	RTVL-H3	6.3	32	cH-7	6.2	31
Silva44	7.7	38	Vero1	10.1	51	RTVL-H2	6.7	34	18103	10.8	54
Silva12	7.8	39				Sol1	6.8	34	cPj-LTR	11.7	59
Silva4	8.1	41				18102	6.9	35	Vero12	7.5	37
Silva11	8.3	42				Sol2	7.4	37	Vero22	9.8	49
Silva20	8.3	42				18321	7.6	38	SilvaT2	5.6	28
Silva9	8.5	43				RGH2	8.1	41			
Silva48	8.5	43				cN10-14	8.3	42			
Silva6	8.9	44				18316	8.4	42			
Silva16	9.0	45				RGH1	8.7	44			
Silva5	9.0	45				Vero2	7.2	36			
Silva33	9.6	48									
Silva10	9.6	48									
Silva1	9.8	49									
Silva43	9.8	49									
Silva42	10.0	50									
Silva35	10.0	50									
Silva7	10.8	54									
Silva18	11.5	57									
Silva15	11.6	58									

[a] The Kimura two-parameter method (Kimura, 1980) and a[b] divergence rate of 0.2% nucleotide differences/MYr were used. Separate consensus sequences were calculated for the marmoset Type I LTRs, the Type I-2-4 LTRs and the Type II LTRs. Seventeen marmoset LTR sequences were included in the multialignment and calculation of the consensus for the Type I-1 data set. The average divergence calculated for marmoset Type I LTRs was 8.6 $\pm$ 1.6% corresponding to an average evolutionary age of 43 $\pm$ 8 MYr. For the human and African green monkey Type I LTRs (Type I-1 and Type I-2-4 columns) the average divergence was 6.9 $\pm$ 2% (35 $\pm$ 10 MYr), while the Type II LTRs showed an average divergence of 6.3 $\pm$ 2.9%, giving an average evolutionary age of 32 MYr.

in three groups, indicating three time frames of integration/expansion (Table 1). The first is around the divergence between New- and Old World monkeys, the second around the divergence of Old World monkeys from the apes, and the third around the divergence between the orangutang and the gorilla lineages. The ages of the marmoset elements (Table 1), however, are clustered around the time of divergence between New- and Old World monkeys (35–40 MYr). This may suggest, as speculated (see, i.e., Li and Graur, 1991) that transposition events may have been involved in the speciation of primates.

When both phylogenetic distribution, sequence divergence (evolutionary age), and structural organization of the ERV-H LTRs are taken into account the Type I-1 LTRs (see Fig. 3) can be considered the archetypical ERV-H LTRs. Type I-2, -I-4, and -I-5 have arisen through duplication(s) of the Type I repeat while Type I-3 are derived from Type I-2 by a deletion removing the 5′ half of the first Type I repeat and part of the upstream region (Figs. 2 and 3).

As is appearant from Fig. 3A, type II LTRs contain 3 to 6 Type II repeats. Subtypes 8 and 9, represented by only one LTR each, have probably arisen by deletion of the Type II unique region. The likelihood of such events would increase with evolutionary age and these two elements are among the "oldest" Type II LTRs.

The different structures of the two Type Ia subtypes (6 and 7 in Fig. 3A) suggest that Type Ia elements may have arisen at least twice by independent recombination events between Type I and Type II LTRs.

Promoter active LTRs are found both in Old- and New World monkeys

To evaluate the HERV-H LTRs for promoter activity the five Type I LTRs and the three Type II LTRs from chromosome 18 were inserted upstream of the promoterless CAT gene in pBLCAT3 and assayed by transient transfection in three different human cell lines. The three LTRs found to display promoter activity were of Type I (Figs. 5A and 5B). The activities varied both between the

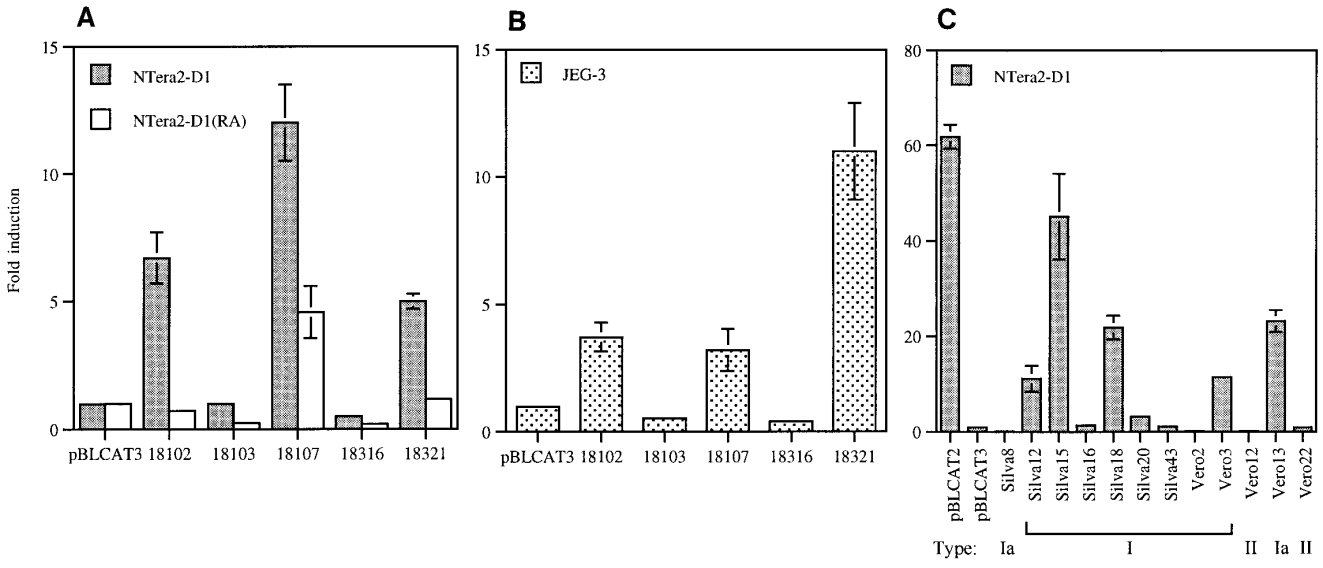


FIG. 5. ERV-H LTRs of Type I and Type Ia display promoter activity in NTera2-D1 and JEG-3 cell lines. (A) The transcriptional activity of HERV-H LTRs is reduced upon differentiation of NTera2-D1 cells. Reporter vectors (8 μg) containing four human Type I LTRs (18102, 18107, 18316, and 18321) and one Type II LTR (18103) inserted upstream of the CAT gene in the promoterless pBLCAT3 vector were transfected into untreated NTera2-D1 cells and NTera2-D1 cells treated for 5 days with retinoic acid (RA). The CAT activity of the promoterless pBLCAT3 vector without any insert was assigned a value of 1.0. The data represent the means of three independent experiments using different plasmid preparations. The error bars indicate standard errors of the mean. (B) The reporter vectors used in (A) were transfected into JEG-3 cells as descibed for the NTera2-D1 cells in (A). (C) Transcriptional active LTRs were found among both New World monkey and Old World monkey Type I and Ia LTRs. Seven New World monkey LTRs (Silva8, Silva12, Silva15, Silva16, and Silva18) and five Old World monkey LTRs (Vero2, Vero3, Vero12, Vero13, and Vero22) were inserted upstream of the CAT gene in pBLCAT3 and transfected into NTera2-D1 cells. The CAT activity of the promoterless pBLCAT3 without any insert was assigned a value of 1.0. The pBLCAT2 plasmid, containing the herpes simplex virus thymidine kinase promoter in front of the CAT gene, was used as positive control. The data represent the means of three independent experiments using different plasmid preparations. The error bars indicate standard errors of the mean.

cell lines and with the differentiation state of the NTera2-D1 cells. Undifferentiated NTera2-D1 cells induced the highest promoter activity with a 5- to 13-fold induction depending on the specific LTR analyzed. Retinoic acid treatment of the NTera2-D1 cells for 5 days, which leads to differentiation (Andrews, 1984), resulted in a nearly threefold reduction of the promoter activity (Fig. 5A). This is in agreement with Northern blot analyses showing that the expression of HERV-H sequences is greatly reduced in NTera2-D1 cells induced to differentiate (data not shown and Wilkinson *et al.,* 1994). High-level expression of HERV-H elements are observed in normal placenta (Johansen *et al.,* 1989; Wilkinson *et al.,* 1990). Consistently, we found that the three chromosome 18 LTRs showed promoter activity in the placental choriocarcinoma cell line JEG-3. LTR 18321, in particular, showed a 10- to 12-fold induction in this cell line (Fig. 5B). In HeLa cells, however, all the chromosome 18 LTRs tested displayed a very low promoter activity, with only a twofold induction of the three active LTRs (data not shown). The low activity in HeLa cells is completely consistent with Northern analyses performed with a Type I-specific probe (Goodchild *et al.,* 1993). Thus, the promoter activity of Type I HERV-H LTRs seems to be cell-specific and to correlate with the expression pattern of HERV-H mRNAs.

Transient transfection assays in NTera2-D1 cells, including seven of the marmoset and five of the African green monkey ERV-H LTRs revealed that three of the marmoset and two of the African green monkey LTRs induced a 5- to 45-fold induction of promoter activity (Fig. 5C). Especially, the marmoset LTR Silva15 displayed a strong promoter activity, close to the activity of the herpes simplex virus thymidine kinase promoter in pBLCAT2. The three promoter active marmoset LTRs all belong to Type I LTRs, with Silva12 and Silva15 of subtype 1 and Silva18 of subtype 2. One of the promoter active African green monkey LTRs was of Type I, subtype 1 (Vero3), while the other was of Type Ia, subtype 6. Thus, as for the human chromosome 18 LTRs, the two African green monkey Type II LTRs did not show promoter activity. This is in agreement with Northern blot analyses reported by Goodchild *et al.* (1993), showing nearly no expression of Type II elements in a variety of cell lines. In addition, they demonstrated high level expression of Type I elements in embryonal teratocarcinoma cells, like NTera2-D1, but low or no expression in the other cell lines tested. Type Ia elements, on the other hand, were found to be expressed in a wide range of cell lines.

We have recently shown that transcriptional activation of HERV-H LTRs isolated from human chromosome 18 is dependent on Sp1 family proteins binding to the GC/GT box located 3′ to the TATA box (Sjøttem *et al.,* 1996). The lack of promoter activity for the Type I LTR 18316 and the Type II LTR 18103 is due to the absence of a

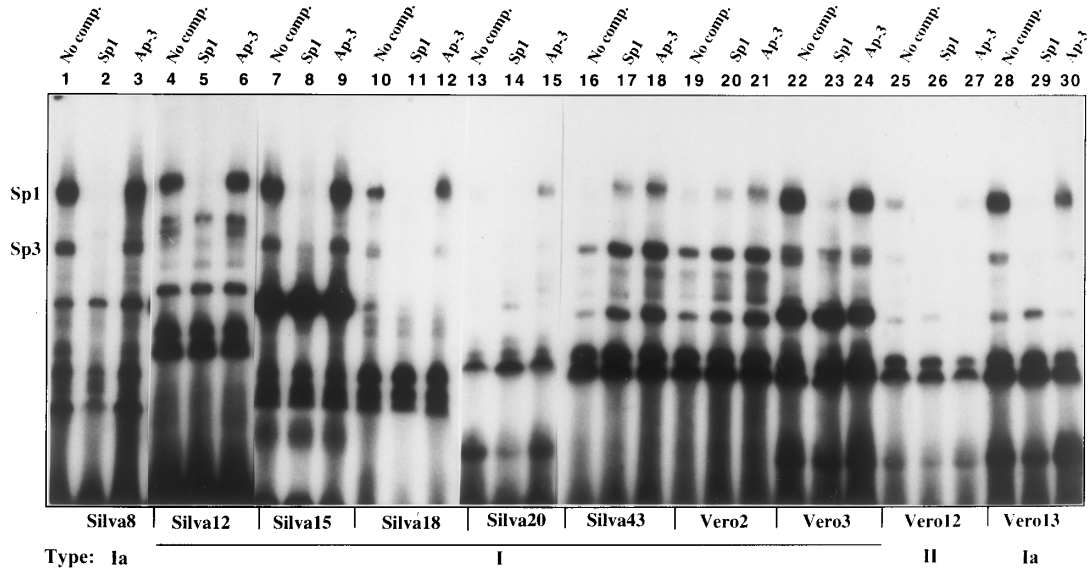


FIG. 6. Sp1 family proteins bind to the TATA-proximal GC/GT box of the transcriptional active marmoset and African green monkey LTRs. Nuclear extracts (2 μg) from NTera2-D1 cells were incubated with 65-bp-labeled fragments spanning the TATA box and the adjacent GC/GT box of marmoset (lanes 1 to 18) and African green monkey (lanes 19 to 30) LTRs. Oligonucleotide competitors (100 ng) containing consensus binding sites for the transcription factors Sp1 and Ap-3 were added as indicated. The complexes were separated on a 4% (39:1) polyacrylamide gel. The specific protein–DNA complexes inhibited by competition with the Sp1 oligonucleotide are indicated (Sp1 and Sp3). These complexes have previously been shown to be due to binding of Sp1 and Sp3 by supershift analyses of chromosome 18 LTRs using specific antibodies (Sjøttem *et al.,* 1996).

functional GC/GT box at this position. In order to determine if Sp1 family proteins bound to the basal promoter elements of the transcriptionally active African green monkey- and marmoset LTRs, we performed GMSA with nuclear extracts from NTera2-D1 cells incubated with labeled 65-bp fragments spanning from the TATA box to the transcription initiation site of the LTRs (Fig. 6). Generally, we found that the binding of Sp1 family proteins correlated well with the transcriptional activity. The transcriptionally active African green monkey LTRs Vero3 and Vero13 showed high intensity Sp1–DNA complexes, while no or very weak complexes could be seen for the promoter inactive Vero2 and Vero12. Similarly, the transcriptionally active marmoset LTRs Silva12, Silva15, and Silva18 contained a Sp1–DNA complex, while no such complex was seen for the promoter inactive marmoset LTRs with one exception. Silva8, the only isolated Type Ia element from marmoset, displayed a high intensity Sp1–DNA complex but was found to be promoter inactive. Thus, for these LTRs binding of Sp1 family proteins seems to be necessary for promoter activity, but is not always sufficient.

Strong promoter activity may be due to synergism between Sp1 binding sites in the U3 region

As mentioned above, we have previously found that the promoter active chromosome 18 LTRs contain a GC/GT box just 3′ to the TATA box, and that this GC/GT box is required for promoter activity. However, recently the group of Mager (Nelson *et al.,* 1996) employed 5′ deletions to demonstrate that the highly promoter active Type Ia LTR H6 contained two Sp1 binding sites in the U3 enhancer region that seemed to be required for promoter activity, and they suggested that the GT box 3′ to the TATA box was of much lower importance. To delineate the role of the different Sp1 binding sites more directly we constructed three distinct mutants of the chromosome 18 promoter active LTR 18321 (a Type I LTR) and four different mutants of the Type Ia LTR H6 (see Fig. 7A). Transient transfections in NTera2-D1 cells showed that generation of a consensus GC box in the Type I repeat of 18321 had little significant effect on the promoter activity (see 18321-mI in Fig. 7A). However, mutation of the GC box 3′ to the TATA box reduced its transcriptional activity to background levels (18321-mII in Fig. 7A), and insertion of a consensus Sp1 binding site in the Type I repeat was not able to compensate for this loss of activity (18321-mIII in Fig. 7A). In fact, the activity increased only slightly above background levels. Gel mobility shift assay with nuclear extracts showed that Sp1 bound to the consensus GC boxes, and not at all or very weakly to the mutated sites (Figs. 7B and 7C). Hence, for LTR 18321 the Sp1 binding site 3′ to the TATA box seems to be required for promoter activity, which is in agreement with our previous findings (Sjøttem *et al.,* 1996). The promoter activity of the Type Ia LTR H6, however, showed a behavior nearly opposite to LTR 18321 (Fig. 7A). Mutation of the GC box located in its Type I repeat reduced its transcriptional activity with about 30% (H6-mI in Fig. 7A), while mutation of both GC boxes located in the U3 enhancer region reduced the promoter activity with about 75% (H6-mII in Fig. 7A). In contrast, mutation of the GT box 3′ to the TATA box did not negatively affect the transcriptional

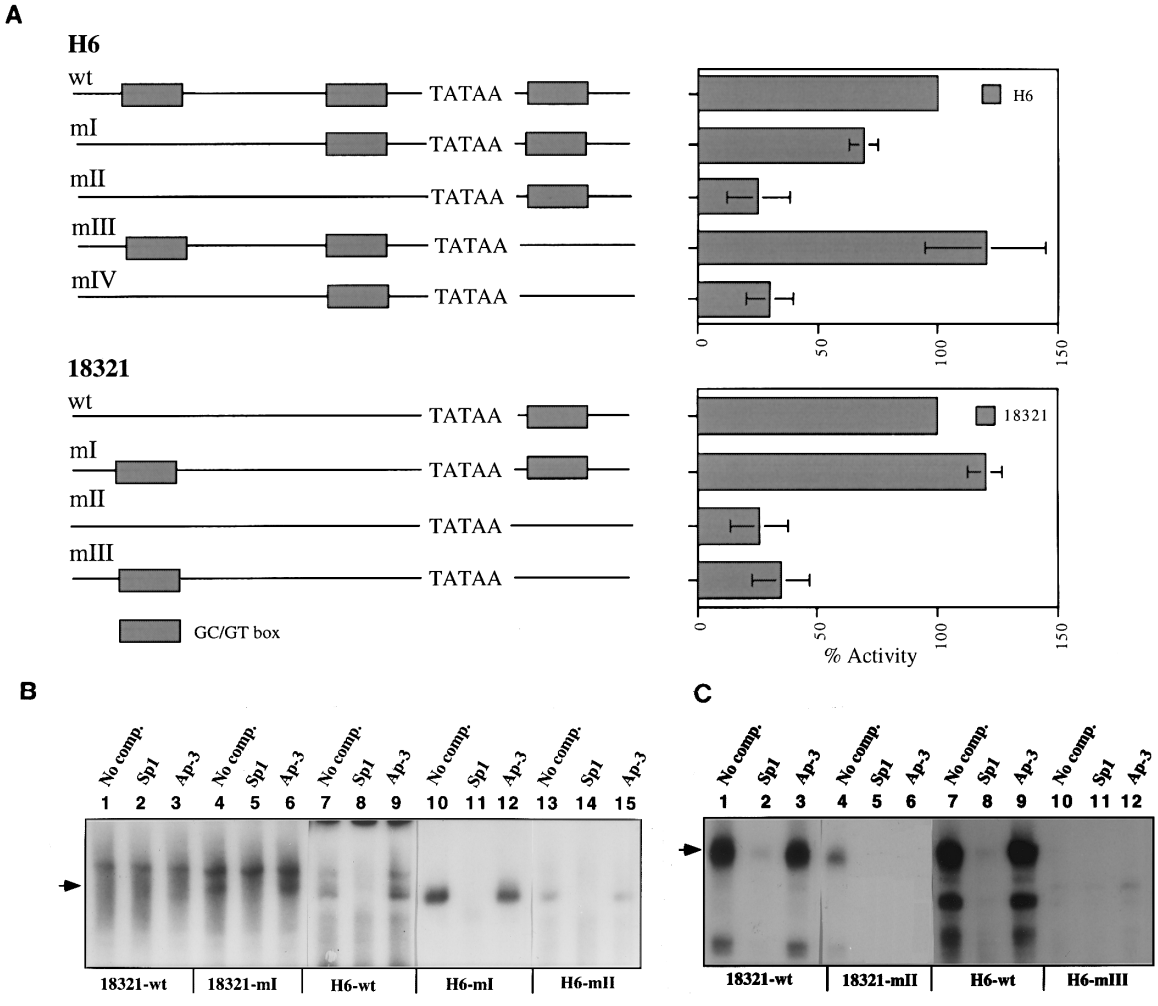


FIG. 7. Sp1 binding sites located in the U3 enhancer region or just 3′ to the TATA box are important for promoter activity of HERV-H LTRs. (A) The Type Ia LTR H6 has achieved high promoter activity due to synergistically acting GC boxes located in the U3 enhancer region, while the Type I LTR 18321 is critically dependent on the GC box 3′ to the TATA box for its relatively weaker promoter activity. Reporter vectors (8 μg) containing wild-type and mutated H6 and 18321 LTRs inserted upstream of the CAT gene in pBLCAT3 were transfected into NTera2-D1 cells. The location of functional GC/GT boxes (grey rectangles) in the different constructs are indicated to the left. The CAT activity of wild-type H6 is set to 100% in the upper graph, while the CAT activity of wild-type 18321 is set to 100% in the lower graph. The values represent the means from six independent experiments using different preparations of the plasmids. Error bars indicate standard errors of the mean. (B) Sp1 binds to the consensus GC/GT boxes but not at all, or very weakly, to the mutated ones. Nuclear extracts (2 μg) from Sp1-deficient *Drosophila* SL-2 cells transfected with the Sp1 expression plasmid pPac-Sp1 were incubated with labeled fragments of the U3 region of LTR 18321 (lanes 1–6) or LTR H6 (lanes 7–15), spanning from position 1 to the TATA box. Oligonucleotide competitors (50 ng) containing consensus binding sites for Sp1 or Ap-3 were added as indicated. The arrowhead to the left indicates the location of the Sp1-DNA complexes. (C) Nuclear extracts (2 μg) from NTera2-D1 cells were incubated with labeled fragments of the U3 region of LTR 18321 (lanes 1–6) or LTR H6 (lanes 7–12) spanning from the TATA box to the R region. Oligonucleotide competitors (50 ng) containing consensus binding sites for Sp1 or Ap-3 were added as indicated. The arrowhead to the left indicates the Sp1–DNA complexes. The faster-migrating complexes competed by the Sp1 competitor oligonucleotide is due to binding of Sp3 as previously determined (Sjøttem *et al.,* 1996).

activity, but rather resulted in a slightly positive effect (H6-mIII in Fig. 7A). Hence, these results indicated that for the H6 LTR a synergistic interaction between the two GC boxes in the U3 enhancer region is important for full promoter activity, while the GT box 3′ to the TATA box was dispensable and even affected the transcriptional activity slightly negatively. Gel mobility shift assays with probes spanning the H6 U3 enhancer region showed that a Sp1 oligonucleotide competitor removed two complexes from the H6-wt probe, and only one complex from the mutated ones (Fig. 7B), confirming that two or more Sp1 proteins could bind simultanously to the H6-wt U3 enhancer region. Surprisingly, mutation of the GT box 3′ to the TATA box of the H6-mI construct, resulting in H6-mIV harboring only one functional GC box in the U3 region, reduced the transcriptional activity with 50% compared to H6-mI (Fig. 7A). This suggests that the Sp1 binding site 3′ to the TATA box is important for transcriptional activity when there are no synergistically acting Sp1 binding sites in the U3 enhancer region. Also, it must be mentioned that the transcriptional activity of the H6 LTR is four- to fivefold higher than the transcriptional

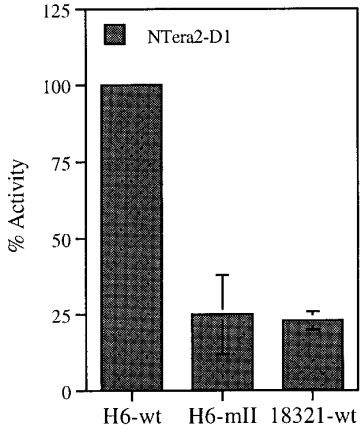


FIG. 8. The promoter activity of the Type Ia LTR H6 is four- to fivefold higher than that of the Type I element LTR 18321. Reporter vectors (8 μg) containing wild-type and mutated H6 and wild-type 18321 LTRs inserted upstream of the CAT gene in pBLCAT3 were transfected into NTera2-D1 cells. The CAT activity of wild-type H6 is set to 100%. The values represent the means from two independent experiments using different preparations of the plasmids. Error bars indicate standard errors of the mean.

activity of LTR 18321. Hence, the H6-mII mutant with no consensus GC boxes in the U3 enhancer region, but retaining the GT box 3′ to the TATA element, has an activity similar to wild-type 18321 (Fig. 8).

Taken together these results suggest that the Type Ia LTR H6 has achieved strong promoter activity due to synergistically acting GC boxes in its U3 enhancer region. Conversely, the relatively weaker Type I LTR promoters, without functional Sp1 binding sites in their U3 regions, are dependent on Sp1 family proteins binding to a high-affinity GC/GT box just 3′ to the TATA box for transcriptional activity.

DISCUSSION

In the present work we have isolated and sequenced 45 LTRs of the endogenous retrovirus family ERV-H from human chromosome 18, an Old World monkey, and a New World monkey. Comparative analyses of these sequences together with previously published HERV-H LTRs, and 35 EST sequences containing HERV-H LTRs, showed that the 3 ERV-H LTR types can be further divided into 13 subtypes. Type I elements were widely distributed both in the New World monkey, the Old World monkey, and man. Calculation of their relative evolutionary age showed that both very old and very young elements belonged to this subtype, and their average age indicated that they have expanded in the genomes around the time when Old World monkeys diverged from New World monkeys. However, the presence of elements of 55–70 MYr suggests that ERV-H elements may have integrated in the genomes even before the divergence of prosimians and New World monkeys. The great abundance and presence of very young Type I elements indicate that there has been and perhaps still is transpositionally active Type I elements. This is in line with the promoter activity analyses, showing that several Type I elements contained transcriptional activity. Interestingly, the marmoset Type I LTR estimated to be the oldest one, Silva15, showed the highest promoter activity. Another expansion of ERV-H elements appears to have occured after the time Old World monkeys diverged from New World monkeys. This expansion has also involved Type II elements, which were found to be present in both monkey species. About 20% of the elements in humans and African green monkey were found to be of Type II. Their relatively low abundance together with the fact that none of the promoter active LTRs were found to be of Type II, suggest that these elements have had a significant lower activity than Type I elements. The Type Ia elements are suggested to have amplified more recently (Goodchild *et al.*, 1993), and their activity in transient transfections indicates that there still are active elements of this subtype in the genomes.

HERV-H elements with Type II LTRs have been reported to represent 30–35% of all HERV-H sequences (Goodchild *et al.,* 1993, 1995), while we found they to comprise only 20%. This may in part be due to the observation that Type II LTRs often have a PBS most closely related to $tRNA^{Phe}$ (Goodchild *et al.,* 1995). Since we have used a 3′ primer annealing to $tRNA^{His}$-related PBS when isolating LTRs from Old World monkeys and from humans, the number of Type II elements may be underestimated. In fact, Type II elements from human chromosome 18 were obtained when we used a 3′ primer annealing to the U5 region of the LTRs (primerset 1, Fig. 1A), but not when we used the $tRNA^{His}$ related primer (primerset 2, Fig. 1A), indicating that the number of Type II elements with a PBS related to $tRNA^{His}$ is low in humans. However, the majority of the EST sequences, which are isolated independently of the nature of their PBS, are of Type I (about 65%), while 20% are of Type Ia and 15% of Type II. Hence, the majority of LTRs seems to be of Type I in all three species. This dominance is much more prominent in the marmoset than in Old World monkeys and man consistent with their estimated age and the proposed amplification of Type II elements after the divergence of New- and Old World monkeys.

Southern hybridization performed by another group (Goodchild *et al.,* 1993) showed Type Ia elements to be present only in humans, chimpanzee, and gorilla and not in Old World monkeys. Hence, they suggested Type Ia elements to be a relatively young, ape-specific subfamily. Our isolation of a Type Ia LTR from both an Old- and a New World monkey may be due to the increased sensitivity obtained with PCR compared to Southern hybridization, since the number of Type Ia elements is clearly very low in these species. The significant different structures of the two subtypes of Type Ia elements (see subtype 6 and 7 in Fig. 3A) suggest that two independent recombination events have occured. Interestingly, subtype 7 was

only found in humans, suggesting that this subtype arose later in primate evolution than subtype 6.

We based our calculations on the individual divergence of the LTRs on consensus sequences generated from alignments of different LTR subtypes. Except for the gaps, which are eliminated from the analysis, the nucleotide differences within each subtype are distributed fairly randomly, indicating a subsequent accumulation of mutations. Hence, determining the average divergence from the progenitor, here approximated by the consensus, should indicate the time since the LTR element was inserted into the genome. For Alu elements, this approach has been shown to correlate well with the ages of the sequences (Shen *et al.,* 1991; Zietkiewicz *et al.,* 1994). For six of the Type I elements, their relative ages have been calculated based on sequence divergences in their *pol* sequences (Mager and Freeman, 1995). Except for LTR cN10-14, this age corresponded reasonably well with the age calculated by us, with discrepancies within 3–10 MYr. This is also the case for three Type I and two Type II LTRs that Goodchild *et al.* (1993) have traced during primate evolution. Thus, our consensus approach should give a useful idea of the approximate ages of the individual elements.

Transient transfections in NTera2-D1 cells showed that strong pomoter activity, as displayed by H6, seems to require synergistically acting Sp1 binding sites in the U3 enhancer region. Weaker promoter activity, as shown by the Type I elements from human chromosome 18, is critically dependent on the high-affinity Sp1 binding site located just 3′ to the TATA box. Can these results be generalized for all the HERV-H LTR elements? None of the Type II LTRs, which lack the GC/GT box 3′ to the TATA box, or any Type I LTR harboring mutations at this site (i.e., 18316, Vero2, Silva20, and -43 in Fig. 5), were found to display promoter activity. Thus, even if this site seemed to be dispensable and even somewhat inhibitory for the strong H6 promoter, it may be necessary for the basal promoter activity of the HERV-H LTR promoters. However, why is it that the GC box introduced into the LTR 18321 U3 enhancer region, and putative GC boxes located in the U3 enhancer region of several of the other LTRs (such as the promoter inactive 18103 and 18316), does not stimulate the transcriptional activity? During our studies we have consistently observed that Sp1 binds with lower affinity to these sites than to the site 3′ to the TATA box. This could be due to a nonoptimal sequence context at these sites, other proteins competing for the same or overlapping binding sites, or to the specific DNA structures at these sites. Recent reports have shown that the actual DNA conformation at a protein binding site is important for target site selection (Parvin *et al.,* 1995; Starr *et al.,* 1995; Grove *et al.,* 1996). Since we found the Type I repeats to contain an intrinsic curvature, this could lead to an unfavorable conformation of DNA at the adjacent Sp1 binding site. In addition, positively and negatively acting transcription factors other than the Sp1 family proteins may bind to additional sites and affect the promoter activity. GMSA with nuclear extracts from various cell-lines have revealed protein–DNA complexes that were not competed by the Sp1 competitor oligonucleotide (Fig. 7 and Sjøttem *et al.,* 1996). The different expression levels of HERV-H elements in various tissues and cell-lines indicate tissue-specific regulation of HERV-H LTR promoters. The fact that Sp1 is a ubiquitously expressed transcription factor raises the question whether additional factors are needed to achieve cell-specific regulation of the HERV-H LTRs. Despite its ubiquitous expression Sp1 is involved in the tissue-specific regulation of several genes. Sp1 is also important for the inducible expression of specific genes and for the expression of a variety of cell-cycle regulated genes through its interaction with E2F (Karlseder *et al.,* 1996; Lin *et al.,* 1996). Interestingly, the DNA binding activity of Sp1 is down-regulated upon terminal differentiation of the liver (Leggett *et al.,* 1995). Hence, Sp1 may be involved in the tissue-specific expression observed for the HERV-H elements, and in the down-regulation of LTR promoters observed upon retinoic-acid induced differentiation of NTera2-D1 cells.

In addition to ERV-H, some low copy ERV families (Kröger and Horak, 1987; Cohen *et al.,* 1988; Perl *et al.,* 1989; Widegren *et al.,* 1996) and the high copy family ERV-K (Simpson *et al.,* 1996) are reported to be present in New World monkeys. The other known ERV families appeared after the split between New- and Old World monkeys. Similar to the ERV-H family, the ERV-K elements are represented as a low copy family in New World monkeys and have expanded in the genomes at the time Old World monkeys split from the hominoids (Steinhuber *et al.,* 1995). This is at the same time as we found the Type II ERV-H LTRs to have amplified. These results together with the proposed amplification of Type I elements around the time when New- and Old World monkey diverged enables the speculation that endogenous retrovirus-like elements may have been involved in the speciation of primates (Li and Graur, 1991; Travis, 1992).

ACKNOWLEDGMENTS

We are greatful to P. Andrews for NTera2-D1 cells and to D. Mager for the H6 LTR. D. Mager is also greatfully acknowledged for suggesting the use of a $tRNA^{Phe}$ PBS primer to detect marmoset Type II LTRs. This work was funded by grants from the Norwegian Cancer Society and the Aakre Foundation to T.J. E.S. is a fellow of the Norwegian Cancer Society. E.S. and S.A. contributed equally to this work.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Andrews, P. W. (1984). Retinoic acid induces neuronal differentiation of a cloned human embryonal carcinoma cell line in vitro. *Dev. Biol.* 103, 285–293.

Church, G. M., and Gilbert, W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA* 81, 1991–1995.

Cohen, M., Kato, N., and Larsson, E. (1988). ERV-3 human endogenous

provirus mRNA are expressed in normal and malignant tissues and cells, but not in choriocarcinoma tumor cells. *J. Cell. Biochem.* 36, 121–128.

Dangel, A. W., Mendoza, A. R., Baker, B. J., Daniel, C. M., Carroll, M. C., Wu, L. C., and Yu, C. Y. (1994). The dichotomous size variation of human complement C4 genes is mediated by a novel family of endogenous retroviruses, which also establishes species-specific genomic patterns among Old World primates. *Immunogenetics* 40, 425–436.

Feuchter, A., and Mager, D. (1990). Functional heterogeneity of a large family of human LTR-like promoters and enhancers. *Nucleic Acids Res.* 18, 1261–1270.

Feuchter, A. E., Freeman, J. D., and Mager, D. L. (1992). Strategy for detecting cellular transcripts promoted by human endogenous long terminal repeats: Identification of a novel gene (CDC4L) with homology to yeast CDC4. *Genomics* 13, 1237–1246.

Feuchter-Murthy, A. E., Freeman, J. D., and Mager, D. L. (1993). Splicing of a human endogenous retrovirus to a novel phospholipase A2 related gene. *Nucleic Acids Res.* 21, 135–143.

Fraser, C., Humphries, R. 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., Freeman, J. D., and Mager, D. L. (1995). Spliced HERV-H endogenous retroviral sequences in human genomic DNA: evidence for amplification via retrotransposition. *Virology* 206, 164–173.

Goodchild, N. L., Wilkinson, D. A., and Mager, D. L. (1992). A human endogenous long terminal repeat provides a polyadenylation signal to a novel, alternatively spliced transcript in normal placenta. *Gene* 121, 287–294.

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.

Goodman, M., Bailey, W. J., Hayasaka, K., Stanhope, M. J., Slightom, J., and Czelusniak, J. (1994). Molecular evidence on primate phylogeny from DNA sequences. *Am. J. Phys. Anthropol.* 94, 3–24.

Goodman, M., Tagle, D. A., Fitch, D. H. A., Bailey, W., Czelusniak, J., Koop, B. F., Benson, P., and Slightom, J. L. (1990). Primate evolution at the DNA level and a classification of hominoids. *J. Mol. Evol.* 30, 260–266.

Goodsell, D. S., and Dickerson, R. E. (1994). Bending and curvature calculations in B-DNA. *Nucleic Acids Res.* 22, 5497–5503.

Grove, A., Galeone, A., Mayol, L., and Geiduschek, E. P. (1996). Localized DNA flexibility contributes to target site selection by DNA-bending proteins. *J. Mol. Biol.* 260, 120–125.

Haltmeier, M., Seifarth, W., Blusch, J., Erfle, V., Hehlmann, R., and Leib-Mösch, C. (1995). Identification of S71-related human endogenous retroviral sequences with full-length pol genes. *Virology* 209, 550–560.

Hirose, Y., Takamatsu, M., and Harada, F. (1993). Presence of env genes in members of the RTVL-H family of human endogenous retrovirus-like elements. *Virology* 192, 52–61.

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, 6343–6348.

Karlseder, J., Rothender, H., and Wintersberger, E. (1996). Interaction of Sp1 with the growth- and cell cycle-regulated transcription factor E2F. *Mol. Cell. Biol.* 16, 1659–1667.

Kim, J., Zwieb, C., Wu, C., and Adhya, S. (1989). Bending of DNA by gene-regulatory proteins: Construction and use of a DNA bending vector. *Gene* 85, 15–23.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120.

Koo, H. S., Wu, H. M., and Crothers, D. M. (1986). DNA bending at adenine-thymine tracts. *Nature* 320, 501–506.

Kröger, B., and Horak, I. (1987). Isolation of novel human retrovirus-related sequences by hybridization to synthetic oligonucleotides complementary to the tRNA(Pro) primer-binding site. *J. Virol.* 61, 2071–2075.

Lebedev, Y. B., Volik, S. V., Obradovic, D., Ermolaeva, O. D., Ashworth, L. K., Lennon, G. G., and Sverdlov, E. D. (1995). Physical mapping of sequences homologous to an endogenous retrovirus LTR on human chromosome 19. *Mol. Gen. Genet.* 247, 742–748.

Leggett, R. W., Armstrong, S. A., Barry, D., and Mueller, C. R. (1995). Sp1 is phosphorylated and its DNA binding activity down-regulated upon terminal differentiation of the liver. *J. Biol. Chem.* 270, 25879–25884.

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.

Levinson, G., and Gutman, G. A. (1987). Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* 4, 203–221.

Li, M. D., Bronson, D. L., Lemke, T. D., and Faras, A. J. (1995). Restricted expression of new HERV-K members in human teratocarcinoma cells. *Virology* 208, 733–741.

Li, W.-H., and Graur, D. (1991). Evolution by transposition. *In* "Fundamentals of Molecular Evolution," pp. 172–203. Sinauer, Sunderland, MA.

Lin, S.-Y., Black, A. R., Kostic, D., Pajovic, S., Hoover, C. N., and Azizkhan, J. C. (1996). Cell cycle-regulated association of E2F1 and Sp1 is related to their functional interaction. *Mol. Cell. Biol.* 16, 1668–1675.

Liu, A. Y., and Abraham, B. A. (1991). Subtractive cloning of a hybrid human endogenous retrovirus and calbindin gene in the prostate cell line PC3. *Cancer Res.* 51, 4107–4110.

Luckow, B., and Schütz, G. (1987). CAT constructions with multiple unique restriction sites for the functional analysis of eukaryotic promoters and regulatory elements. *Nucleic Acids Res.* 15, 5490.

Mager, D. L. (1989). Polyadenylation function and sequence variability of the long terminal repeats of the human endogenous retrovirus-like family RTVL-H. *Virology* 173, 591–599.

Mager, D. L., and Freeman, J. D. (1987). Human endogenous retroviruslike genome with type C pol sequences and gag sequences related to human T-cell lymphotropic viruses. *J. Virol.* 61, 4060–4066.

Mager, D. L., and Freeman, J. D. (1995). HERV-H endogenous retroviruses: Presence in the new world branch but amplification in the old world primate lineage. *Virology* 213, 395–404.

Mager, D. L., and Goodchild, N. L. (1989). Homologous recombination between the LTRs of a human retrovirus-like element causes a 5-kb deletion in two siblings. *Am. J. Hum. Genet.* 45, 848–854.

Mager, D. L., and Henthorn, P. S. (1984). Identification of a retrovirus-like repetitive element in human DNA. *Proc. Natl. Acad. Sci. USA* 81, 7510–7514.

Mariani-Costantini, R., Horn, T. M., and Callahan, R. (1989). Ancestry of a human endogenous retrovirus family. *J. Virol.* 63, 4982–4885.

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.

Medstrand, P., Lindeskog, M., and Blomberg, J. (1992). Expression of human endogenous retroviral sequences in peripheral blood mononuclear cells of healthy individuals. *J. Gen. Virol.* 73, 2463–2466.

Mizuno, T. (1987). Random cloning of bent DNA segments from Escherichia coli chromosome and primary characterization of their structures. *Nucleic Acids Res.* 15, 6827–6841.

Nelson, D. T., Goodchild, N. L., and Mager, D. L. (1996). Gain of Sp1 sites and loss of repressor sequences associated with a young, transcriptionally active subset of HERV-H endogenous long terminal repeats. *Virology* 220, 213–218.

Parvin, J. D., McCormick, R. J., Sharp, P. A., and Fisher, D. E. (1995). Prebending of a promoter sequence enhances affinity for the TATA-binding factor. *Nature* 373, 724–727.

Pearson, W. R., and Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* 85, 2444–2448.

Perl, A., Rosenblatt, J. D., Chen, I. S., 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.

Radic, M. Z., Lundgren, K., and Hamkalo, B. A. (1987). Curvature of mouse satellite DNA and condensation of heterochromatin. *Cell* 50, 1101–1108.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," 2 ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Shen, M. R., Batzer, M. A., and Deininger, P. L. (1991). Evolution of the master Alu gene(s). *J. Mol. Evol.* 33, 311–320.

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

Simpson, G. R., Patience, C., Löwer, R., Tönjes, R. R., Moore, H. D. M., Weiss, R. A., and Boyd, M. T. (1996). Endogenous D-Type (HERV-K) related sequences are packaged into retroviral particles in the placenta and possess open reading frames for reverse transcriptase. *Virology* 222, 451–456.

Sjøttem, E., Anderssen, S., and Johansen, T. (1996). The promoter activity of long terminal repeats of the HERV-H family of human retrovirus-like elements is critically dependent on Sp1 family proteins interacting with a GC/GT box located immediately 3′ to the TATA box. *J. Virol.* 70, 188–198.

Starr, D. B., Hoopes, B. C., and Hawley, D. K. (1995). DNA bending is an important component of site-specific recognition by the TATA binding protein. *J. Mol. Biol.* 250, 434–446.

Steinhuber, S., Brack, M., Hunsmann, G., Schwelberger, H., Dierich, M. P., and Vogetseder, W. (1995). Distribution of human endogenous retrovirus HERV-K genomes in humans and different primates. *Hum. Genet.* 96, 188–192.

Travis, J. (1992). Possible evolutionary role explored for "jumping genes". *Science* 257, 884–885.

Widegren, B., Kjellman, C., Aminoff, S., Sahlford, L. G., and Sjögren, H. O. (1996). The structure and phylogeny of a new family of human endogenous retroviruses. *J. Gen. Virol.* 77, 1631–1641.

Wilkinson, D. A., Freeman, J. D., Goodchild, N. L., Kelleher, C. A., and Mager, D. L. (1990). Autonomous expression of RTVL-H endogenous retroviruslike elements in human cells. *J. Virol.* 64, 2157–2167.

Wilkinson, D. A., Goodchild, N. L., Saxton, T. M., Wood, S., and Mager, D. L. (1993). Evidence for a functional subclass of the RTVL-H family of human endogenous retrovirus-like sequences. *J. Virol.* 67, 2981–2989.

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

Wu, H. M., and Crothers, D. M. (1984). The locus of sequence-directed and protein-induced DNA bending. *Nature* 308, 509–513.

Yeh, K. W., Yang, W. K., Huang, H. C., Feng, Y. N., Liu, J. C., Wu, F. Y., and Wu, C. W. (1995). Cloning and characterization of the endogenous retroviral-tRNA(Glu) multigene family from human genomes of different racial backgrounds. *Gene* 155, 247–252.

Zietkiewicz, E., Richer, C., Makalowski, W., Jurka, J., and Labuda, D. (1994). A young Alu subfamily amplified independently in human and African great apes lineages. *Nucleic Acids Res.* 22, 5608–5612.