Nucleic Acids Research, 1995, Vol. 23, No. 15 **2823–2830**

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

Antonio Di Cristofano, Maria Strazzullo, Letizia Longo[1,+] and Girolama La Mantia*

Department of Genetics, General and Molecular Biology, University of Naples, Via Mezzocannone 8, 80134 Naples, Italy and [1]Department of Haematology, Hammersmith Hospital, RPMS, Du Cane Road, London W12 ONN, UK

Received May 26, 1995; Revised and Accepted June 27, 1995

EMBL accession nos X65233, X83496 and X83497

ABSTRACT

ERV9 is a low repeated family of human endogenous retroviral elements, which has close to 50 members, in addition to at least 4000 solitary LTRs. Previous work has shown that randomly selected LTRs can promote transcription of reporter genes, raising the possibility that these sequences may affect the expression of adjacent cellular genes. We describe here the structure of the ZNF80 cDNA clone putatively coding for a zinc-finger protein, whose 5′ terminus starts from within an ERV9-LTR. Characterization of the single copy genomic locus indicates that a complete ERV9-LTR element is present upstream of the ZNF80 coding region and that this element acts as a functional promoter in both *in vivo* and *in vitro* experiments. A 2.6 kb long transcript is selectively expressed only in some hematopoietic cell lineages. Interestingly we mapped the ZNF80 locus to the 3q13.3 band, a region involved in karyotype rearrangements associated with myelocytic disorders. We have also analyzed the ZNF80 genomic organization in African green monkey and we show that this lower primate does not harbour an ERV9 element at this locus. Our findings strongly suggest that the expression of a zinc finger gene, which is highly conserved during evolution of primates, is regulated in humans by an LTR element of the ERV9 family.

INTRODUCTION

Like the genomes of other vertebrates, human genome harbours several distinct families of endogenous retroviral sequences (HERV) that, altogether, comprise at least 0.6% of the human genome and are dispersed over the whole human genome (reviewed by 1,2). Irrespective of their evolutionary origins, the high copy number and dispersion of closely related endogenous retroviral sequences to various human chromosomes should have occurred primarily through repeated chromosomal insertions of reverse transcripts.

The nature of the LTR transcriptional regulatory elements influences both the capability to transpose and to affect expression of neighboring cellular genes. In several different cases, in mouse, insertions of intracisternal A-type (IAP) retroviral sequences near or within regulated genes resulted in their constitutive expression (3–5). Conversely, the androgen responsiveness of the mouse Slp gene seems to be conferred by a provirus-like LTR inserted 2 kb upstream of Slp (6). It is worthwhile to underline that in this latter case the phenotype is imposed by a highly defective element that has maintained functional LTR sequence.

Human endogenous retroviral sequences (HERV) have been shown to contribute regulatory sequences to some genes in humans (7–10). In one case (11), at some point in the evolutionary history, a HERV element altered regulation of a duplicated human gene for amylase, allowing the gene to be expressed in the salivary glands as well as in its original site, the pancreas.

We have been studying a human endogenous retrovirus family (ERV9), which has close to 50 members besides at least 4000 solitary LTRs and is conserved in primates but not in rodents (12–14). The prototype ERV9 genomic locus is ~8 kb long and the regions hybridizing to *gag*, *pol* and *env* sequences are flanked by LTRs (14). Since ERV9-LTRs are present in such large numbers in the genome, they may represent an important reservoir of potentially functional promoter sequences that could affect the pattern of expression of cellular genes. Support for this hypothesis has been provided by experiments showing that ERV9 LTRs derived from randomly selected units are capable of driving the expression of a reporter gene in a transient transfection assay (12). Moreover, we have characterized the basal promoter of the ERV9 LTRs and we have demonstrated that they can be differently stimulated by various transcription activation domains (15,16). It is therefore possible to hypothesize that the interactions with cellular sequences could result in functionally diverse promoters that can impose new patterns of regulation on cellular genes.

* To whom correspondence should be addressed

+Present address: Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, Box 110, New York, NY 10021, USA

We describe here the structural and functional characterization of the ZNF80 locus suggesting that the expression of this zinc finger gene is driven by a solitary ERV9-LTR promoter. We also report that this gene shows selective RNA expression in some hematopoietic cell lineages. Interestingly we mapped the ZNF80 locus to the 3q13.3 band, a region involved in karyotype rearrangements associated with myelocytic disorders.

Finally, we have also analyzed the ZNF80 genomic organization in African green monkey and we show that this lower primate does not harbour an ERV9 LTR upstream the conserved ZNF80 coding region.

MATERIALS AND METHODS

DNA probes

Probe A is a 1210 bp *SalI–XbaI* fragment, probe B a 360 bp *EcoRI* fragment and probe C a 300 bp *PstI-SalI* fragment, all of them derived from the cloned ZNF80 cDNA.

DNA libraries screening

The cDNA library from the T-lymphoma Peer cell line (cat. #HL 1078a, Clontech, Palo Alto, CA) was screened (7.5×10^5 plaques) with the ZNF80 specific probe A under stringent conditions, as previously described (12). Five cDNA clones were further analyzed after subcloning in pGEM3 plasmids. DNA sequencing was determined using the Sequenase kit (US Biochemical Co.). Genomic clones were isolated after screening with the ZNF80 specific probe (probe A) a lambda Fix human lung fibroblast genomic library (Stratagene). 5×10^5 plaques were screened and four positive plaques were isolated. The recombinant phages λFix1.2 and λFixPP1 were further analyzed by restriction enzymes analysis and DNA sequencing of relevant regions was performed after subcloning of appropriate fragments in pGEM-3 plasmid.

Oligonucleotides

RT1: 5′-TAGGGATCCTTGCTCCTGGAA-3′
RT2: 5′-AACGAGCTCCCACATTCCTTG-3′
RT3: 5′-AGCACTCTGAATTCAGTGGCAG-3′
RT4: 5′-TTCACCATCATTCTCTCGTCCC-3′

Genomic PCR and product cloning

Reaction mixtures (50 µl) containing ~200 ng DNA, 20 pmol each oligonucleotide primer (RT1/2), $MgCl_2$ 0.2 mM, dNTP 2.5 mM, were incubated with 2.5 U AmpliTaq DNA polymerase (Perkin-Elmer Cetus) under standard ionic conditions. Cycling was as follows: 3′ at 95°C, 25 cycles of 30″ at 94°C, 30″ at 57°C, 30″ at 72°C, using the Gene Amp PCR System 9600 (Perkin-Elmer Cetus). Amplified products were gel purified and cloned in pGEM-3 plasmid using the *Bam*HI and *Sac*I sites included in the oligonucleotides sequences.

RNA analysis

Premade RNA blots (MTN) were purchased from Clontech (Palo Alto, CA). Total RNA was isolated by the guanidine isothiocyanate method and poly(A)$^+$ mRNA was selected using oligo(dT) chromatography.

For Northern blot, 10 µg of poly(A)$^+$ RNA was separated on a 1.5% agarose gel and transferred to Hybond N$^+$ nylon membranes. Filters were hybridized to probe A according to standard procedures and washed with 0.1× SSPE, 0.1% SDS at 55°C.

For RT–PCR, 1 µg of RNA was reverse transcribed and amplified using the RNA PCR Kit from Perkin Elmer and the RT3 and RT4 oligonucleotides, amplifying a fragment encompassing the intronic sequence.

Amplification conditions were: 30″ denaturation at 94°C, 30″ annealing at 54°C and 30″ extension at 72°C for 40 cycles. The products were separated on a 1.5% agarose gel, transferred on nylon membrane and probed with the fragment corresponding to the amplified region (probe B).

In vitro translation

The 1420 bp long *Eco*RI cDNA fragment containing both the R-U5 region and the entire ORF was subcloned into plasmid pGEM3, downstream of the T7 promoter. Supercoiled plasmid (1 µg) was transcribed and translated *in vitro* in the presence of ^{35}S cysteine (Amersham) using the Promega TNT™ Coupled Reticulocyte Lysate System according to the conditions indicated by the supplier.

Five microlitre samples were separated on a 12% SDS-polyacrylamide gel. The gel was then fixed, dried and exposed overnight to X-ray film.

In vitro transcription and primer extension

The *in vitro* transcription experiments were performed using the 'Hela Cell Extract transcription System' Kit from Promega. Transcription products were analyzed by primer extension using a CAT primer as previously described (15). Reaction products were resolved on a 6% polyacrylamide–8 M urea gels. Sequencing ladders were generated by dideoxy-chain termination method using double strand DNAs as templates and the Sequenase kit from US Biochemical Corp. (Cleveland, OH).

CAT plasmid construction

The 10.1#2D and 10.1#2I-CAT plasmids were constructed by subcloning a 1.9 kb *Hin*dIII fragment derived from λFix1.2 and containing the U3-R and part of U5 regions in the promoterless p8CAT-0 (12) in both orientations.

Transfection and CAT assays

Jurkat T-leukemia cells were grown in RPMI1640 medium supplemented with 10% fetal calf serum. Transfections using DEAE dextran were performed as described (17). Thirty-six hours later the cells were harvested for CAT assays. For normalization of transfection efficiencies a β-Gal expression plasmid was included in the cotransfections. The ratio of acetylated form to the total was then normalized by comparison with the β-galactosidase activity measured using the β-galactosidase assay System Kit by Promega. The CAT activity was quantified by counting the amount of ^{14}C chloramphenicol converted to the mono-acetylated forms using the Molecular Dynamics PhosphorImager™ system.

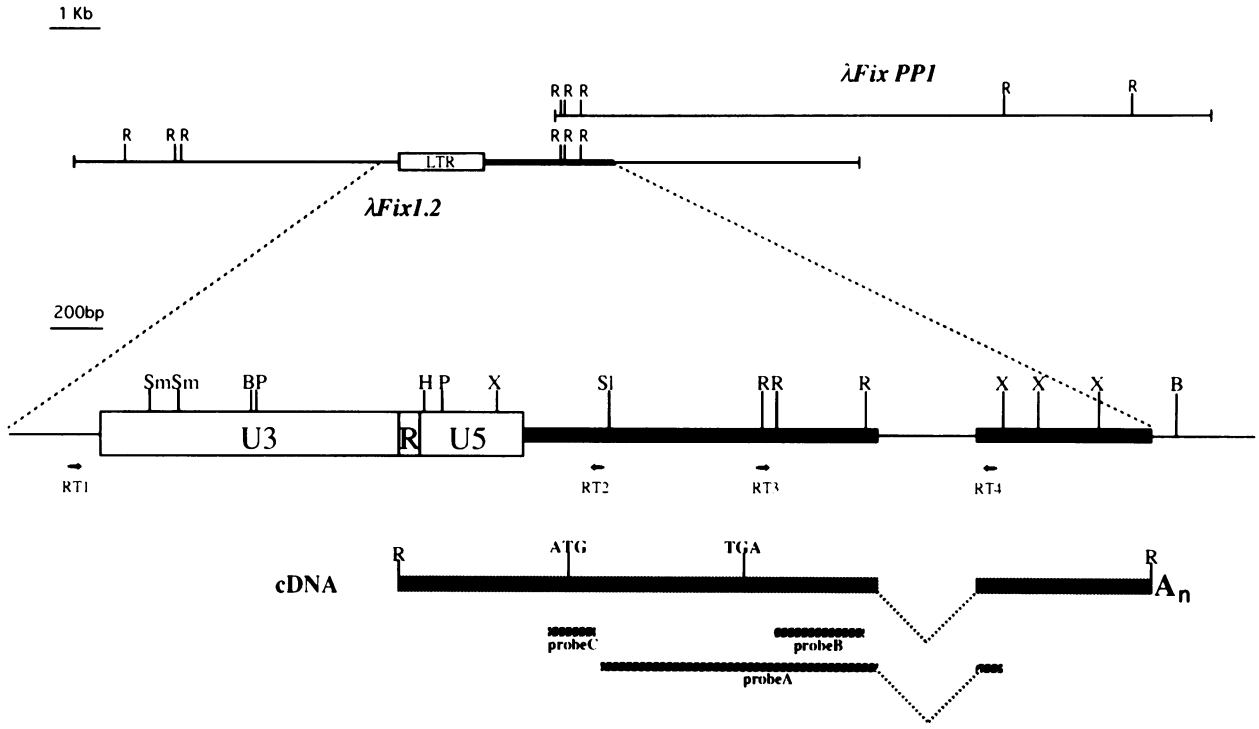

Figure 1. Genomic organization of the ZNF80 locus. Two recombinant phages are shown at the top; a schematic representation of the genomic locus is shown indicating the LTR element (boxed), the transcribed region (thick line) and the 390 bp intron (thin line). At the bottom is a representation of the ZNF80 cDNA and the probes utilized in this work (see Materials and methods). The positions of the oligonucleotides utilized in PCR amplifications (RT 1–4) are indicated by arrows. Sm, *Sma*I; B, *Bam*HI; P, *Pvu*II; H, *Hin*dIII; X, *Xba*I; Sl, *Sal*I; R, *Eco*RI.

Flourescence *in situ* hybridization (FISH)

Chromosome slides were prepared from cultures of human lymphocytes. FISH experiments were performed as described (18) with slight modifications, using the λFix1.2 phage probe. Slides were stained with propidium iodide (0.5 μg/ml) and counterstained with 4′,6′-diamidino-2-phenylindole (DAPI, 1 μg/ml). Fluorescent hybridization signals and DAPI images were captured by a cooled CCD camera (Photometrics, Tucson, AZ) and analyzed using the Smartcapture Imaging System (Vysis, IL).

RESULTS

Structure of ZNF80 cDNA and genomic locus

The ZNF80 cDNA clone, isolated from a human T-cell lymphoma (Peer) cDNA library (19), is 2580 bp long, contains a single 1053 bp long open reading frame, that begins at position 279 and extends up to an in-frame termination codon at position 1332; a putative ATG initiation codon is located at position 513 (Fig. 1). The predicted product contains 273 amino acids residues with a calculated molecular mass of 31 229 Da. Seven tandemly repeated zinc finger domains, already described (19) can be identified, preceded and followed by 50 and 34 amino acids respectively. The termination codon is followed by a 1250 bp long 3′ untranslated region and contains a canonical polyadenylation signal (AATAAA) 19 bp upstream from the poly(A) tail. The ZNF80 cDNA clone sequence can be retrieved from the EMBL library using the accession no. X65233.

To determine whether the cDNA clone could direct the synthesis of a protein of the predicted molecular mass, the 1.4 kb *Eco*RI fragment comprising the entire putative ORF was cloned into pGEM3 under the T7 promoter and then subjected to *in vitro* translation. The results show that a ~31 kDa polypeptide, thus similar in size to the predicted protein from the 273 amino acids long ORF, was synthesized (data not shown).

Computer-assisted analysis of the sequences preceding the ORF within the cDNA revealed the presence of a 446 bp long fragment corresponding to ERV9-LTR R and U5 region. Moreover, the first 16 nucleotides of the cDNA (from 7 to 22) derive from the U3 region (14–16).

To investigate the structure of the ZNF80 locus in the genome, Southern blot analysis of human DNA samples from different cell lines were performed, indicating that a single copy of ZNF80 gene is present in human genome. Moreover, digestions with *Hin*dIII, whose recognition site is present in the LTR region of the cDNA clone, gave bands of equal sizes in all cell lines, suggesting that the genomic arrangement predicted from the cDNA clone is not specific to the Peer cell line (data not shown, see Fig. 1).

As the isolation of the genomic ZNF80 locus may help to clarify the nature of ERV9-LTR involvement in the transcription of this gene, we screened with the non-LTR region from ZNF80 cDNA clone (probe A, Fig. 1) a human genomic library (WI38) and four independent clones were isolated. Two of them, Fix1.2 and FixPP1 clones (Fig. 1), overlapped in a genomic region of ~6 kb. ZNF80 genomic structure was determined by partial sequencing and various Southern blot hybridizations. Genomic and cDNA sequences were found colinear along the whole

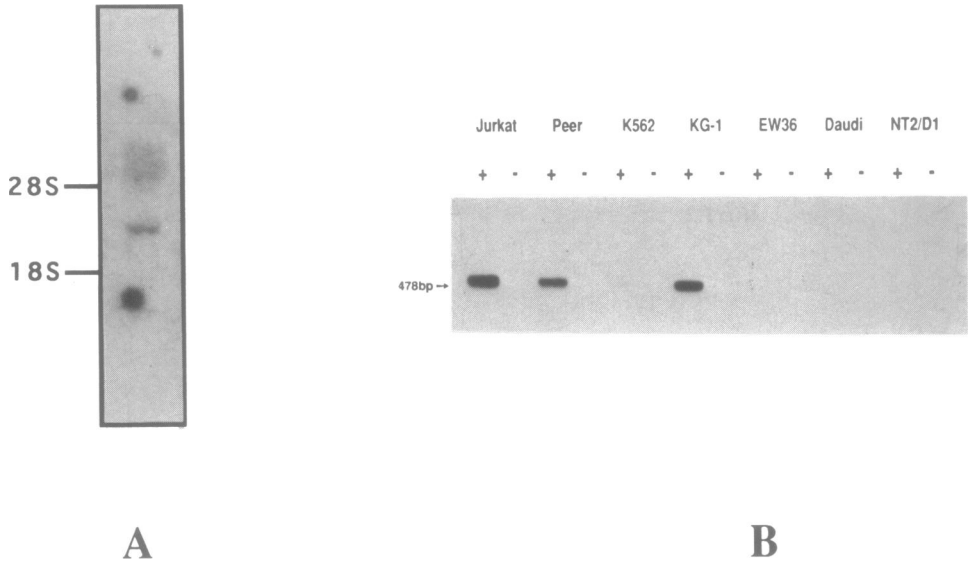

Figure 2. (**A**) Northern blot analysis of 10 μg polyA$^+$ from Peer cell line hybridized with the non-LTR region from the cDNA clone (probe A). (**B**) RT–PCR analysis of ZNF80 expression pattern in various cell lines using the oligonucleotides RT 3/4, hybridized with probe B; (+) with RT, (–) no RT. The RNAs utilized are from the following cell lines: Jurkat and Peer (T-cells), K562 (erythroleukemia), KG-1 (acute myeloblastic leukemia), EW36 and Daudi (B-cells), Ntera2/D1 (teratocarcinoma).

transcribed region, but the presence of a 390 bp intron in the 3′ untranslated region was deduced by hybridization of cDNA fragments to Southern blots of lambda genomic clones and was confirmed by sequencing. The intron/exons junction sequences were GAAAG/GTA and TTTTCTCAG/GGA, that are in good agreement with the complete consensus splice junction sequences, suggesting that this region is indeed a splicing position.

The 5′ flanking genomic region was sequenced, showing the presence of a 1709 bp long ERV9-LTR element, bordered by a 4 bp long inverted repeat (TGTG-CACA) with the characteristic features of other cloned ERV9-LTRs (12,14). The nucleotide sequence of the 540 bp long segment flanking the LTR was also obtained and computer-assisted analysis revealed no significant similarity with known sequences. Genomic LTR and flanking sequences have been submitted to the EMBL database and can be retrieved using accession no. X83497.

The Fix1.2 and FixPP1 clones were then analyzed by Southern blot experiments, using probes derived from ERV9-gag-pol and env regions (12) and the results clearly indicate that in the 23 kb long genomic region encompassed by the lambda clones no retroviral sequences were present (data not shown). These results strongly suggest that the 5′ end of the ZNF80 cDNA originates from within a solitary LTR.

ZNF80 expression pattern

A proviral LTR contains three functional regions, U3, R and U5 with the transcriptional regulatory elements located within the U3 region, while the transcription start site defines the beginning of the R region. Because of this arrangement, the 5′ end of a transcript driven by an LTR should not contain sequences from the U3 region.

The finding of U3 sequences at the 5′ end of the ZNF80 cDNA clone could indicate that this cDNA is not representative of the entire ZNF80 mRNA. It is possible to hypothesize the existence of a larger transcript, promoted by cellular sequences, containing a solitary ERV9-LTR, perhaps in an unspliced intron region. In this case we would expect to find transcripts longer and/or lacking LTR sequences.

However, we rescreened ~7.5 × 10^5 phages from the Peer cDNA library and we only found clones that, by restriction analysis and partial sequencing, resulted identical to the ZNF80 cDNA clone and two others, representing shorter transcripts (data not shown). Moreover, we performed a Northern blot analysis using polyA$^+$ RNA from the Peer cell line. The results obtained using as probe the non-LTR region from ZNF80 cDNA clone (probe A), show that a single 2.6 kb long transcript of a size similar to the cDNA clone is expressed, although at a very low level (Fig. 2A).

We have also investigated ZNF80 expression in a variety of normal tissues. Human multiple tissues Northern blots from Clontech Laboratory were used for this analysis, but a ZNF80 transcript was detected at very low amount only in peripheral blood leukocytes (data not shown).

To investigate further the tissue specificity of this gene, we performed a RT–PCR analysis. PolyA$^+$ RNA from Ntera2/D1, Jurkat and Peer (T-cells), K562 (erythroleukemia cell line), KG-1 (acute myeloblastic leukemia), EW36 and Daudi (B-cells) was used with RT3 and RT4 primers (shown in Fig. 1), complementary to sequences respectively upstream and downstream the intron. Figure 2B shows the products of these amplification reactions hybridized to probe B (Fig. 1), clearly indicating a selective expression in T-cells and in myeloid cells, but not in B-cells.

ZNF80-LTR promoter activity

The data reported so far seem to suggest that an endogenous LTR promotes the transcription of the ZNF80 gene. Obviously, the transcriptional competence of the regulatory region housed in the LTR is a prerequisite to such a hypothesis.

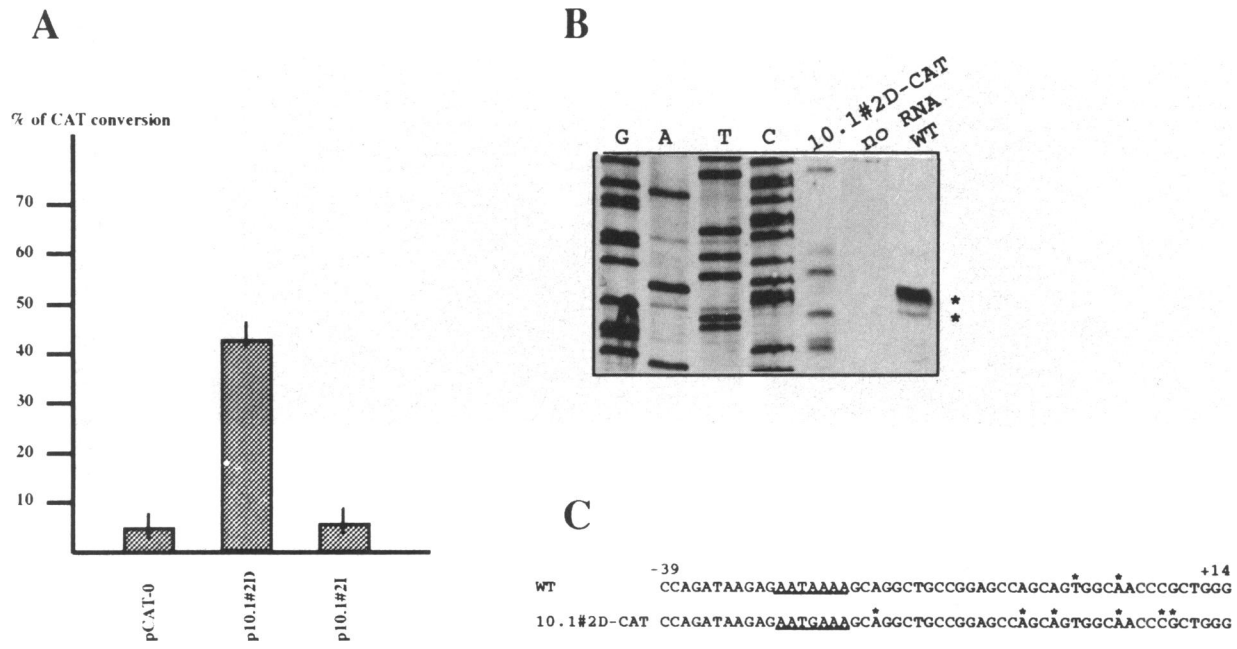

Figure 3. (**A**) Transient transfection analysis of the activity of the ZNF80-LTR promoter in Jurkat cell line: each column represents the mean value of three independent experiments. Standard deviations were <10%. (**B**) Primer extension mapping of the transcription start site. G, A, T, C: sequence ladder used to determine the exact position of the start sites; 10.1#2D: ZNF80-LTR clone; WT: ERV9 wild type LTR (16); the two wild type start sites are marked by asterisks. (**C**) Alignment of the sequences surrounding ERV9 and ZNF80 LTR start sites: non-canonical TATA-boxes are underlined, start sites are marked by asterisks.

To examine the functional capacity of the ZNF80-LTR, we tested the ability of a fragment bearing the U3 and R region to drive expression of a reporter CAT gene. The 10.1#2D and 10.1#2I-CAT plasmids (see Materials and methods), containing the 1.9 kb *Hin*dIII genomic fragment, in the right and opposite orientation respect to the CAT gene, were transfected in the Jurkat cell line. Clone 10.1#2D displayed 43% of CAT activity, while the putative promoter in the opposite orientation did not. These findings indicate that the ZNF80 LTR possesses a promoter activity (Fig. 3A).

We have previously defined all the elements required for the correct transcriptional activity of an ERV9-LTR promoter and we have shown, using an *in vitro* transcription assay, that for faithful transcription is necessary the integrity of a non-canonical TATA element (AATAAA) present at –28 relative to the major start site (16). In particular, mutation of the AATAAA sequence resulted in a greater heterogeneity of initiation sites, some of them located upstream the major start site and, therefore, upstream the functionally defined R region. As sequence analysis of the TATA corresponding region from the ZNF80-LTR showed the sequence AATGAA, we performed an *in vitro* transcription experiment, using the 10.1#2D-CAT plasmid. The products of the *in vitro* transcription were analyzed by extension from a primer complementary to the CAT mRNA sequences. The results show the presence of several extension products, indicating that ZNF80 promoter, unlike the previous analyzed ERV9-promoter, drives transcription initiation at numerous start sites, some of them located in the region previously defined as U3. These findings again suggest that the ZNF80 cDNA clone is representative of a transcript starting *in vivo* from upstream the wild-type major start site (Fig. 3B,C).

Chromosomal localization of the ZNF80 locus

ZNF80 locus was previously mapped by rodent-human somatic cell hybrid analysis to 3p12-3qter (19). This regional assignment has been now confirmed and further defined by fluorescence *in situ* hybridization (FISH). We observed in 70 out of 85 (82.3%) metaphases analyzed two symmetrical spots on both sister chromatids on the long arm of one of the two homologous chromosome 3 and in 15 out of 85 (17.6%) two specific yellow signals on both sister chromatids on the long arm of one of the two homologous chromosome 3. By Q-banding of the same metaphase preparations the phage Fix1.2 probe was mapped to the 3q13.3 band (Fig. 4).

ZNF80 genomic organization in African green monkey

We have previously reported that other primate species, but not rodents, contain ERV9 elements (12). We now decided to characterize the ZNF80 locus in African green monkey, in order to investigate if a lower primate contains an ERV9 element at the ZNF80 locus.

Initially, a Southern blot experiment was performed using DNA from African green monkey (CV-1 cell line) and humans. At high stringent conditions, using the probe A, a single hybridizing fragment is present in CV1 DNA digested with several restriction enzymes, suggesting a good conservation of ZNF80 sequences (data not shown).

PCR experiments were subsequently carried out using pairs of primers complementary to different sequences in the coding region and amplified products of similar size were obtained with both human and green monkey DNAs, indicating a strong

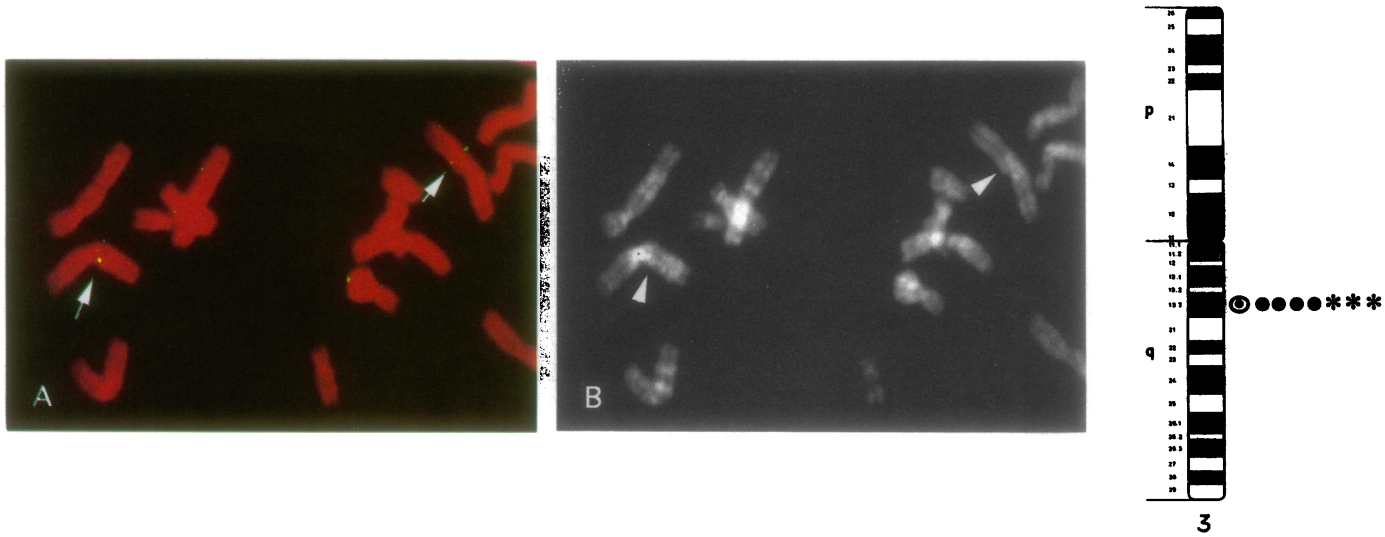

Figure 4. Chromosomal localization of ZNF80 locus: (**A**) is a propidium iodide-stained partial metaphase showing fluorescent hybridization signals (arrows). (**B**) is the same partial metaphase stained with DAPI to reveal chromosome banding pattern. The bands at which the hybridizations had occurred are identified by arrowheads. On the right side of the figure is the idiogram of chromosome 3 showing the distribution of FISH signals. Circles indicate hybridizations on both chromatids of the two homologous chromosomes, stars indicate hybridizations on both chromatids of one of the two homologous chromosomes. The large circle with a dot in the middle represents 50 hybridization events, each small circle and each star represent five hybridization events.

conservation of the transcribed sequences (data not shown). In particular an amplified product of the expected size (868 bp) hybridizing with probe B was obtained with primers RT3–RT4 (Fig. 1) in both DNAs (Fig. 5A).

PCR done with primers RT1 and RT2 (Fig. 1), flanking the LTR element in human DNA, yielded a 2022 bp long band when the probe D was used, as expected. Conversely an amplified product of 311 bp was found in DNA from CV-1 cells (Fig. 5A). This fragment was cloned (see Materials and Methods) and sequenced. As shown in Figure 5B the first 75 bp of monkey DNA are nearly identical to the sequence immediately preceding the LTR element in human (only three mismatches). The remaining 236 bp show more than 90% homology with human sequences located just downstream the LTR, thus clearly indicating that no ERV9 elements are present in green monkey ZNF80 locus. The ZNF80-ERV9 element, therefore, integrated into the apes genome 30 million years ago, surely after the divergence from the Old World monkey. Further experiments are in progress to better date the integration event.

DISCUSSION

We describe here the structure of the human ZNF80 locus, putatively coding for a zinc-finger protein and we present strong evidence showing that the transcription of this gene is driven by a solitary ERV9-LTR.

In principle, chimeric transcripts containing endogenous retroviral LTR sequences and cellular sequences can be produced by different mechanisms. Several cellular transcripts polyadenylated by an LTR have been identified (20–22). Moreover, as endogenous retroviral elements have been found in intronic region (2), one could expect to find unspliced transcripts bearing retroviral sequences. Furthermore, at least three reports have described transcripts initiating in the 5′ LTR of an endogenous provirus and splicing into a cellular downstream gene (8,9,23). Finally, a 3′ proviral LTR or a solitary LTR can directly promote the transcription of unrelated sequences (7).

ZNF80 cDNA contains at the 5′ terminal region a 446 bp long fragment corresponding to ERV9-LTR sequences. Genomic and cDNA sequences are colinear, except for the presence of a small intron in the 3′ untranslated region (Fig. 1). Moreover, sequence analysis of the genomic 5′ flanking region reveals the occurrence of a complete 1.7 kb long ERV9-LTR, while no other retroviral related sequences are present in a 23 kb long genomic region. These results strongly suggest that the 5′ end of the ZNF80 cDNA originates from within a solitary LTR.

Furthermore, the ZNF80-LTR possesses a promoter ability, as demonstrated by both *in vivo* and *in vitro* experiments. A fragment bearing the LTR-U3 and R region is, in fact, able to drive expression of a reporter CAT gene in Jurkat cell line (Fig. 3A), and an *in vitro* transcription assay indicates that the same fragment drives transcription initiation at numerous start sites (Fig. 3B), some of them located in the region previously defined by us (15) as U3.

The 5′ end of a retroviral transcript, that is, a transcript promoted by an LTR, does not contain U3 sequences, but only R and U5 sequences, as the 5′ terminus of R region is defined by the transcription initiation. Our previous experiments have demonstrated that for the selection of the correct start sites in an ERV9 promoter is important the integrity of the TATA-box sequence, AATAAA and that mutation of the wild type sequence results in a greater heterogeneity of initiation sites (16). The multiple start sites shown by the *in vitro* transcription assay are therefore justified by the observation that the TATA-box corresponding sequence in the ZNF80 U3 region is AATGAA, instead of AATAAA. The ZNF80 cDNA clone may represent, therefore, a transcript starting *in vivo* from upstream the wild-type start site, that is, in the U3 region.

However, it is also possible that the ERV9-LTR resides in an intron and, in this case, the ZNF80 cDNA clone could represent or an unspliced transcript, or a transcript occasionally starting from the intronic LTR. In both cases we would expect to find cDNA clones longer and/or lacking LTR sequences. No such clones were detected on rescreening the Peer cDNA library, and

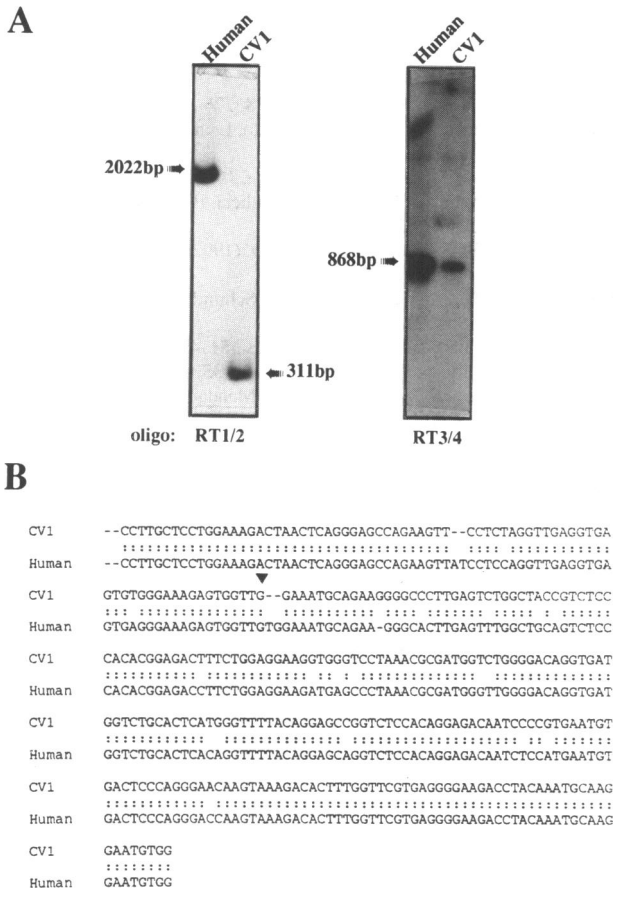

Figure 5. Analysis of ZNF80 structure in CV1 cell line. (**A**) PCR amplification of genomic DNA using oligonucleotides flanking the LTR in human DNA (RT 1/2, left) and oligonucleotides in the 3′ UTR region (RT 3/4, right) hybridized respectively, with probes C and B. (**B**) Alignment of CV1 and human DNA sequences amplified using RT 1/2: LTR insertion site (in human DNA) is marked by an arrow head.

more importantly, a single band of the expected size was detected in a Northern assay, giving confirmatory evidence of our hypothesis that an endogenous LTR plays a role in the transcriptional regulation of a cellular gene.

The possible biological significance of this occurrence is particularly intriguing, as the ZNF80 gene is a member of the zinc finger family of transcriptional regulatory proteins, often involved in controlling gene activity in development and differentiation (24). Interestingly, our expression analysis clearly indicated a selective expression of the ZNF80 gene only in some hematopoietic cell lineages. We don't know yet if the cell type specificity is conferred only by the ZNF80-LTR sequence or by the juxtaposition of cellular and retroviral sequences. Further experiments are necessary to exploit both this point and the possible regulative role of this finger gene in hematopoietic cell lineages.

Moreover, we mapped the ZNF80 locus to the 3q13.3 band, a region involved in karyotype rearrangements associated with myelocytic disorders (25,26). Experiments are currently in progress to determine whether leukemias containing abnormalities of this region of chromosome 3 have rearrangements of the ZNF80 locus.

The finding of a cellular transcript promoted by an ERV9-LTR is not unexpected, as ERV9 family contains ~50 integrated proviruses, but at least 4000 solitary LTRs, dispersed in many genomic loci and potentially functional promoters/enhancers (15).

The generation of a solo LTR can be explained principally by recombination between 5′ and 3′ LTRs, leading to excision of proviral sequences (27,28). However, this mechanism has to be preceded by the integration of a new retroviral element by intracellular retrotransposition. The events leading to a new solitary LTR may involve stable alterations to the germ line, transmitted vertically, as well as somatic occurrence, altering the expression of a cellular gene and perhaps affecting cell growth. Data presented here indicate that the ERV9-LTR element is a normal part of the ZNF80 locus in humans, as we found the same genomic arrangement in several unrelated normal human DNAs tested.

The extent to which retrotransposons have affected evolution of their host genomes, especially with respect to host gene expression, is an intriguing topic. While new transposition events can be deleterious, it is possible that proviral association with cellular genes could alter gene expression in a benign way, resulting in an evolutionary preserved genetic variation (6,11).

In this paper we have analyzed the structure of the ZNF80 finger gene in the green monkey genome, in an attempt to determine the evolutionary age of the insertion. Although the coding and flanking sequences seem to be highly conserved, a major structural difference between the two related loci is that no ERV9 elements are present in green monkey ZNF80 locus, indicating, therefore, that a retroviral element integrated into the apes genome less than 30 million years ago, after the divergence from Old World monkeys. Further experiments are in progress to trace the origin of the transposition event and to evaluate the possible impact of this event on both gene structure and expression. However, the finding that the ERV9 insertion upstream the ZNF80 gene is not present in green monkey is an indication that this family of elements has been transpositionally active since the two species diverged.

In conclusion we report here that the expression of a zinc finger gene, which is highly conserved during evolution of primates, is regulated in humans by an LTR element of the ERV9 family. Examinations of other members of this proviral family may lead to other genes that are candidates for acquired changes in gene expression due to neighboring transposable elements.

ACKNOWLEDGEMENTS

We thank Prof. L. Lania for useful discussions and critical reading of the manuscript. We would like also to thank Miss R. Terracciano for skilful technical help. This work was paid for by grants from European Community (EC Grant no. Gene-CTP3-0019), the Italian Association for Cancer Research (AIRC) and from the AIDS program of the Istituto Superiore di Sanit (Rome) and MURST to G.L.M.; M.S. is recipient of an AIRC fellowship.

REFERENCES

1 Leib-Mosch, C., Brack-Werner, R., Werner, T., Bachmann, M., Faff, O., Erfle, V. and Hehlmann, R. (1990) *Cancer Res.*, **50**, 5636–5642.
2 Wilkinson, D.A., Mager, D.L. and Leong, J.C. (1994) in Levy, J. (ed.), *The Retroviridae*. Plenum Press, New York, pp. 465–535.

3 Kongsuwan, K., Allen, J. and Adams, J.M. (1989) *Nucleic Acids Res.*, **17**, 1881–1892.
4 Leslie, K.B., Lee, F. and Schrader, J.W. (1991) *Mol. Cell. Biol.*, **11**, 5562–5570.
5 Ymer, S., Tucker, W.Q.J., Sanderson, C.J., Hapel, A.J., Champbell, H.D. and Young, I.G. (1985) *Nature*, **317**, 225–258.
6 Stavenhagen, J.B. and Robins, D.M. (1988) *Cell*, **55**, 247–254.
7 Feuchter, A., Freeman, J.D. and Mager, D.L. (1992) *Genomics*, **13**, 1237–1245.
8 Feuchter-Murthy, A.E., Freeman, J.D. and Mager, D.L. (1993) *Nucleic Acids Res.*, **21**, 135–143.
9 Kato, N., Shimotohno, K., VanLeeuwen, D. and Cohen, M. (1990) *Mol. Cell. Biol.*, **10**, 4401–4405.
10 Suzuki, H., Hosokawa, Y., Tod, H., Nishikimi, M. and Ozawa, T. (1990) *J. Biol. Chem.*, **265**, 8159–8164.
11 Ting, C.N., Rosenberg, M.P., Snow, C.M., Samuelson, L.C. and Meisler, M.H. (1992) *Genes Dev.*, **6**, 1457–1465.
12 La Mantia, G., Maglione, D., Pengue, G., Di Cristofano, A., Simeone, A., Lanfrancone, L. and Lania, L. (1991) *Nucleic Acids Res.*, **19**, 1513–1520.
13 La Mantia, G., Pengue, G., Maglione, D., Pannuti, A., Pascucci, A. and Lania, L. (1989) *Nucleic Acids Res.*, **17**, 5913–5922.
14 Lania, L., Di Cristofano, A., Strazzullo, M., Pengue, G., Majello, B. and La Mantia, G. (1992) *Virology*, **191**, 464–468.
15 La Mantia, G., Majello, B., Di Cristofano, A., Strazzullo, M., Minchiotti, G. and Lania, L. (1992) *Nucleic Acids Res.*, **16**, 4129–4136.
16 Strazzullo, M., Majello, B., Lania, L. and La Mantia, G. (1994) *Virology*, **200**, 686–695.
17 Leung, K. and Nabel, G. (1988) *Nature*, **338**, 776–778.
18 Bessler, M., Hillmen, P., Longo, L., Luzzatto, L. and Mason, P. (1994) *Hum. Mol. Genet.*, **3**, 751–757.
19 Huebner, K., Druck, T., LaForgia, S., Lasota, J., Croce, C.M., Lanfrancone, L., Donti, E., Pengue, G., La Mantia, G., Pelicci, P.G. and Lania, L. (1993) *Hum. Genet.*, **91**, 217–222.
20 Goodchild, N., Wilkinson, D. and Mager, D. (1992) *Gene*, **121**, 287–294.
21 Mager, D. (1989) *Virology*, **173**, 591–599.
22 Paulson, K.E., Matera, A.G., Deka, N. and Schmid, C.W. (1987) *Nucleic Acids Res.*, **15**, 5199–5215.
23 Liu, A. and Abraham, B. (1991) *Cancer Res.*, **51**, 4107–4110.
24 El-Baradi, T. and Pieler, T. (1991) *Mech. Dev.*, **35**, 155–169.
25 Billstrom, R., Heim, S., Kristoffersson, U., Mandahl, N. and Mitelman, F. (1990) *Eur. J. Haematol.*, **45**, 150–152.
26 Michalova, K., Bartsch, O., Stary, J., Jelinek, J., Wiegant, J. and Bubanska, E. (1993) *Cancer Genet. Cytogenet.*, **71**, 67–70.
27 Mager, D. and Goodchild, N. (1989) *Am. J. Hum. Genet.*, **45**, 848–854.
28 Paulson, K., Deka, N., Schmid, C., Misra, R., Schindler, C., Rush, M., Kadyk, L. and Leinwand, L. (1985) *Nature*, **316**, 359–361.