GENOMICS **54,** 542–555 (1998)
ARTICLE NO. GE985608

A Long Terminal Repeat of the Human Endogenous Retrovirus ERV-9 Is Located in the 5′ Boundary Area of the Human β-Globin Locus Control Region

Qiaoming Long,*,[1] Chikh Bengra,*,[1] Chunhua Li,* Ferdane Kutlar,† and Dorothy Tuan*,[2]

*Department of Biochemistry and Molecular Biology and †Hemoglobin Laboratory, Department of Medicine, Medical College of Georgia, Augusta, Georgia 30912

Received July 31, 1998; accepted September 28, 1998

Transcription of the human β-like globin genes in erythroid cells is regulated by the far-upstream locus control region (LCR). In an attempt to define the 5′ border of the LCR, we have cloned and sequenced 5 kb of new upstream DNA. We found an LTR retrotransposon belonging to the ERV-9 family of human endogenous retroviruses in the apparent 5′ boundary area of the LCR. This ERV-9 LTR contains an unusual U3 enhancer region composed of 14 tandem repeats with recurrent GATA, CACCC, and CCAAT motifs. This LTR is conserved in human and gorilla, indicating its evolutionary stability in the genomes of the higher primates. In both recombinant constructs and the endogenous human genome, the LTR enhancer and promoter activate the transcription of *cis*-linked DNA preferentially in erythroid cells. Our findings suggest the possibility that this LTR retrotransposon may serve a relevant host function in regulating the transcription of the β-globin LCR.

INTRODUCTION

The human endogenous retroviruses (HERVs) were inserted into the germ cells of primates millions of years ago and have remained an integral part of the primate genomes during evolution. In addition to the proviruses, solo LTRs are dispersed throughout the human genome (Wilkinson *et al.*, 1994; Lower *et al.*, 1996). The solo LTRs contain the U3, R, and U5 regions (Temin, 1982) but no internal gag, pol, and env genes. Together, the HERVs and the solo LTRs constitute approximately 5% of the human genome and belong to the category of middle repetitive DNAs characterized as retrotransposons (Smit, 1996; Henikoff *et al.*, 1997). The ERV-9 proviruses, containing 30–50 members, constitute one of many families of the HERVs (Wilkinson *et al.*, 1994; Lower *et al.*, 1996). In addition to the proviruses, solo ERV-9 LTRs with a copy number of 3000–4000 have been found in the human genome (Henthorn *et al.*, 1986; La Mantia *et al.*, 1991; Schlessiger, 1992). The ERV-9 retrotransposons were inserted into the primate genome probably as early as 10 million years ago (Di Cristofano *et al.*, 1995b). The retrotransposons have been suggested to be selfish DNAs irrelevant to the cellular functions of the hosts (Dolittle and Sapienza, 1980). However, recent findings indicate that the enhancer and promoter elements in the U3 region of the LTRs (Lenz *et al.*, 1984; Speck *et al.*, 1990) initiate and promote the transcription of host genes located immediately downstream of the LTRs and may thus serve relevant cellular functions (Stravenhagen and Robins, 1988; Feuchter *et al.*, 1992; Goodchild *et al.*, 1992; Ting *et al.*, 1992; Schulte *et al.*, 1996).

The human β-like globin genes consist of the embryonic ε, the fetal Gγ and Aγ, and the adult δ and β genes located on chromosome 11 in the transcriptional order 5′ ε-Gγ-Aγ-δ-β 3′ (Efstratiadis *et al.*, 1980). The transcription of these genes is regulated by the far upstream locus control region (LCR), which is defined by four erythroid-specific, DNase I-hypersensitive sites, HS 1, 2, 3, and 4 (Tuan *et al.*, 1985; Forrester *et al.*, 1987; Grosveld *et al.*, 1987; Dhar *et al.*, 1990). The LCR between HS1 and HS4 is present in other mammals from mouse to galago and composes the major functional component of the LCR (reviewed by Hardison *et al.*, 1997). A ubiquitous HS5 site has been identified further upstream of the HS 1–4 sites (Tuan *et al.*, 1985; Dhar *et al.*, 1990) in the apparent 5′ boundary area of the LCR. In an attempt to define the 5′ border of the LCR and study its functions, we have cloned and sequenced 5 kb of DNA further upstream of the HS5 site that has not been previously characterized.

Our sequencing data for this region reveal a solitary ERV-9 LTR with the characteristics of a retrotransposon in a location near the HS5 site (see Fig. 1). This 5′ HS5 LTR possesses an unusual sequence feature in the U3 enhancer region that is composed of 14 tandem

[1] These authors contributed equally to this work.
[2] To whom correspondence should be addressed. Telephone: 706-721-0272. Fax: 706-721-6608. E-mail: dtuanlo@mail.mcg.edu.

0888-7543/98 $25.00

repeats of a consensus DNA of 41 bases. These U3 repeats as well as the downstream promoter contain recurrent GATA, CACCC, and CCAAT motifs. This LTR retrotransposon is conserved with 98–99% sequence identities in people of different races and in the gorilla, except that some people have 11 instead of 14 U3 repeats and the gorilla has only 5 U3 repeats. Functional tests with the CAT reporter gene assays demonstrate that the human 5′ HS5 LTR activates the *cis*-linked CAT gene and possesses enhancer and promoter activities in erythroid cells. In the CAT reporter gene assays, the LTR also synergized with and activated the *cis*-linked HS5 site. Consistent with these results, RT-PCR studies of cellular RNAs isolated from human primary cells and cell lines indicate that the endogenous LTR activates transcription of the downstream R, U5, and the genomic DNA at a higher level in erythroid cells than in nonerythroid cells. Taken together, these results suggest the possibility that the 5′ HS5 LTR may serve a relevant cellular function in regulating the transcriptional status of the β-globin LCR.

MATERIALS AND METHODS

Isolation of 5′ ϵ1.4 phage clone and DNA sequencing. The 5′ ϵ1.4 phage clone spanning 12 kb of DNA 5′ of the HS4 site was obtained from a K562 genomic DNA library constructed in EMBL phage (Weber-Benarous *et al.,* 1988). The library was screened with a unique DNA probe, 5′ ϵ1.4, located near the HS4 site in the LCR (Li *et al.,* 1985). The genomic DNA insert contained 8 kb of DNA spanning the HS5 site whose sequence was subsequently reported (Yu *et al.,* 1994) and 5 kb of further-upstream new DNA. The 8 kb of DNA was cleaved by *Hin*dIII into four subfragments of 2.7 kb spanning the HS5 site and 1.5, 1.6, and 2 kb spanning the new DNA. They were subcloned into a plasmid vector (Tuan *et al.,* 1990) and sequenced with the dideoxy terminator method (Sanger *et al.,* 1977) using a Sequenase or Taquenase kit (USB Corp.). This sequence strategy produced unambiguous DNA sequencing ladders for the entire 8 kb of DNA except for the 1 kb of DNA in the junction area between the 1.5- and 1.6-kb subclones, which contained the repetitive sequences of the ERV-9 LTR. The junction DNA was recloned into a phagemid vector, Bluescript II SK(±) (Stratagene), and the single-stranded DNA was sequenced as described above. The sequences were assembled and analyzed using the GCG DNA analysis software.

Purification of genomic DNAs from gorilla and from people of different races. Genomic DNAs were isolated anonymously from human blood samples collected by the Hemoglobin Laboratory at the Medical College of Georgia for diagnosis of thalassemia and sickle cell disease. African samples were from patients homozygous for sickle cell disease or hereditary persistence of fetal hemoglobin (HPFH). Arabic and Asian samples were from people hemizygous for α-thalassemia, and the Caucasian samples were from normal individuals or patients with β-thalassemia. The gorilla blood sample was obtained from the Yerkes Primate Center of Emory University. High-molecular-weight genomic DNAs were purified from nucleated blood cells (Poncz *et al.,* 1982).

PCR amplification of the 5′ HS5 LTR in genomic DNAs and sequence analysis of the amplified LTR. The 5′ HS5 LTRs were amplified from genomic DNAs with primer pair 3 used also for RT-PCR in Fig. 5 (forward primer, positions 595–616, and reverse primer, 1807–1831, Fig. 2a). PCR conditions consisted of an initial denaturation at 95°C for 4 min followed by 32 cycles of denaturation at 95°C for 1.5 min, annealing at 59°C for 1 min, and extension at 72°C for 2 min, and a final extension step at 72°C for 15 min. The amplified LTR fragments were purified with a Quantum Plasmid Miniprep Kit (Bio-Rad) and sequenced by the Molecular Biology Core Laboratory of the Medical College of Georgia using the cycle sequencing technique with fluorescent dideoxy terminators.

Construction of recombinant CAT plasmids. For LTR-CAT (Construct 1), the 1-kb LTR was amplified from K562 genomic DNA by PCR with forward primer 5′ TACTGTCGACCTGAGTTTGCTGGGGATG 3′ (positions 3250–3271 in the 8-kb GenBank Accession No. AF064190 corresponding to positions 595–616 in Fig. 2a) and reverse primer 5′ GATGGATCCTGTGTCCGGAATTGGTGG 3′ (positions 4332–4349 in GenBank Accession No. AF064190; positions 1677–1694 in Fig. 2a). A *Sal*I cloning site and a *Bam*HI cloning site (underlined) were added to the forward and reverse primers respectively. The PCR fragment was cleaved with *Sal*I and *Bam*HI enzymes and together with a *Bam*HI–*Hin*dIII adapter was spliced into a promoterless CAT vector derived from ϵp-CAT (Construct 3), in which the ϵ-globin promoter (ϵp) was removed with *Sal*I and *Hin*dIII digestions. Ups-CAT (Construct 2) contains a 1-kb PCR fragment amplified from the genomic DNA located 2 kb further upstream of the LTR and was created with the same cloning strategy. The respective forward and reverse primers were 5′ ACTGTCGACTTATGTATTCAAGTTGC 3′ (positions 50–66 in the GenBank sequence) and 5′ GATGGATCCAATAGATTTTTGTCATCT 3′ (positions 1203–1220 in the GenBank sequence). ϵp-CAT (Construct 3) and HS2-ϵp-CAT (Construct 4) were previously made (Tuan *et al.,* 1989). LTR-ϵp-CAT (Construct 5) was created with the above-mentioned 1-kb LTR DNA obtained by PCR, which was cleaved at the *Sal*I and *Bam*HI cloning sites and spliced into ϵp-CAT (Construct 3), which was also cleaved at the *Sal*I and *Bam*HI sites located 5′ of the ϵp. HS5-ϵp-CAT (Construct 6) was created with the same cloning strategy as for LTR-ϵp-CAT (Construct 5). The 1.2-kb HS5 fragment was generated by PCR from forward primer 5′ ACT-GTCGACAAGCTTCTGACAAATTATTCTT 3′ (positions 5481–5505, GenBank Accession No. AF064190) and reverse primer 5′ GATGGATCCACTGAAAGGGCTCATGCAAC 3′ (positions 6707–6726, GenBank Accession No. 064190). LTR-HS5-ϵp-CAT (Construct 7) was made from LTR-ϵp-CAT (Construct 5), which was linearized at the *Bam*HI site 3′ of the LTR. The above-mentioned 1.2-kb HS5 fragment obtained by PCR was cleaved at the 5′ end with *Hin*dIII (a natural site) and at the 3′ end with *Bam*HI and together with a *Bam*HI–*Hin*dIII adapter was spliced into the *Bam*HI site in LTR-ϵp-CAT.

Transient and stable transfections and CAT assays. Transfection host cells K562, HL60, and MEL cells were cultured and transfected as described (Tuan *et al.,* 1989) with modifications. In transient transfections, 10 μg of each of the CAT plasmids was mixed with 5 μg of a reference CMV β-gal plasmid and transfected into the host cells by electroporation. CAT assays were carried out as described (Tuan *et al.,* 1989) with two modified steps of normalizations. The CAT extracts were normalized first with respect to the total protein in the extract determined with the BCA (bicinchoninic acid) protein kit (Pierce) and then with respect to the β-galactosidase level of the cotransfected CMV β-gal plasmid to ensure that the CAT assays of different samples were carried out on extracts containing similar levels of β-gal activities and, therefore, similar amounts of the transfected tested plasmids. The β-gal enzyme levels were determined with the β-gal assay kit (Promega). The CAT enzymatic activities were analyzed by thin-layer chromatography and quantified with a PhosphorImager (Molecular Dynamics). The results were presented as percentages of conversion calculated from the ^{14}C counts in the acetylated chloramphenicol divided by the total input ^{14}C counts of the chloramphenicol substrate. In stable transfection, pooled cell populations were studied. The CAT activities were normalized with respect to the copy numbers of the integrated plasmids determined by Southern blots (not shown).

Isolation of total cellular RNAs and RT-PCR. Total cellular RNAs were purified from freshly harvested, nontransfected human erythroid K562, promyelocytic HL60, embryonic teratocarcinoma N-Tera (obtained from ATCC) and murine erythroleukemia MEL cell lines, adult human peripheral blood CFU-E and T-lymphocytes

(Wickrema *et al.*, 1992), and full-term human placenta. The RNAs were purified with the Totally RNA kit (Ambion). For a semiquantitative comparison of the RT-PCR bands generated by different primer pairs, each RNA was first reverse transcribed with random hexamers as primers into a cDNA master stock. Equal aliquots of the cDNA master stock were then added to separate tubes for PCR with different primer pairs including a β-actin reference primer pair. The β-actin primer pair amplified from the constitutively expressed β-actin mRNA presumably present at similar levels in different cell types to produce RT-PCR bands of similar intensities (Weiss *et al.*, 1994; Kong *et al.*, 1997), which served as the common quantitative references for comparing the relative intensities of the RT-PCR bands of the LTR and thus the relative levels of the LTR transcripts in different cell types. The 5′ → 3′ sequences of the respective forward and reverse primers are marked in Fig. 2a: primer pair 1, CTGAGTTTGCTGGGGATGCGAA (positions 595–616) and GATTTAGTGACTCATATTGTTTCTGA (positions 1700–1726); primer pair 2, TGCTGCTGCTCACTGTTTGGGTCTA (positions 1349–1373), with the reverse primer the same as that of primer pair 1. Primer pairs 3 and 4 contained the same forward primers as the respective forward primers of primer pairs 1 and 2. Primer pairs 3 and 4 contained a common reverse primer: 5′GGGCACTCTGCCTTAGGGAGTAACA 3′ (positions 1807–1831). The human β-actin primer pair was obtained from Stratagene. Before RT-PCR, the abilities of the primer pairs to produce amplification fragments were confirmed by PCR with genomic DNA templates.

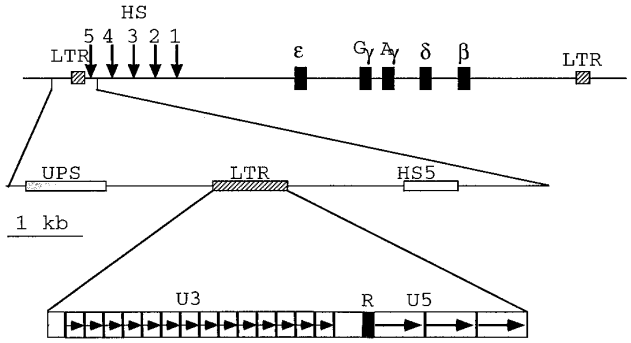


FIG. 1. Location of the ERV LTR in the boundary area of the β-globin LCR. (**Top**) The human β-like globin gene locus. Solid boxes: the embryonic ε-, fetal γ-, and adult δ- and β-globin genes. The vertical arrows: locations of the DNase I-hypersensitive sites HS 1, 2, 3, 4, and 5. The hatched box 5′ of the HS5 site: a solo ERV-9 LTR. The hatched box 3′ of the β-globin gene: a second copy of the ERV-9 LTR located 30 kb 3′ of the β-globin gene (Henthorn *et al.*, 1986; Anagnou *et al.*, 1995). (**Middle**) The enlarged 5′ boundary area drawn to scale according to the 1-kb scale bar. Open, hatched, and gray boxes, respective locations of the HS5 site, ERV-9 LTR, and an arbitrary upstream region (Ups), which was used as a control sequence for the LTR in reporter gene assays (Fig. 4). (**Bottom**) The structure of the LTR. Short horizontal arrows, the 14 short tandem repeats in the U3 region. Solid bar, the R region. Long horizontal arrows, the 3 longer repeats in the U5 region.

RESULTS

An LTR Retrotransposon of the ERV-9 Family of Human Endogenous Retroviruses Is Located Proximal to the HS5 Site in the 5′ Boundary Area of the LCR

To study the sequence and function of DNA in the boundary area of the LCR, we screened a K562 DNA library (Weber-Benarous *et al.*, 1988) and obtained a clone containing 8 kb of DNA sequence that spans the HS5 site and 5 kb of new further upstream DNA. Because the sequence features of the upstream DNA were previously unknown, we sequenced the 5-kb new DNA as well as the 3-kb DNA spanning the HS5 site (GenBank Accession No. AF064190). The DNA sequence of the 3-kb DNA spanning the HS5 site was in general agreement with the DNA sequence of this region reported earlier (Yu *et al.*, 1994), except for a number of polymorphic base differences. In the new DNA, sequence matches using the GCG and BLAST programs revealed the existence of a solitary LTR at a location within 2 kb 5′ of the HS5 site (Long *et al.*, 1995) (Fig. 1). Comparison with a few selected homologous sequences in the GenBank database, including the LTR sequence located 5′ of the ZNF80 protein gene (Di Cristofano *et al.*, 1995a, GenBank Accession No. X83497), showed that the 5′ HS5 LTR spans 1.7 kb of DNA (Fig. 2a) and belongs to the ERV-9 family of human endogenous retroviruses (La Mantia *et al.*, 1991; Lania *et al.*, 1992).

Consistent with a common property of the retrotransposons, the 5′ HS5 LTR is flanked by 4 bases of direct repeats GTAT in the genomic DNA immediately 5′ and 3′ of the LTR sequence (Fig. 2a). This indicates that the 5′ HS5 LTR was inserted into the human ancestral genome at the GTAT site some time during evolution. In line with the general LTR structure of mammalian retroviruses (Temin, 1982), the 5′ HS5 LTR contains the U3, R, and U5 regions and is bracketed by the dinucleotides TG and CA, respectively, at the 5′ and 3′ ends (Fig. 2a). The U3 region contains the viral enhancer spanning tandemly repeated DNA sequences and the viral promoter (Lenz *et al.*, 1984; Golemis *et al.*, 1990; La Mantia *et al.*, 1991; Anagnou *et al.*, 1995). The R region starts with the viral transcription initiation site (La Mantia *et al.*, 1992) and is followed by the U5 region (Fig. 1). In the U3 region, the 600 DNA bases preceding the U3 repeats are composed of 70% G and C bases. This GC-rich region is found in many of the homologous ERV-9 LTRs in the database but is not present in the LTR of the ERV-9 provirus (La Mantia *et al.*, 1991). The U3 enhancer repeats and the promoter in the 5′ HS5 LTR show 80–90% base identities with other ERV-9 LTRs found in the human genome (Yang *et al.*, 1983; La Mantia *et al.*, 1991; Lania *et al.*, 1992; Di Cristofano *et al.*, 1995b).

It is of interest to note that in addition to the 5′ HS5 LTR located approximately 25 kb 5′ of the ε-globin gene, another ERV-9 LTR is located at a position approximately 25 kb 3′ to the β-globin gene (Fig. 1). The repetitive DNA in the region 3′ of the β-globin gene was first reported by Henthorn *et al.* (1986) and subsequently studied by Anagnou *et al.* (1995). Although neither of those groups recognized that the repetitive DNA was part of an endogenous LTR, sequence matches carried out by us as shown above revealed that the repetitive DNA of this region shares 80–90% sequence identities with the U3, R, and U5 regions of the 5′ HS5 LTR. Thus, two copies of the ERV-9 LTRs exist in flanking positions of the β-globin gene cluster.

Sequence Analysis of the U3 Enhancer Region in the 5′ HS5 ERV-9 LTR

The U3 enhancer region of the 5′ HS5 LTR shows an interesting sequence structure. It is composed of 14 tandem repeats of a consensus DNA sequence of 37–41 bases (Fig. 2a). Sequence matches show that the tandem repeats are composed of four subtypes, 1, 2, 3, and 4, which are arranged in the LTR in the order 1-2-3-4-1-2-3-4-1-2-3-4-4-1 (Fig. 2b). Among the four subtypes, the sequence identities are 60–80%, using subtype 2 as the reference. Among the U3 repeats of each subtype, the sequence identities are 80–98% (Fig. 2b). The consensus sequence of the 14 U3 repeats (Fig. 2b) reveals recurrent sequence motifs that can potentially bind to the GATA (Ko and Engel, 1993; Merika and Orkin, 1993), CCAAT (Johnson and McKnight, 1989), and CACCC (Miller and Bieker, 1993; Crossley *et al.*, 1996) transcription factors. Altogether, the U3 enhancer region contains within 600 bases of DNA eight GATA, nine CCAAT, three CACCC, and four CCACC sites. The consensus sequence of the 14 U3 repeats shows higher than 90% sequence identity with that of the 7 U3 repeats in the 3′ β LTR (Henthorn *et al.*, 1986) and that of the 6 U3 repeats in LTR2, a random clone of the ERV-9 LTR (Lania *et al.*, 1992) (Fig. 2b).

Sequence Analysis of the U3 Promoter Region

The promoter sequence in the LTR is located in the U3 region at the 3′ end of the 14 U3 repeats. It is located immediately upstream of the R region, whose 5′ border marks the transcriptional initiation site for retroviral RNA synthesis (Temin, 1982) (Fig. 2b). The promoter of the 5′ HS5 LTR shows a sequence homology of 80% with the promoter of the 3′ β LTR and of over 90% with the promoter of LTR2 (Fig. 2c). The transcriptional initiation site of LTR2 has been determined by primer extension to be located 28 bases downstream of the AATAAAA box (La Mantia *et al.*, 1992; Strazzullo *et al.*, 1994). Because of extensive sequence identities between the 5′ HS5 LTR and the LTR2 promoters, especially the 100% sequence homology in the 70 DNA bases flanking the AATAAAA box, the presumptive transcriptional initiation site of the 5′ HS5 LTR was placed at the identical T base 28 bases downstream of the AATAAAA box (Fig. 2c). All three LTR promoters contain the GATA, CACCC, and CCAAT motifs located at identical sites, −36, −46, and −63 bases, respectively, relative to the retroviral transcriptional initiation site (Fig. 2c).

The 5′ HS5 LTR promoter also shares similarities with the promoters of the further downstream ε-, γ-, and β-globin genes (Baralle *et al.*, 1980; Shen *et al.*, 1981; Poncz *et al.*, 1983; Li *et al.*, 1985) in that a combination of similar GATA, CACCC, and CCAAT motifs is found also upstream of the AATAAAA boxes in the globin promoters (Nienhuis *et al.*, 1984). In particular, the LTR promoter and the ε-globin promoter share additional sequence identities in the region immediately 5′ of the transcriptional initiation site (Fig. 2c). The above sequence and structural homologies suggest that, like the globin promoters, the 5′ HS5 LTR promoter would be active in erythroid cells.

The 5′ HS5 ERV-9 LTR Is Conserved in the Genomes of Gorilla and of People of Different Racial Lineages

Because the 5′ HS5 LTR is apparently a retrotransposon and is located not near but far upstream of the β-like globin genes, it is possible that the 5′ HS5 LTR might have resulted from a recent, random insertional event in the K562 genome during cell culture. Should this be the case, the 5′ HS5 LTR would not be present in the genome of the gorilla, which diverged from the human genome approximately 10 million years ago (Sibley and Ahlquist, 1987), nor in the genomes of people of different racial lineages, which diverged approximately 100,000 years ago (Vogel and Motulsky, 1986). To examine these possibilities, we used PCR to detect the presence or absence of the 5′ HS5 LTR in the genomic DNAs isolated from the blood samples of the gorilla and of people of different races. The PCR primers were synthesized according to the K562 DNA sequence, which amplified 1.2 kb of 5′ HS5 LTR, including 130 bases of genomic DNA downstream of the LTR (see Materials and Methods and Fig. 2a).

The PCR results indicate that the 5′ HS5 ERV-9 LTR is conserved in the genomes of the gorilla and people across racial lines. Fifteen of a total of 17 human DNAs isolated from Africans, Arabs, Asians, and Caucasians and from human cell lines K562 and HL60 produced amplicons of the anticipated length of 1.2 kb. However, 2 of the 9 African DNAs produced either a shorter amplicon of 1.1 kb or a longer 1.4-kb and a shorter 1.1-kb amplicon, while the gorilla DNA produced an even shorter amplicon of 0.9 kb (Fig. 3a).

It is possible that the observed amplicons might be spurious PCR products amplified by the primer pair from other ERV-9 LTRs in the human or the gorilla genome, because the 5′ primer was located within the U3 region immediately upstream of the enhancer repeats—a region present also in some of the other ERV-9 LTRs even though the 3′ primer was located in the unique genomic DNA region (see Fig. 2a). Therefore, the authenticity of the amplicons was further confirmed by DNA sequencing. We sequenced four standard amplicons of 1.2 kb from two Caucasian and two African DNAs, two shorter amplicons of 1.1 kb from the African DNAs (GenBank Accession No. AF064191), and the 0.9-kb amplicon of the gorilla DNA (Fig. 3b). The electropherograms of the DNA sequences showed sharp DNA sequence ladders with only a couple of ambiguities where two different bases occupied the same sequence positions (not shown), indicating that the two homologous chromosomes contained base polymorphisms at these positions. All the sequenced amplicons showed base identities of 98–99% in both

the LTR and the 3′ flanking genomic DNA; the only exception was the smaller number of U3 repeats in some people and in the gorilla (Figs. 3a and 3b). If the sequenced amplicons contained amplification products generated also from other homologous ERV-9 LTRs, the electropherograms would have contained too many sequence ambiguities to generate clearly readable sequences. The above observations indicate that the amplicons were genuine products of the 5′ HS5 LTR in the human and gorilla genomes.

In both the shorter human amplicons containing 11 U3 repeats from two different individuals, the deletion of 3 complete U3 repeats was generated apparently by the same in-phase deletion event so the subtype organizations of both amplicons were identical, 1-2-3-4-1-2-3-4-1-2-1 (Figs. 3a and 3b). This finding again confirms that the amplicons were not spurious PCR products. In the gorilla amplicon with 5 U3 repeats, the subtype organization is 1-2-3-4-1 (Figs. 3a and 3b). The apparent genomic insertion site of the LTR—the GTAT sequence—is conserved in both the human and the gorilla amplicons (Fig. 3b).

The sequence identities in the 5′ HS5 LTR between human and gorilla and among people of different races indicate that this LTR was probably inserted into the 5′ boundary area of the β-globin LCR at least 10 million years ago before the divergence of the human and apes, and it has been conserved in the genomes of the higher primates during the ensuing years of evolution. These observations suggest the possibility that this 5′ HS5 LTR retrotransposon may be conserved for the preservation of a relevant cellular function of the host.

The 5′ HS5 ERV-9 LTR Possesses Enhancer and Promoter Activities in Erythroid Cells

To investigate whether the enhancer and promoter regions in the 5′ HS5 LTR possess enhancer and promoter activities, seven recombinant CAT plasmids were made (Fig. 4a). LTR-CAT (Construct 1) contained the 1-kb LTR spanning the 14 U3 enhancer repeats, U3 promoter, R, and U5 spliced 5′ of the CAT gene in the absence of a promoter in the vector. To determine whether a randomly chosen region of the 5′ boundary area of the LCR also possessed enhancer and promoter activities, the control Ups-CAT plasmid (Construct 2) contained a 1-kb DNA (Ups) located 1.5 kb 5′ of the LTR (Fig. 1). The HS2-ϵp-CAT plasmid (Construct 4) that contained the strong HS2 enhancer of the LCR (Tuan *et al.*, 1989) coupled to the ϵ-globin promoter served as the standard with which to compare the

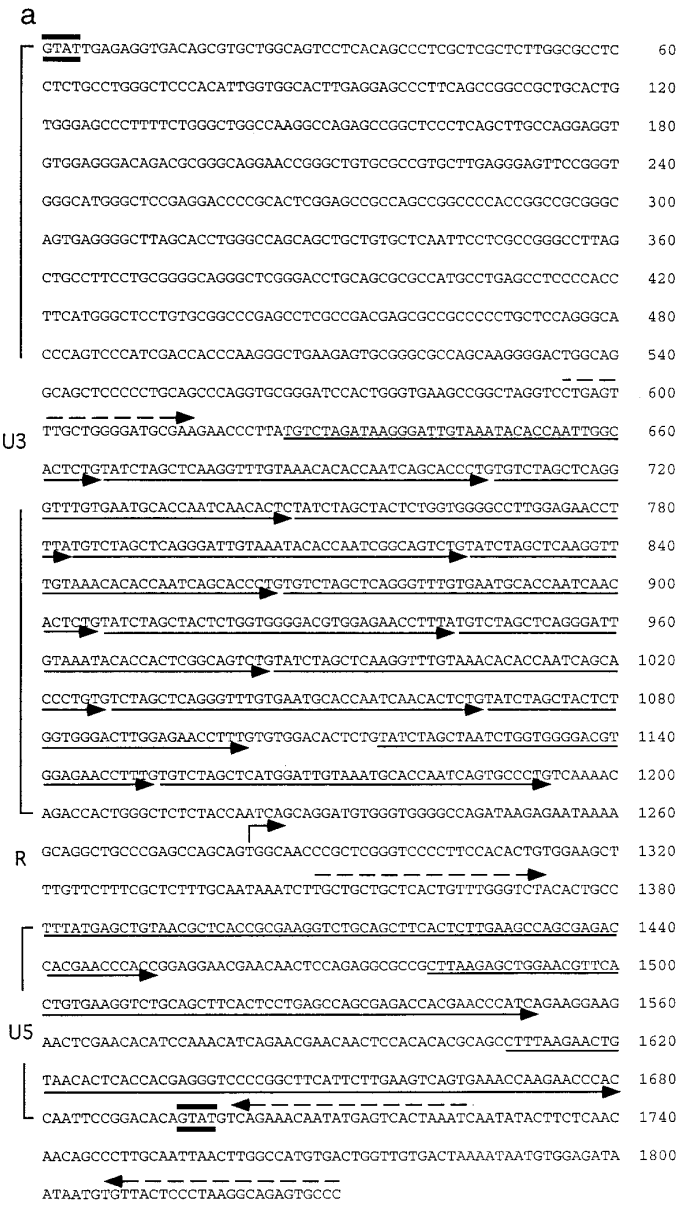
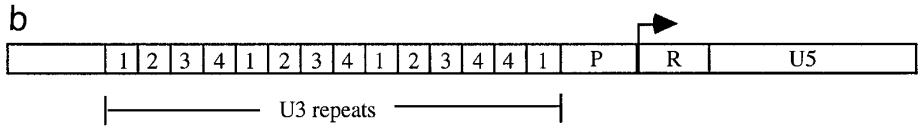


FIG. 2. Sequence analysis of the 5′ HS5 LTR in K562 cells. (**a**) DNA sequence of the 5′ HS5 LTR in the 5′ ϵ1.4 phage DNA clone from K562 cells. The four bases GTAT with the heavy overline and underline located at the 5′ and 3′ ends of the LTR, the presumed integration site of the LTR in the human genomic DNA. The horizontal arrows in U3, the 14 tandem repeats of 37–41 bases in the U3 region. Angled arrow, the presumed transcriptional initiation site in the LTR, marking the beginning of the R region. The long horizontal arrows in the U5 region, the 3 repeats of 70 bases in U5. Arrowheads connected to dotted overlines, locations of the PCR primers used in DNA PCR and RT-PCR studies in Figs. 3 and 5; directions of the arrowheads, the 5′ to 3′ direction of the primers. (**b**) Comparison of the sequences of the U3 repeats. (**Top**) The organization of the four subtype U3 repeats 1, 2, 3, and 4 in 5′ HS5 LTR. P, The promoter in the U3 region. (**Middle**) The sequences of the second set of subtype repeats 1–2–3–4 in (**a**) Underlined bases, the GATA, CCAAT, CACCC, or CCACC motifs. (**Bottom**) Consensus sequences of the U3 repeats in different ERV-9 LTRs. 5′ HS5, 3′ β, and LTR2, the 5′ HS5 LTR, the LTR at 25 kb 3′ of the β-globin gene (Henthorn *et al.*, 1986; Anagnou *et al.*, 1995), and the LTR in a random human DNA clone (Lania *et al.*, 1992), respectively. Lowercase letters separated by slashes, polymorphic bases in the U3 repeats. (**c**) Sequence comparison of three U3 promoters and the ϵ-globin promoter. (**Top**) 5′ HS5, the U3 promoter of the 5′ HS5 LTR. The overlined bases, the equivalent of the TATA box (Strazzullo *et al.*, 1994). Underlined bases, the DNA motifs found also in the U3 repeats. Angled arrow, the transcriptional initiation site in LTR2 (La Mantia *et al.*, 1992; Strazzullo *et al.*, 1994) and the presumed transcriptional initiation site in the 5′ HS5 LTR. (**Bottom**) Sequence alignment of the four promoters in the 5′ HS5 LTR, 3′ β LTR, and LTR2, respectively. Dashes, DNA base deletions.



U3 repeat subtypes:

1. TGTCTAGCTC AGGGATTGTA AATACACCAA TCGGCAGTCT G
2. TATCTAGCTC AAGGTTTGTA AACACACCAA TCAGCACCCT G
3. TGTCTAGCTC AGGGTTTGTG AATGCACCAA TCAACACTCT G
4. TATCTAGCTA CTCTGGTGGG GACGTGGAGA ACCTTTA

Consensus sequences of U3 Repeats:

5'HS5 Tg/aTCTAGCT C Ag/t/aGG T TTGTa/gAAc/t a/gCACCAATCAGCACTCTG

3'β T G TCTAGCTa/cA A GG T TTGT A AA T G CACCAATCAGCACTCTG

LTR2 Tg/aTCTAGCTc/aA g/a GGa/tTTGT A AA C a/gCACCAATCAGCACTCTG

c

U3 and ε-globin Promoter sequences:

5'HS5 TCAAAACAGA CCACTGGGCT CTCTACCAAT CAGCAGGATG TGGGTGGGGC

CAGATAAGAG AATAAAAGCA GGCTGCCCGA GCCAGCAGTG GCAA

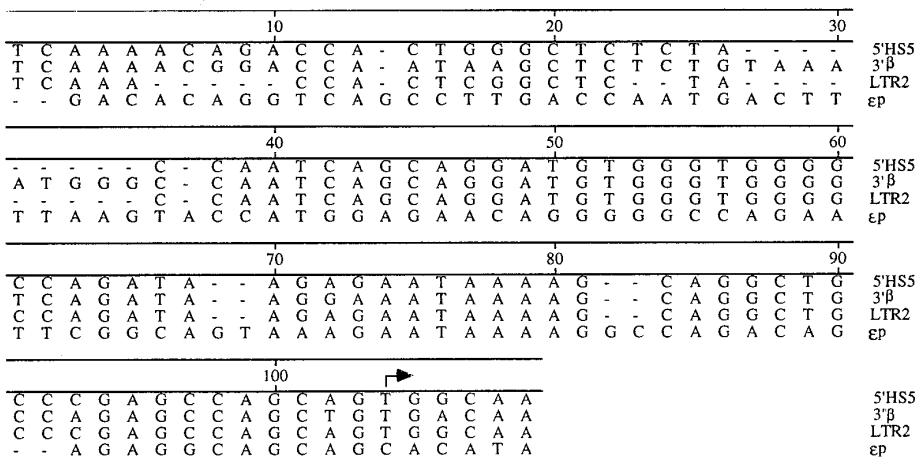
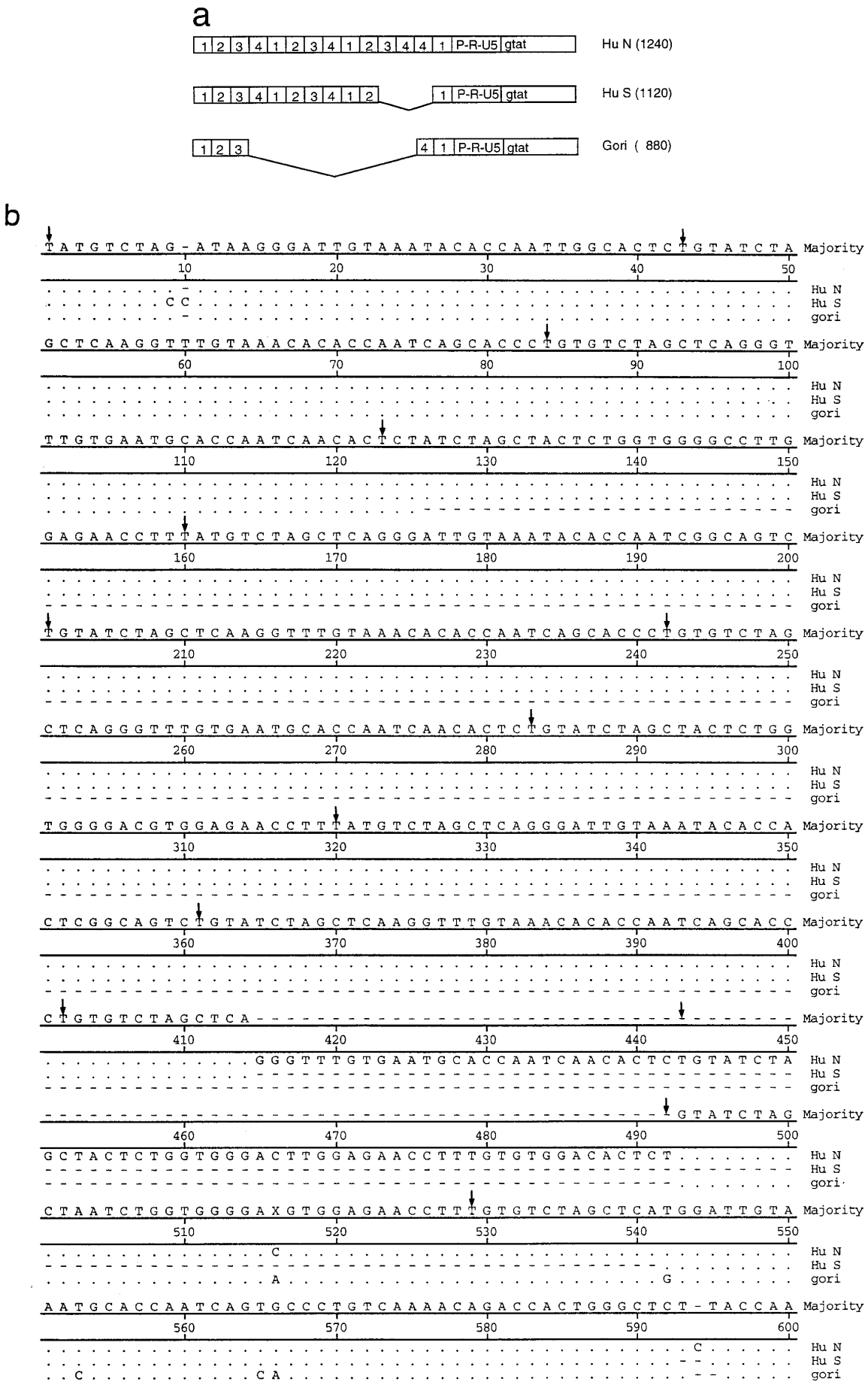


FIG. 2—*Continued*

enhancer and promoter activities of the 5′ HS5 LTR. To test whether the enhancer in 5′ HS5 LTR can synergize with and activate the HS5 site located naturally downstream of and proximal to the LTR, LTR-εp-CAT, HS5-εp-CAT, and LTR-HS5-εp-CAT (Constructs 5, 6, and 7, Fig. 4a) contained the LTR and HS5 site spliced either separately or together into εp-CAT (Construct 3). The plasmids were transiently transfected into erythroid K562 and MEL cells and nonerythroid HL60 cells and stably integrated into K562 cells.

Transient transfection results indicate that in human erythroid K562 cells, the LTR in the LTR-CAT plasmid displayed enhancer and promoter activities that were approximately 50% of the combination of the HS2 enhancer and the ε-globin promoter in the HS2-εp-CAT plasmid. In contrast, in murine erythroid MEL cells and human nonerythroid HL60 cells, both LTR-CAT and HS2-εp-CAT displayed much lower enhancer and promoter activities (Fig. 4b). The low enhancer activity of the HS2 enhancer in MEL cells was due apparently to the inactivity of the *cis*-linked embryonic ε-globin promoter in MEL cells expressing the adult globin program (not shown); when linked to the more permissive adult β-globin promoter, the HS2 enhancer displayed much higher enhancer activity in MEL cells (Cavallesco and Tuan, 1997). Likewise, the U3 enhancer in the LTR may also be potentially active in MEL cells; its apparently low enhancer activity may be


a
1 2 3 4 1 2 3 4 1 2 3 4 4 1 P-R-U5 gtat
Hu N (1240)
1 2 3 4 1 2 3 4 1 2
1 P-R-U5 gtat
Hu S (1120)
1 2 3
4 1 P-R-U5 gtat
Gori (880)
b
Majority
Hu N
Hu S
gori

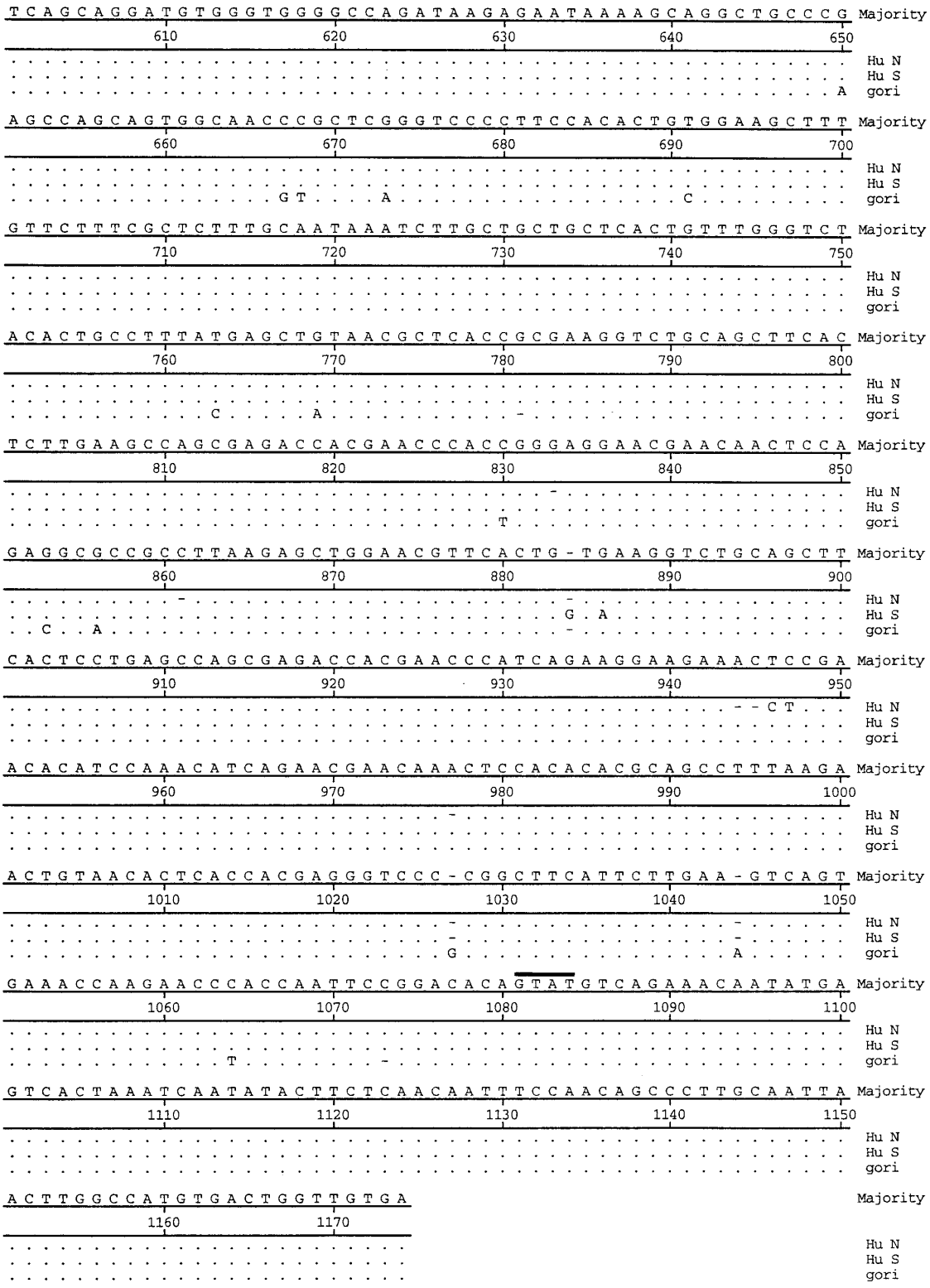


FIG. 3. The 5′ HS5 LTR is conserved in the genomes of human and gorilla and in people of different racial lineages. (**a**) The 5′ HS5 LTRs amplified by primer pair 3 (Figs. 2a and 5a) from the genomic DNAs of human and gorilla. Hu N, human LTR of the normal length with 14 U3 repeats; Hu S, human LTR of a shorter length with 11 U3 repeats; Gori, gorilla LTR with 5 U3 repeats. Numbers in parentheses, total number of bases in the amplified LTRs, including 140 bases of genomic DNAs downstream of the LTR insertion site—the gtat bases—that were amplified by the PCR primers. Bent lines in Hu S and Gori, deletions of three and nine complete U3 repeats in the truncated human and gorilla LTRs, respectively. (**b**) Sequence alignment of the normal human, truncated human, and gorilla LTRs. Majority bases, the consensus DNA sequence among the three LTRs. Numbers between two horizontal lines, the DNA base ruler with base 1 being the first base of the first U3 repeat in the LTRs. Vertical arrows, the positions of the first base in the U3 repeats. Dots, the same bases in the human or gorilla DNAs as those in the consensus sequence. Dashes, base deletions. The GTAT bases at positions 1081–1084 marked with heavy overline, the integration site of the 5′ HS5 LTR in both human and gorilla DNAs.

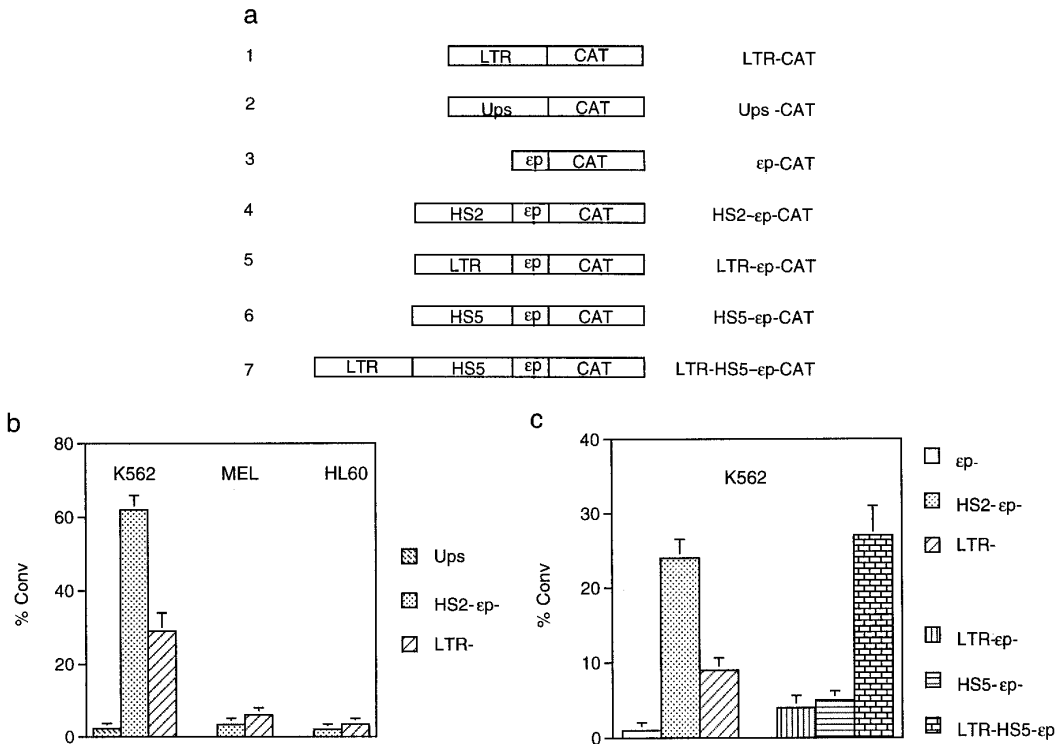


FIG. 4. Enhancer and promoter activities of the 5′ HS5 LTR in recombinant CAT constructs. (**a**) Recombinant CAT constructs. LTR, a 1-kb LTR sequence; Ups, 1.2 kb of DNA upstream of the LTR (see Fig. 1); εp, a 200-bp ϵ-globin promoter; HS2, a 0.74-kb HS2 enhancer; HS5, a 1.2-kb sequence spanning the HS5 site. (**b**) CAT activities of Ups-CAT, HS2-εp-CAT, and LTR-CAT plasmids transiently transfected into K562, MEL, and HL60 cells. % Conv, percentage conversion of the [^{14}C]chloramphenicol substrate by the CAT enzyme produced by the transfected test plasmid after normalization with respect to a common level of a cotransfected CMV-β-gal plasmid. (**c**) CAT activities of the εp-CAT, HS2-εp-CAT, LTR-CAT, LTR-εp-CAT, HS5-εp-CAT, and LTR-HS5-εp-CAT plasmids integrated into the genome of K562 cells. % Conv, percentage conversion of the [^{14}C]chloramphenicol substrate by the CAT enzyme produced by the integrated plasmids after normalization with respect to the per-cell copy numbers of the plasmids.

due to the low activity in MEL cells of the U3 promoter which shares certain sequence identities with the ϵ-globin promoter (Fig. 2c). The developmental stage specificity of the U3 enhancer and promoter is currently being investigated.

When stably integrated into the genome of K562 cells, the LTR displayed enhancer and promoter activities that were approximately 30% of those of the HS2-εp-CAT plasmid (Fig. 4c). However, in integrated LTR-HS5-εp-CAT plasmid, the LTR enhancer synergized with the HS5 site and activated the CAT gene to a level comparable to that displayed by the HS2 enhancer in HS2-εp-CAT (Fig. 4c). These results indicate that the 5′ HS5 LTR possesses enhancer and promoter activities in erythroid cells, and it synergized with and activated the HS5 site.

The Endogenous 5′ HS5 LTR Activates the Transcription of Downstream DNA Preferentially in Erythroid Cells

We next determined whether the endogenous 5′ HS5 LTR also exhibits enhancer and promoter activities and can activate the transcription of the downstream R region and the flanking genomic DNA in the β-globin LCR. The transcriptional statuses of the 5′ HS5 LTR and downstream genomic DNA were determined by RT-PCR in erythroid K562 and nonerythroid HL60 and N-Tera cell lines and in nontransformed, primary erythroid cells CFU-E and nonerythroid T-lymphocytes and placental cells.

Four PCR primer pairs were made (Fig. 5a). Primer pair 1 determined whether the entire LTR between the U3 enhancer and the U5 regions as well as the genomic DNA immediately downstream of it was transcribed. Primer pair 2 detected retroviral mRNA transcripts of the R and U5 regions that were activated by the U3 enhancer and promoter. To ensure that primer pair 2 detected the RNA transcribed specifically from the 5′ HS5 LTR and not RNAs transcribed from other ERV-9 LTRs, the forward primer was located in the R region that contains a number of polymorphic bases among the ERV-9 LTRs (Fig. 2a; Henthorn *et al.*, 1986; Lania *et al.*, 1992), and the reverse primer was located in the genomic DNA immediately downstream of the LTR. Primer pairs 3 and 4 were synthesized to confirm that the RNAs detected by primer pairs 1 and 2 were indeed transcribed from the 5′ HS5 ERV-9 LTR. These two primer pairs contained the same respective forward primers as primer pairs 1 and 2 but shared a common reverse primer located in the genomic DNA 110 bases further downstream of the reverse primer of primer pairs 1 and 2. Hence, the authentic RT-PCR bands of

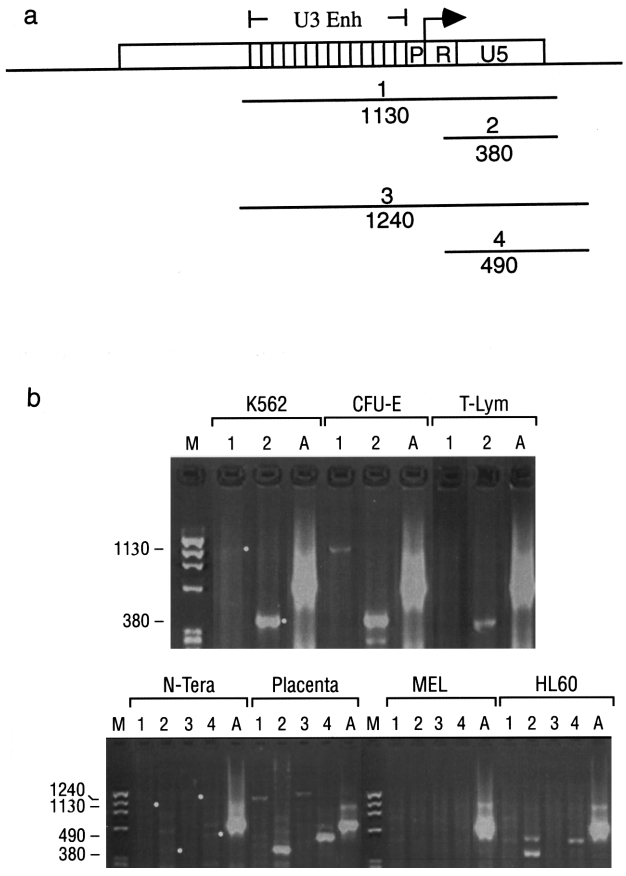


FIG. 5. Transcription of the endogenous 5′ HS5 LTR in the genomes of human erythroid and nonerythroid cells as determined by RT-PCR. (**a**) (**Top**) The 5′ HS5 LTR in normal human DNA with 14 U3 enhancer repeats. (**Bottom**) The four horizontal lines 1, 2, 3, and 4, anticipated RT-PCR fragments amplified, respectively, by primer pairs 1–4, synthesized according to the K562 sequence in Fig. 2a. Numbers below the lines, anticipated sizes in basepairs of the amplified cDNA fragments. (**b**) RT-PCR bands generated by RNAs isolated from erythroid cells K562, CFU-E, and MEL cells and nonerythroid cells N-Tera, HL60, T-lymphocytes, and placenta. Lanes 1–4 contain RT-PCR bands generated by the corresponding primer pairs 1–4 as shown in a. White dots on the right of the marked lanes, electrophoretic positions of the anticipated RT–PCR fragments. Lanes **A,** RT–PCR bands generated by the β-actin primer pair, a common quantitative reference. Each lane contained 15 μl of the RT-PCR product except lanes **A** in the lower panel, which contained 3 μl of the RT-PCR products. **M,** size marker, *Hae*III cleaved ϕx DNA. Numbers in the left margin, sizes of the RT-PCR bands in basepairs.

the 5′ HS5 LTR generated by primer pairs 3 and 4 were 110 bases longer than those generated, respectively, by primer pairs 1 and 2 (Fig. 5a).

Consistent with the design of the primer pairs (Fig. 5a), the sizes of the RT-PCR bands produced by primer pairs 3 and 4 were indeed longer by 110 bases than those produced by primer pairs 1 and 2 (Fig. 5b). This indicates that the RT-PCR bands generated by these four primer pairs were genuine products amplified from the 5′ HS5 LTR and not from other ERV-9 LTRs in the human genome. In addition, the authenticity of the PCR band produced by primer pair 3 had been confirmed by direct DNA sequencing (Fig. 3).

For a semiquantative comparison of the intensities of RT-PCR bands generated by primer pairs 1–4 in different RNA samples, a β-actin primer pair spanning a region in the constitutively expressed β-actin mRNA was included in the RT-PCRs. Indeed, the intensities of the β-actin band generated by the same amount of RNAs from different cell types were similar (Fig. 5b, Lanes A). The relative intensities of the LTR bands with respect to the intensity of the β-actin band generated from equal aliquots of the same cDNA master stock were then compared.

The RT-PCR results indicate that the endogenous 5′ HS5 LTR promoted the transcription of the R and U5 regions. In both erythroid and nonerythroid cells, primer pairs 2 and 4 generated amplification bands of the R and U5 regions (Fig. 5b, Lanes 2 and 4). However, the LTR enhancer and promoter appear to be more active in erythroid cells than in nonerythroid cells because the amplification bands generated from RNAs of K562 cells and CFU-E were relatively stronger than those of nonerythroid T-lymphocytes, N-Tera, and HL60 cells. An apparent exception to the above observation was the nonerythroid placenta, which also generated strong LTR bands (Fig. 5b). This may be due to contamination of abundant maternal and fetal blood erythroid cells, in which the 5′ HS5 LTR enhancer and promoter were active. On the other hand, the 5′ HS5 LTR enhancer and promoter may also be active in the placenta, because many HERVs and their solitary LTRs have been found to be capable of initiating viral RNA synthesis from the R region in placental cells (Wilkinson *et al.*, 1994; Lower *et al.*, 1996).

Further upstream of the R region in the LTR, no additional transcriptional initiation sites appear to exist in the majority of the cell types tested, since primer pairs 1 and 3 did not generate detectable bands from RNAs of erythroid K562 and nonerythroid T-lymphocytes, N-Tera, and HL60 cells (Fig. 5b, Lanes 1 and 3). However, primer pairs 1 and 3 generated faint amplification bands from erythroid CFU-E and nonerythroid placenta RNAs. This suggests that CFU-E and placenta may contain additional transcriptional initiation sites proximal to the 5′ HS5 LTR.

The above RT-PCR results indicate that the endogenous 5′ HS5 LTR possesses apparent enhancer and promoter activities and is capable of promoting the transcription of the R and U5 regions in the LTR and of further downstream genomic DNA in the LCR.

DISCUSSION

In this report, we show that a solitary ERV-9 LTR with the characteristics of a retrotransposon is located proximal to the HS5 site in the apparent 5′ boundary area of the β-globin LCR. This 5′ HS5 ERV-9 LTR possesses unusual sequence features in the multiple tandem repeats of the U3 enhancer region. The U3 repeats and the immediately downstream U3 promoter contain within 700 DNA bases 9 GATA, 4 CACCC, and 10 CCAAT sites. These DNA motifs can bind, respec-

tively, to the cognate GATA (Orkin, 1992) and CACCC (Miller and Bieker, 1993; Crossley *et al.*, 1996) transcription factors expressed abundantly in erythroid cells and to the CCAAT factors C/EBP (Johnson and McKnight, 1989) and NF-Y (Bi *et al.*, 1997) expressed in many hematopoietic and nonhematopoietic cells. The high concentration of these motifs in the U3 region suggests that the 5′ HS5 ERV-9 LTR, and probably also other homologous ERV-9 LTRs dispersed throughout the human genome, may be preferentially active in erythroid cells.

The 5′ HS5 LTR is conserved in the gorilla and in people of different racial lineages, indicating that this LTR was probably inserted into its location at the 5′ boundary area of the LCR before species divergence between human and gorilla approximately 10 million years ago. The conservation of the 5′ HS5 LTR during evolution of the higher primates suggests that this LTR retrotransposon is not detrimental to the host and may be conserved to serve a host function.

Functional tests with the CAT reporter gene assays show that the 5′ HS5 LTR, in line with its component sequence motifs, possesses enhancer and promoter activities preferentially in erythroid cells. Moreover, the LTR enhancer activity can synergize with and activate the *cis*-linked HS5 site in the LCR.

It is of interest to note that, like the domain boundary elements of the chicken lysozyme gene (Steif *et al.*, 1989) and the *Drosophila* heat-shock gene locus (Kellum and Schedl, 1991), the HS5 site has been reported to be associated with the nuclear matrix (Jarman and Higgs, 1988) and to possess insulator activity. When DNA fragments containing the HS5 site were spliced in a position between a reporter gene and the viral polyoma enhancer (Yu *et al.*, 1994), the HS2 enhancer (Chung *et al.*, 1993), or the HS3 enhancer (Li and Stamatoyannopoulos, 1995), the HS5 site was reported to exhibit insulator activity and to block these enhancers from activating the transcription of the *cis*-linked gene. In contrast, we found that when the HS5 site was spliced in a LTR-HS5-ϵp-CAT construct between the CAT gene and the LTR enhancer in its natural genomic position downstream of the LTR, the HS5 site did not exhibit insulator property, as it did not block the LTR enhancer activity but synergized with and activated the LTR enhancer. These observations suggest that whether or not the HS5 site exhibits insulator property depends on the nature of the *cis*-linked enhancer in the reporter gene constructs.

Consistent with our finding that the 5′ HS5 LTR possesses enhancer activity, other ERV-9 LTRs cloned from random loci in the human genome, including LTR2 (see Figs. 2b and 2c), have been reported to possess enhancer and promoter activities in CAT reporter gene assays (La Mantia *et al.*, 1991; Lania *et al.*, 1992). In addition, the 3′ β LTR located far downstream of the human β-globin gene (see Figs. 1, 2b, and 2c), when linked to either the CAT or the human γ-globin gene, exhibits enhancer activity in erythroid cell lines and in the fetal liver cells of transgenic mice (Anagnou *et al.*, 1995). It has been suggested that the 3′ β LTR enhancer juxtaposed to the γ-globin genes as a result of the large DNA deletions in hereditary persistence of fetal hemoglobin, HPFH-3 and -4, may promote the expression of the fetal γ-globin genes in erythroid cells (Henthrorn *et al.*, 1986; Anagnou *et al.*, 1995).

In agreement with the erythroid preference of the enhancer and promoter activities of the 5′ HS5 LTR in reporter gene assays, the endogenous 5′ HS5 LTR activates the transcription of the downstream R and U5 regions and flanking genomic DNA in the LCR at higher levels in erythroid cells than in nonerythroid cells. Taken together, our results suggest that this 5′ HS5 LTR-retrotransposon may regulate the transcription of the proximal LCR DNA.

The ERV-9 LTRs of other genomic loci have been reported to regulate the transcription of nearby genes. ZNF80, a zinc-finger protein gene, is found to be transcribed from the R region of an upstream ERV-9 LTR, suggesting that the U3 enhancer and promoter of the ERV-9 LTR initiate mRNA synthesis of the ZNF80 gene (Di Cristofano *et al.*, 1995b). This LTR apparently plays a relevant role in the ZNF80 gene locus because the LTR is conserved in an identical position 5′ of the ZNF80 gene in the genome of gorilla and probably also of chimpanzee (Di Cristofano *et al.*, 1995a). cDNA clones of other genes that also contain the R region of the ERV-9 LTRs at the 5′ ends of the gene transcripts have been identified, suggesting that the transcription of these gene sequences was initiated by the ERV-9 LTR enhancer and promoter located 5′ of these gene sequences (Strazzullo *et al.*, 1998).

The transcription of an increasing number of human genes has been observed to be initiated within the proximal LTRs of the HERVs. The list includes the genes encoding cytochrome c_1 (Suzuki *et al.*, 1990), CDC4L (Feuchter *et al.*, 1992), phospholipase A_2-related gene (Feuchter-Murthy *et al.*, 1993), and Ku80 in the DNA-dependent protein kinase complex (Ludwig *et al.*, 1997). Furthermore, the tissue-specific expression of the salivary amylase and the pleitrophin genes has been shown to be conferred by the LTR enhancer and promoter sequences (Ting *et al.*, 1992; Schulte *et al.*, 1996).

The above observations indicate that the LTRs may be capable of regulating the synthesis of mRNAs of nearby genes in different tissues and may thus serve relevant cellular functions. However, in the β-like globin gene locus, the 5′ HS5 LTR is located in a position near the far upstream locus control region at a long distance from the β-like globin genes. Moreover, the ERV-9 LTRs were apparently not found in lower primates and nonprimates, including rodents (Di Cristofano *et al.*, 1995b), which nevertheless properly regulate their globin genes. It is thus possible that this 5′ HS5 LTR may not serve a relevant cellular function and that the LTR-initiated RNAs were sterile tran-

scripts devoid of any biological significance. However, the alternative possibilities that the 5′ HS5 LTR may serve a cellular function and that in the nonprimate species, including rodents, other related endogenous LTRs may be present in the β-globin LCR in place of the ERV-9 LTR cannot be excluded. The genomes of animals, including rodents, contain abundant endogenous retroviruses and LTRs (Coffin, 1984). Many of these LTRs affect the transcription of host genes (Fan, 1994) and have been suggested to provide a cellular reservoir of control elements with diverse transcriptional specificities (Keshet *et al.*, 1991).

A possible function of the 5′ HS5 LTR may be to regulate the transcription of the β-globin LCR in erythroid cells. We as well as others have recently reported that the β-globin LCR is preferentially transcribed in erythroid cells (Tuan *et al.*, 1992; Kong *et al.*, 1997; Ashe *et al.*, 1997). We suggested that this transcription process may serve important regulatory roles in globin gene expression in opening up the chromatin structure of the β-globin gene domain in erythroid cells and/or delivering transcription factors recruited by the LCR to the far downstream globin genes to activate globin mRNA synthesis (Tuan *et al.*, 1992; Cavallesco and Tuan, 1997; Kong *et al.*, 1997). In this respect, the transcriptional activation by the 5′ HS5 LTR of the DNA in the apparent 5′ boundary area of the LCR may be an upstream event of the transcriptional activation of the β-globin LCR. Whether or not the transcriptional process promoted by the 5′ HS5 LTR regulates the transcription of the far downstream β-like globin genes will be tested by LTR-knockout experiments in appropriate cellular systems.

ACKNOWLEDGMENTS

We thank Dr. S. Krantz for the CFU-E and T-lymphocytes, Dr. Y. Fei for the placenta RNA, Drs. M. Anderson, S. Lin, and H. Wang for assistance with computer graphics, and Dr. B. Whitney for helpful discussion and critical review of the manuscript. This work was supported in part by NIH Grants HL39948 and 48374.

REFERENCES

Anagnou, N. P., Perez-Stable, C., Gelinas, R., Costantini, F., Liapaki, K., Constantopoulou, M., Kosteas, T., Moschonas, N. K., and Stamatoyannopoulos, G. (1995). Sequences located 3′ to the breakpoint of the hereditary persistence of fetal hemoglobin-3 deletion exhibit enhancer activity and can modify the developmental expression of the human fetal Aγ-globin gene in transgenic mice. *J. Biol. Chem.* **270:** 10256–10263.

Ashe, H., Monks, J., Wijgerde, M., Fraser, P., and Proudfoot, N. (1997). Intergenic transcription and transinduction of the human β-globin locus. *Genes Dev.* **11:** 2494–2509.

Baralle, F., Shoulders, C., and Proudfoot, N. (1980). The primary structure of the human ϵ-globin gene. *Cell* **21:** 621–626.

Bi, W., Wu, L., Coustry, F., Crombrugghe, B., and Maity, S. (1997). DNA binding specificity of the CCAAT-binding factor CBF/NF-Y. *J. Biol. Chem.* **272:** 26562–26572.

Cavallesco, R., and Tuan, D. (1997). Modulatory subdomains of the HS2 enhancer differentially regulate enhancer activity in erythroid cells at different developmental stages. *Blood Cells Mol. Dis.* **23:** 8–26. [World Wide Web URL: http://www.scripps.edu/bcmd]

Chung, J. H., Whiteley, M., and Felsenfeld, G. (1993). A 5′ element in the chicken β-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell* **74:** 505–514.

Coffin, J. (1984). Endogenous retroviruses. *In* "RNA Tumor Viruses" (R. Weiss, N. Teich, H. Varmus, and J. Coffin, Eds.), Vol. 1, pp. 1109–1203, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Crossley, M., Whitelaw, E., Perkins, A., Williams, G., Fujiwara, Y., and Orkin, S. (1996). Isolation and characterization of the cDNA encoding BKLF/TEF-2, a major CACCC-box-binding protein in erythroid cells and selected other cells. *Mol. Cell. Biol.* **16:** 1695–1705.

Dhar, V., Nandi, A., Schildkraut, C. L., and Skoultchi, A. I. (1990). Erythroid-specific nuclease-hypersensitive sites flanking the human β-globin domain. *Mol. Cell. Biol.* **10:** 4324–4333.

Di Cristofano, A., Strazzullo, M., Longo, L., and LaMantia, G. (1995a). Characterization and genomic mapping of the ZNF80 locus: Expression of this zinc-finger gene is driven by a solitary LTR of ERV9 endogenous retroviral family. *Nucleic Acids Res.* **23:** 2823–2830.

Di Cristofano, A., Strazzullo, M., Parisi, T., and LaMantia, G. (1995b). Mobilization of an ERV9 human endogenous retroviral element during primate evolution. *Virology* **213:** 271–275.

Dolittle, R., and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genomic evolution. *Nature* **284:** 601–603.

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weismann, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C., and Produfoot, N. J. (1980). The structure and evolution of the human β-globin gene family. *Cell* **21:** 653–668.

Fan, H. (1994). Retroviruses and their role in cancer. *In* "The Retroviridae" (J. Levy, Ed.), Vol. 3, pp. 313–362, Plenum, New York.

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

Forrester, W. C., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G., and Groudine, M. (1987). Evidence for a locus activating region: The formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res.* **15:** 10159–10177.

Golemis, E. A., Speck, N. A., and Hopkins, N. (1990). Alignment of U3 region sequences of mammalian type C viruses: Identification of highly conserved motifs and implications for enhancer design. *J. Virol.* **64:** 534–542.

Goodchild, N., Wilkinson, D. A., and Mager, D. L. (1992). Human endogenous long terminal repeat provides a polyadenylation signal a novel, alternatively spliced transcript in normal placenta. *Virology* **21:** 287–294.

Grosveld, F., Assendelft, G. B. V., Greaves, D. R., and Kollias, G. (1987). Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell* **51:** 975–985.

Hardison, R., Slightom, J., Gumucio, D., Goodman, M., Stojanovic, N., and Miller, W. (1997). Locus control regions of mammalian β-globin gene cluster: Combining phylogenetic analyses and experimental results to gain functional insights. *Gene* **205:** 73–94.

Henikoff, S., Greene, E., Pietrokovski, S., Bork, P., Attwood, T., and Hood, L. (1997). Gene families: The taxonomy of protein paralogs and chimeras. *Science* **278:** 609–614.

Henthorn, P., Mager, D., Huisman, H., and Smithies, O. (1986). A gene deletion ending within a complex array of repeated sequences

3′ to the human β-globin gene cluster. *Proc. Natl. Acad. Sci. USA* **83:** 5194–5198.

Jarman, A. P., and Higgs, D. R. (1988). Nuclear scaffold attachment sites in the human globin gene complexes. *EMBO J.* **7:** 3337–3344.

Johnson, P., and McKnight, S. (1989). Eukaryotic transcriptional regulatory proteins. *Annu. Rev. Biochem.* **58:** 799–839.

Kellum, R., and Schedl, P. (1991). A position-effect assay for boundaries of higher order chromosomal domains. *Cell* **64:** 941–950.

Keshet, E., Schkff, R., and Itin, A. (1991). Mouse retrotransposons: A cellular reservoir of long terminal repeat (LTR) elements with diverse transcriptional specificities. *Adv. Cancer Res.* **56:** 215–251.

Ko, L. J., and Engel, J. D. (1993). DNA-binding specificities of the GATA transcription factor family. *Mol. Cell. Biol.* **13:** 4011–4022.

Kong, S., Bohl, D., Li, C., and Tuan, D. (1997). Transcription of the HS2 enhancer toward a cis-linked gene is independent of the orientation, position, and distance of the enhancer relative to the gene. *Mol. Cell. Biol.* **17:** 3955–3965.

La Mantia, G., Pengue, G., Maglione, D., Pannuti, A., Pascucci, A., and Lania, L. (1989). Identification of new human repetitive sequences: Characterization of the corresponding cDNAs and their expression in embryonal carcinoma cells. *Nucleic Acids Res.* **17:** 5913–5920.

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 embryonal carcinoma cells. *Nucleic Acids Res.* **19:** 1513–1520.

La Mantia, G., Majello, B., DiCristofano, A., Strazzullo, M., Minchiotti, G., and Lania, L. (1992). Identification of regulatory elements within the minimal promoter region of the human endogenous ERV9 proviruses: Accurate transcription initiation is controlled by an Inr-like element. *Nucleic Acids Res.* **20:** 4129–4136.

Lania, L., Di Cristofano, A., Strazzullo, M., Pengue, G., Majello, B., and LaMantia, G. (1992). Structural and functional organization of the human endogenous retroviral ERV9 sequences. *Virology* **191:** 464–468.

Lenz, J., Celander, D., Crowther, R. L., Patarca, R., Perkins, D. W., and Haseltine, W. A. (1984). Determination of the leukaemogenicity of a murine retrovirus by sequences within the long terminal repeat. *Nature* **308:** 467–470.

Li, Q., Powers, P. A., and Smithies, O. (1985). Nucleotide sequence of 16 kilobase pairs of DNA 5′ to the human β-globin gene. *J. Biol. Chem.* **260:** 14901–14910.

Li, Q., and Stamatoyannopoulos, G. (1995). Hypersensitive site 5 of the human β locus control region functions as a chromatin insulator. *Blood* **84:** 1399–1401.

Long, Q., Li, C., Nechtman, J., Tjia, J., Yoo, J., and Tuan, D. (1995). Nucleotide sequence and transcriptional analysis of DNA upstream of hypersensitive site 4 in the human β-globin locus control region. *Blood* **86:** 472a.

Lower, R., Lower, J., and Kurth, R. (1996) The viruses in all of us: Characteristics and biological significance of human endogenous retroviruses. *Proc. Natl. Acad. Sci. USA* **93:** 5177–5184.

Ludwig, D., Dhen, F., Peterson, S., Nussenzweig, A., Li, G., and Chen, D. (1997). Ku80 gene expression is Sp1-dependent and sensitive to CpG methylation within a novel cis element. *Gene* **199:** 181–194.

Merika, M., and Orkin, S. H. (1993). DNA-binding specificity of GATA family transcription factors. *Mol. Cell. Biol.* **13:** 3999–4010.

Miller, I., and Bieker, J. (1993). A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the *Kruppel* family of nuclear proteins. *Mol. Cell. Biol.* **13:** 1776–2786.

Nienhuis, A., Anagnou, N., and Ley, T. (1984). Advances in thalassemia research. *Blood* **63:** 738–758.

Orkin, S. H. (1992). GATA-binding transcription fators in hematopoietic cells. *Blood* **80:** 575–581.

Poncz, M., Sowiejczyk, D., Harpel, B., Mory, Y., Schwartz, E., and Surrey, S. (1982). Constructon of gene libraries from small amounts of peripheral blood. *Hemoglobin* **6:** 27–36.

Poncz, M., Schwartz, E., Ballantine, M., and Surrey, S. (1983). Nucleotide sequence analysis of the δβ-globin gene region in human. *J. Biol. Chem.* **258:** 11599–11609.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:** 5463–5467.

Schulte, A. M., Lai, S., Kurtz, A., Czubayko, F., Riegel, A. T., and Wellstein, A. (1996). Human trophoblast and choriocarcinoma expression of the growth factor pleiotrophin attributable to germ-line insertion of an endogenous retrovirus. *Proc. Natl. Acad. Sci. USA* **93:** 14759–14764.

Shen, S., Slightom, J., and Smithies, O. (1981). A history of the human fetal globin gene duplication. *Cell* **26:** 191–203.

Sibley, C., and Ahlquist, J. (1987). DNA hybridization evidence of hominoid phylogeny: Results from an expanded data set. *J. Mol. Evol.* **26:** 99–121.

Smit, A. (1996) The origin of interspersed repeats in the human genome. *Curr. Opin. Gen. Dev.* **6:** 743–748.

Speck, N. A., Renjifo, B., Golemis, E., Frederickson, T. N., Hartley, J. W., and Hopkins, N. (1990). Mutation of the core or adjacent LVb elements of the Moloney murine leukemia virus enhancer alters disease specificity. *Genes Dev.* **4:** 233–242.

Stavenhagen, J. B., and Robins, D. M. (1988). An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene. *Cell* **55:** 247–254.

Steif, A., Winter, D. M., Stratling, W. H., and Sippel, A. E. (1989). A nuclear DNA-attachment element mediates elevated and position-independent gene activity. *Nature* **341:** 343–345.

Strazzullo, M., Majello, B., Lania, L., and LaMantia, G. (1994). Mutational analysis of the human endogenous ERV9 proviruses promoter region. *Virology* **200:** 686–695.

Strazzullo, M., Parisi, T., DiCristofano, A., Rocchi, M., and LaMantia, G. (1998). Characterization and genomic mapping of chimeric ERV9 endogenous retroviruses–host gene transcripts. *Gene* **206:** 77–83.

Suzuki, H., Hosokawa, Y., Toda, H., Nishikim, M., and Ozawa, T. (1990). Common protein-binding sites in the 5′-flanking regions of human genes for cytochrome *c*1 and ubiquitine-binding protein. *J. Biol. Chem.* **265:** 8159–8163.

Temin, H. (1982). Structure, variation and synthesis of retrovirus long terminal repeat.*Cell* **27:** 1–3.

Ting, C., Rosenberg, M., Snow, C., Samuelson, L., and Meisler, M. (1992). Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev.* **6:** 1457–1465.

Tuan, D., Solomon, W., Li, Q., and London, I. M. (1985). The β-like-globin gene domain in human erythroid cells. *Proc. Natl. Acad. Sci. USA* **82:** 6384–6388.

Tuan, D., Solomon, W. B., London, I. M., and Lee, D. P. (1989). An erythroid-specific, developmental-stage-independent enhancer far upstream of the human β-like-globin genes. *Proc. Natl. Acad. Sci. USA* **86:** 2554–2558.

Tuan, D., Oh, Y. D., Venditti, C., Cavallesco, R., LeBoulch, P., Huang, G., and London, I. (1990). A distant erythroid enhancer in the regulation of human globin genes during erythropoiesis. *In* "Molecular Biology of Hematopoiesis" (N. G. Abraham, G. Konwalinka, L. Sachs, and C. G. Wiedermann, Eds.), Intercept Ltd., England.

Tuan, D., Kong, S., and Hu, K. (1992). Transcription of the hypersensitive site HS 2 enhancer in erythroid cells. *Proc. Natl. Acad. Sci. USA* **89:** 11219–11223.

Vogel, F., and Motulsky, A. (1986). "Human Genetics," pp. 534–538, Springer-Verlag, Berlin/New York.

Weber-Benarous, A., Cone, R., London, I., and Mulligan, R. (1988). Retroviral-mediated transfer and expression of human β-globin

genes in cultured murine and human erythroid cells. *J. Biol. Chem.* **263:** 6142–6145.

Weiss, M., Keller, G., and Orkin, S. (1994). Novel insights into erythroid development revealed through in vitro differentiation of GATA-1^{-} embryonic stem cells. *Genes Dev.* **8:** 1184–1197.

Wickrema, A., Krantz, S., Winklemann, J., and Bondurant, M. (1992). Differentiation and erythropoietin expression in human erythroid progenitor cells. *Blood* **80:** 1940–1946.

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

Yang, R., Fristensky, B., Deutch, A. H., Huang, R. C., Tan, Y. H., Narang, S. A., and Wu, R. (1983). The nucleotide sequence of a new human repetitive DNA consists of eight tandem repeats of 66 base pairs. *Gene* **25:** 59–66.

Yu, J., Bock, J. H., Slightom, J. L., and Villeponteau, B. (1994). A 5′ β-globin matrix-attachment region and the polyoma enhancer together confer position-independent transcription. *Gene* **139:** 139–145.

Zucchi, I., and Schlessinger, D. (1992) Distribution of moderately repetitive sequences pTR5 and LF1 in Xq24–q28 human DNA and their use in assembling YAC contigs. *Genomics* **12:** 264–275.