Nucleic Acids Research, Vol. 18, No. 5

Functional heterogeneity of a large family of human LTR-like promoters and enhancers

Anita Feuchter and Dixie Mager*
Terry Fox Laboratory, BC Cancer Research Centre and Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

Received August 10, 1989; Revised and Accepted January 29, 1990

ABSTRACT

The human genome contains a variety of elements similar in structure to retroviruses and retrotransposons. We have shown that the long terminal repeat (LTR) sequences of a large family of human retrovirus-like elements, RTVL-H, are heterogeneous in their ability to regulate the expression of linked genes. Although all of five LTRs tested could promote expression of the chloramphenicol acetyltransferase (CAT) gene, their relative promoter activities as well as range of activities varied widely. Several of the LTRs tested also exhibited bidirectional promoter activity either alone or when activated by an SV40 early enhancer. One LTR, H6, displayed strong promoter activity in human (NTera2D1, 293, Hep2), monkey (COS-1), and mouse (3T3) cells. In fact, the activity of this LTR was similar to that of the SV40 early promoter/enhancer in 293, COS-1, and 3T3 cells. RNA mapping studies have localized the transcription start site to the expected location in the H6 LTR. RTVL-H LTRs were also shown to contain sequences which could increase transcription from the human β-globin promoter and be influenced by SV40 enhancer sequences. As the human genome contains several hundred related RTVL-H sequences and a similar number of solitary LTRs, these findings raise the possibility that RTVL-H LTRs could have diverse effects on the expression of adjacent cellular genes.

INTRODUCTION

In recent years, several classes of elements structurally related to integrated retroviruses (proviruses) or retrotransposons have been discovered in the human genome (for review, ref. 1). In many cases, the LTR-like sequences associated with these elements contain potential transcriptional regulatory sequences which are analogous to those found in 'true' retroviral LTRs (2–4). These findings raise the possibility that human endogenous LTR-like sequences may function to regulate the expression of linked genes. Such genes could be those contained within the retrovirus-like element itself or they could be adjacent cellular genes. Indeed, in mouse cells, several genetic rearrangements involving LTR-like sequences have been implicated in gene inactivation (by insertional mutagenesis) (5,6) and in gene activation (by acting as mobile promoters or enhancers) (7–10).

A proviral LTR contains three functional regions U_3, R, and U_5, U_3 and U_5 are unique sequences derived from the 3′ and 5′ ends of the viral RNA, while R is a short sequence present at both termini of the RNA genome. The strongest viral transcriptional enhancer sequences are typically found in U_3, as are transcriptional promoter signals such as the CCAAT and TATAA boxes which are located approximately 75 and 22–26 bp upstream of the transcriptional start site, respectively. The transcriptional start site, usually GC, defines the beginning of the R region which, in mammalian retroviruses, also contains the polyadenylation signal followed 15–20 bp downstream by the polyadenylation site, CA (LTR structure is reviewed in refs 11–13). LTRs are found not only in proviruses but also in retrotransposons such as Ty in yeast (14) and copia in Drosophila melanogaster (15) where they serve similar functions.

We have been studying a family of human endogenous retrovirus-like elements, RTVL-H, which has close to 1000 members in addition to a similar number of solitary LTRs (16). Prototypic RTVL-H elements are 5.8 kb in length and contain regions of *gag* and *pol* homology to other retroviruses (17) but do not contain an *env* equivalent region. Therefore, although designated as 'retrovirus-like', these elements may be structurally more similar to retrotransposons (14) or to endogenous murine elements such as VL30 (18) which also lack an *env* region. RTVL-H LTRs contain direct repeats, a TATAA box, and a polyadenylation signal in positions analogous to those found in prototypical LTRs. RNAs homologous to RTVL-H sequences are found in certain human cells (19, Wilkinson et al., submitted) and we have shown that RTVL-H LTRs do promote endogenous expression of these elements (Wilkinson et al., submitted). These observations indicate that RTVL-H LTRs contain transcriptional regulatory sequences that function in particular cell types. Since these LTRs are present in such large numbers in the genome, we have been interested in examining the functional heterogeneity and cell type specificities of these elements. In this study, we have measured the ability of different RTVL-H LTRs to drive expression of the reporter gene chloramphenicol acetyltransferase (CAT) in various mammalian cells. We report here that the five

* To whom correspondence should be addressed at Terry Fox Laboratory, BC Cancer Research Centre, 601 West 10th Ave., Vancouver, BC V5Z 1L3, Canada

LTRs tested had widely different abilities to promote CAT gene expression. We also demonstrate that these LTRs can be activated by heterologous enhancer sequences as well as act to enhance a heterologous promoter. RTVL-H LTRs thus constitute a large family of endogenous sequences with the potential to affect human gene expression.

MATERIALS AND METHODS

Cell Lines and Transfections

The adenovirus-transformed human embryonal kidney cell line 293 was obtained from the American Type Culture Collection. Hep2 cells (a subline of HeLa; 20) were obtained from Dr. Paula Henthorn. The human teratocarinoma cell line NTera2D1 (21) was provided by Dr. Peter Andrews. Mouse Ltk$^-$ and NIH 3T3, and COS-1 (African green monkey kidney cells transformed by SV40) cells were provided by Dr. Keith Humphries. The CV-1 cell line (the non-transformed parent of COS-1 cells) was obtained from Dr. Jurgen Vielkind. All cells were grown in Dulbecco Modified Eagle Medium supplemented with either 10% horse serum (293 cells) or 10% fetal calf serum (others) and were maintained at 37°C with 5% CO_2. Subconfluent cell cultures were transfected with supercoiled plasmid DNA by calcium phosphate precipitation (22). Fresh media was added to the cells 24 hours post transfection.

Description of LTRs

The five RTVL-H LTRs used in this study were obtained from either genomic or cDNA libraries (see figure 1a). In each case, the portion of the LTR analyzed extended from a highly conserved StuI restriction enzyme site, eight bp downstream of the start of the LTR, to a slightly variable point near the polyadenylation (poly A) signal. The genomic clones containing the 3'LTR from the RTVL-H1 element (3'R1) and the 5' LTR from the RTVL-H2 element (5'R2) have previously been described (16, 17). Both of these LTRs have a HindIII site three bp downstream of the poly A signal and were isolated as StuI/HindIII fragments. The remainder of the LTRs, PB-3, N10-14, and H6 were isolated from cDNA libraries. Because each of these LTRs had functioned to polyadenylate transcripts (see below), they terminated in a poly A tail which was removed before the LTRs were tested for promoter activity. Thus, for our analyses we subcloned each LTR into the plasmid pUC18 and deleted the poly A tail with exonuclease III (exo III). The LTR fragments were then isolated by digestion with StuI and a second enzyme specific to the pUC18 polylinker, as described below. For this reason, each LTR subcloned into the CAT construct contains a few nucleotides of downstream pUC18 polylinker sequences.

The PB-3 LTR was isolated from a human phytohemmaglutinin stimulated peripheral blood cDNA library and had functioned to polyadenylate a non-RTVL-H related transcript (23). The region used in our analysis was isolated from a pUC18-PB-3 construct, after exo III treatment, by a StuI/HindIII digestion. It extended to a position one bp downstream of the poly A signal and contained six bp of plasmid derived sequence. The N10-14 LTR is the 3' polyadenylated LTR from an NTera2D1 cDNA clone that appears to represent a spliced RTVL-H transcript (Wilkinson et al., submitted). The LTR fragment used in this study was isolated from a pUC18-N10-14 construct by a StuI/SphI digest and extended from the 5' StuI site to the first A of the poly A signal followed by two bp of plasmid sequence. The H6 LTR is the 3' polyadenylated LTR isolated from a Hep2 cDNA clone that contains approximately 500 bp of upstream internal RTVL-H sequences (23). The portion of the LTR analyzed extended from the StuI site to one bp downstream of the poly A signal. It contained seven bp of polylinker sequences generated from a StuI/BamHI digestion.

Plasmid Constructions

The expression vectors pSVA0CAT(X), pSV2ACAT, pSVA0CAT(LR), pSV232ACAT(LR), pSVAβGCAT(X), and pSVAβGCAT(LR) have previously been described (24−26) and were kindly provided by Dr. Paula Henthorn and Dr. Tom Kadesch. Test vectors were constructed by blunt end ligation of each LTR fragment into the HindIII site of pSVA0CAT(X) or pSVA0CAT(LR) (figure 1) or into the XbaI site of pSVAβGCAT(X) (figure 7a) in both the forward and reverse orientations.

CAT Assays

Cell extracts were isolated 44−50 hours post transfection and assayed for CAT activity (27) with 0.2 μCi of ^{14}C chloramphenicol used as a substrate. Percentage conversion of chloramphenicol into its acetylated forms was determined by scintillation counting of the radioactive spots after their resolution by thin layer chromatography and autoradiography. Experiments were repeated 3 to 8 times and average percentage conversion values determined.

Primer Extension Analysis

RNA was isolated from five confluent 150×25 mm plates of cells by guanidinium isothiocyanate preparation (28) 48 hours after transfection with a total of 150 μg of the appropriate plasmid. A 21 bp oligonucleotide corresponding to a region of the CAT gene 100 bp downstream of the HindIII site was 5' end labelled and primer extension analysis was performed (29) with 50 μg total cellular RNA and 2×10^5 cpm of 5' end labelled oligonucleotide primer. The products were treated with 300 μg/ml RNAse A at 40°C for 60 minutes, phenol/chloroform extracted and run on a sequencing gel alongside the appropriate plasmid which had been sequenced by a modification of the chain termination method (30) using the same oligonucleotide primer.

RESULTS

RTVL-H LTR Promoter Activities Are Heterogeneous

To examine the functional capacities of different RTVL-H LTRs, we have made use of the system in which the bacterial gene coding for CAT can be placed under the transcriptional control of heterologous promoter sequences. Because eukaryotic cells do not contain CAT, the level of this enzyme in cells transfected with a CAT plasmid is a quantitative measurement of the strength of the promoter under consideration. The construction of the vectors used in this analysis is outlined in figure 1b. The parent plasmid pSVA0CAT(X) is promoterless, with the CAT coding sequences followed by SV40 early polyadenylation signals. An SV40 poly A signal dimer (24) has been inserted as indicated to prevent 'readthrough' transcription from cryptic promoters located elsewhere in the plasmid. pSVA0CAT(LR) is a modification of pSVA0CAT(X) in which the SV40 early enhancer region, LR, (24) has been inserted downstream of the CAT gene as shown. The vectors pSV2ACAT and pSVA232CAT(LR) (not shown) were used as positive controls for experiments involving

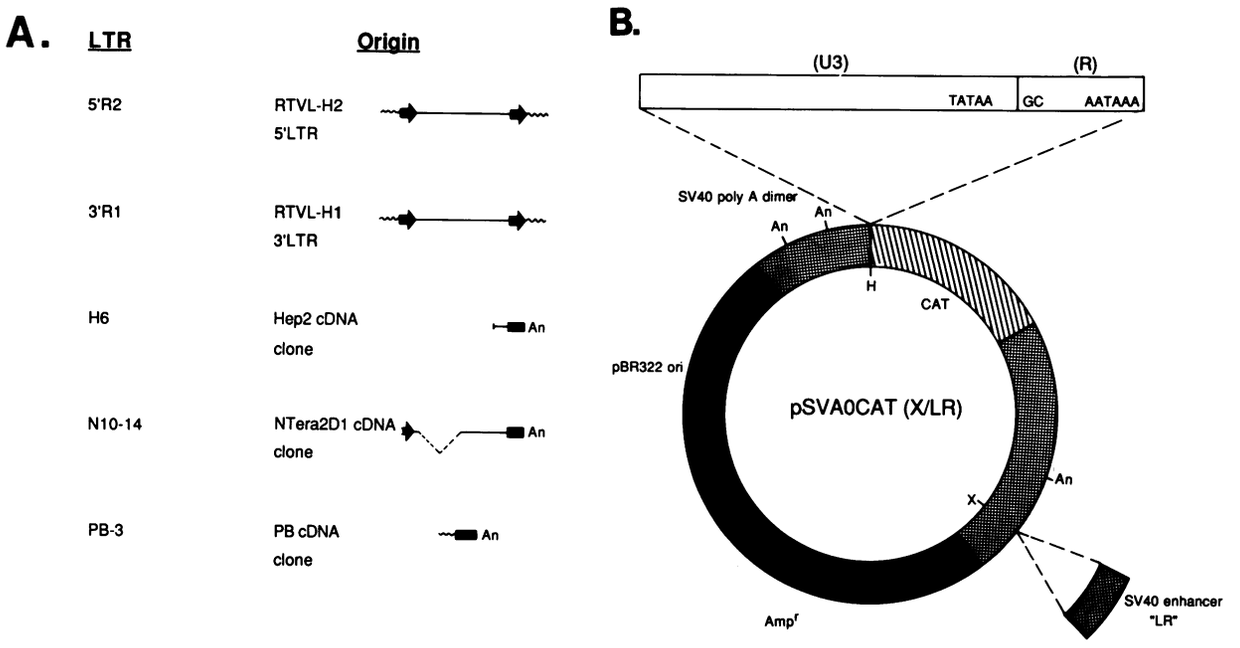

Figure 1. Construction of plasmids used to test the promoter activity of RTVL-H LTRs. (a) The RTVL-H LTRs tested were isolated from genomic or cDNA libraries (see Materials and Methods). Straight lines indicate RTVL-H interior sequences and thick arrows and boxes indicate complete LTR sequences and partial LTR sequences (due to polyadenylation), respectively. 'A_n' indicates that the polyadenylation signal within the LTR has been utilized. Wavy lines represent non-RTVL-H related genomic sequences, and the dotted lines in the N10-14 interior sequences indicate a putative splicing event (Wilkinson et al., submitted). (b) Insertion of the putative U_3 and R regions of the LTRs, in both the forward and reverse orientations, into the CAT expression vectors (forward orientation is shown). The 3' extent of each LTR sequence varied by a few nucleotides, as described in Materials and Methods. The LTR sequences were inserted at the unique HindIII site (H). Striped, CAT coding sequences; black, pBR322 sequences; stippled, SV40 sequences. 'A_n' denotes the location of SV40 derived polyadenylation signals. One set of constructs had the SV40 early enhancer region, LR, inserted as shown. The 'X' refers to the multiple cloning site of the CAT(X) constructs which replaces 545 bp of plasmid sequence in the CAT(LR) constructs.

the CAT(X) and CAT(LR) constructs, respectively. The SV40 early promoter/enhancer region directs transcription in both of these vectors, however in the 232ACAT(LR) construct the enhancer sequences have been deleted from the promoter region and replaced by the LR segment, as indicated in figure 1b.

A unique HindIII site located ~35 bp upstream of the CAT initiation codon was used to insert the portions of the LTRs shown, in both the forward (f) and reverse (r) orientation with respect to analogy with other LTRs (figure 1). The LTRs tested were derived from genomic or cDNA libraries as shown and as described in Materials and Methods. We chose to analyze two randomly isolated genomic LTRs as well as LTRs derived from transcripts expressed in three different cell types (figure 1a) to increase the probability of detecting LTRs with different functional capabilities. In the case of the cDNA clones, LTR containing segments were isolated from the 3' end of the transcript for analysis. We expect these sequences to be similar, if not identical, to the linked 5' LTR as the 5' and 3' LTRs of intact RTVL-H elements show an average of 96% sequence identity, while unlinked RTVL-H LTRs range from between 75–95% homologous (23). The U_3 region of the 5' LTR most probably acted as the promoter for two of these transcripts (see below) but would not be present in the cDNA clone since transcription initiates at the 3' boundary of U_3 (11–13). The NTera2D1 cDNA clone from which the N10-14 LTR was isolated corresponds to a transcript which initiated in the 5' LTR (Wilkinson et al., submitted; figure 1a). The 5' terminus of this cDNA clone mapped to 40 bp downstream of the TATAA box in the LTR. Thus the N10-14 LTR tested here should be very close in sequence to the linked 5' LTR which functions in NTera2D1 cells. The H6 LTR is derived from a partial Hep2 (HeLa) cDNA clone which also most probably represents an LTR initiated transcript since unit-length RNAs are found in these cells (Wilkinson et al., submitted). The PB-3 LTR had functioned to polyadenylate an unrelated transcript and was chosen because it differs substantially in sequence from the other LTRs tested (23).

The various LTR^{f}CAT(X) constructs were transfected into the human embryonal carcinoma line NTera2D1, the human embryonal kidney line 293, and the HeLa subline Hep2 by calcium phosphate mediated precipitation. Cell lysates were measured for CAT activity 44–50 hours later. A typical CAT assay is shown in figure 2 (a,b, and c). The results clearly indicate that some LTRs are capable of directing CAT expression in these cells, although their individual promoter strengths vary considerably. The H6 LTR consistently showed the strongest promoter capabilities; its activity was approximately 84%, 44% and 38% of the SV40 promoter in 293, Hep2 and NTera2D1 cells, respectively. The N10-14 LTR showed substantially weaker activity, 4.4% and 1.3% of SV40, in NTera2D1 and 293 cells respectively and did not promote detectable levels of CAT activity in Hep2 cells. The other 3 LTRs tested showed no activity above background levels in these three human cell lines.

To determine whether the RTVL-H LTR^f promoters showed species or tissue specificity, we tested the constructs in mouse (Ltk$^-$, NIH 3T3) and monkey (COS-1) cells (figure 2d, e, and f). The H6 LTR again showed strong promoter activity in NIH 3T3 and COS-1 cells, but did not function to a detectable level

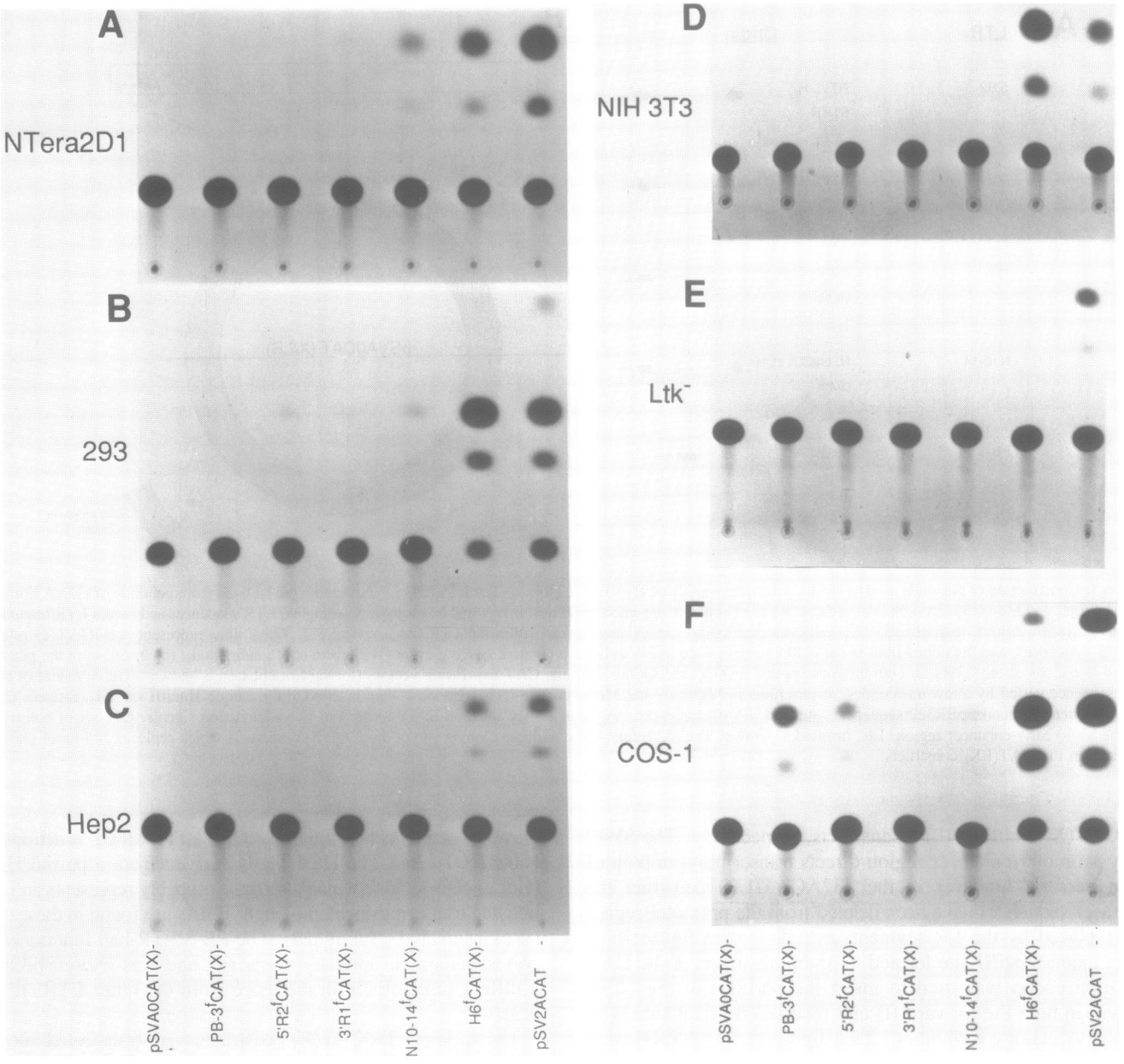

Figure 2. Assay of RTVL-H LTR promoter activity in different cell lines. Cell extracts (100 μl) were prepared 48 h post transfection with the indicated plasmid and assayed at 37°C for 60 min as previously described (25). CAT activity was measured by conversion of ^{14}C chloramphenicol to its acetylated forms. The products were separated by thin layer chromatography and detected by autoradiography. Extracts prepared from a) NTera2D1, b) 293, c) Hep2, d) NIH 3T3, e) Ltk$^-$, and f) COS-1 cells are shown. The bottom spot in each lane is the unreacted chloramphenicol. The spots directly above, in the positive lanes, are the mono (lower two spots) and di-acetylated products.

in Ltk$^-$ cells. No other RTVL-H LTR tested promoted transcription in the two mouse cell lines tested. In COS-1 cells, however, several of the LTRs functioned to some extent (figure 2f). The H6 LTR remained the strongest and the PB-3 LTR, which did not function in any of the previously examined cells, displayed a significant level of activity. Low activities were also promoted by the 5′R2 LTR. Although not evident in the figures shown, the 3′R1 LTR showed low levels of activity in NTera2D1 and COS-1 cells in other experiments (results not shown). Interestingly, the N10-14 LTR, which showed weak activity in the NTera2D1 and 293 cells, did not function to a significant level in the COS-1 cells. Further tests using the H6ᶠCAT(X) construct revealed promoter activity in a variety of cell types, including human primary bone marrow fibroblasts, mouse P-19 embryonal carcinoma cells, and 1 day fish (Orizias latipes) embryos (results not shown). Thus, the H6 LTR contains strong promoter sequences with a wide range of activity. The other LTRs tested here contain weaker promoters with more limited specificities.

This heterogeneity in promoter function may be attributed to sequence differences between the LTRs, since those tested here range from 75 to 93% identical to each other. A sequence comparison of the LTRs used in this study is shown in figure 3. Figure 3a compares the sequences of H6, N10-14, 3′R1, and 5′R2 to a consensus LTR generated from these LTRs plus eight other genomic LTRs (23). The PB-3 LTR is not included because

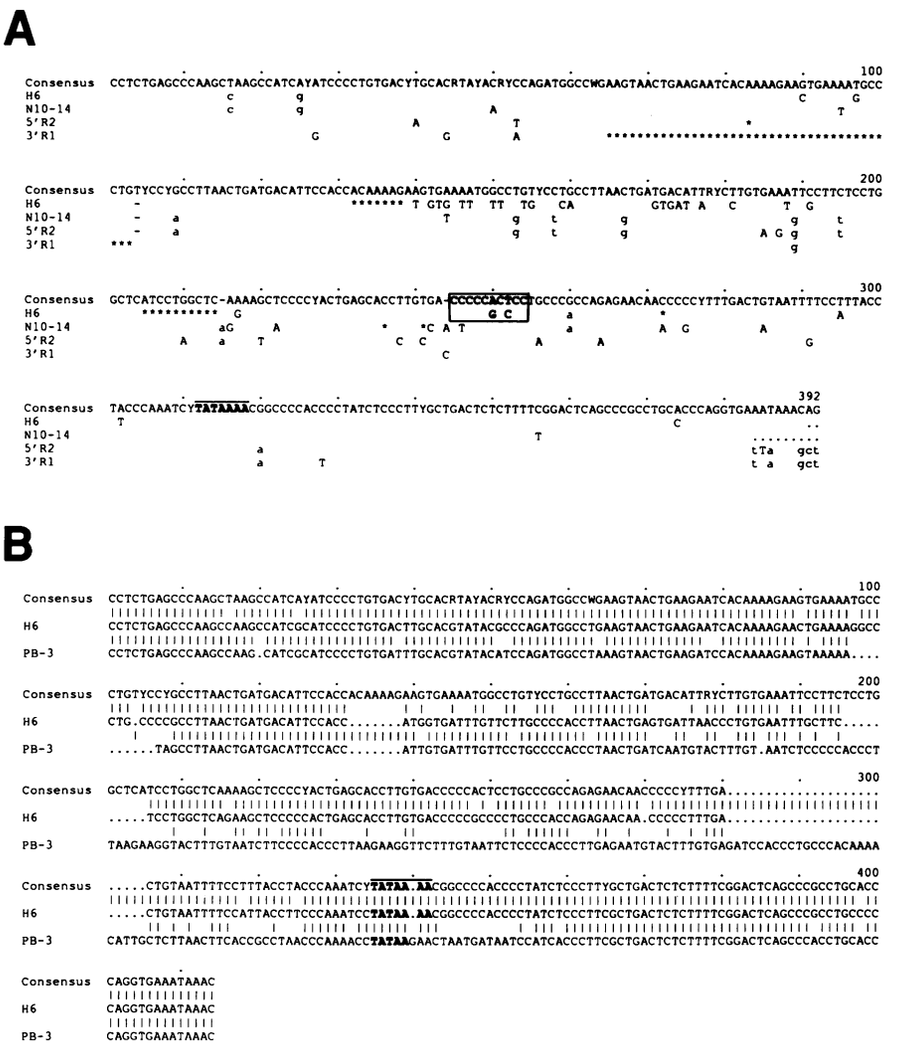

Figure 3. Sequence comparison of RTVL-H LTRs. a) The sequences of the H6, N10-14, 5'R2, and 3'R1 LTRs are compared to a consensus sequence generated from these LTRs plus eight other genomic LTRs (23). Capital letters indicate nucleotide differences, relative to the consensus, which occur only in one of the LTRs shown and lower case letters are nucleotide differences occurring in multiple RTVL-H LTRs. Asterisks represent deletions, relative to the consensus, which occur only in one of the LTRs shown and dashes indicate deletions which occur in multiple LTRs. The dots at the end of the LTRs indicate the extent of the LTR tested in this analysis. A potential SP1 binding site in the H6 LTR is boxed. b) The sequences of the PB-3 and H6 LTRs are compared to the consensus. Dots indicate length differences. The TATAA boxes are overlined. Abbreviations: Y, C or T; R, A or G; W, A or T. The sequences shown here have been reported elsewhere (16–17; 23).

its sequence differs substantially from the others (see below). As expected, figure 3a shows that several single nucleotide differences are scattered throughout the LTR sequences. Interestingly, the H6 LTR differed significantly in sequence from the consensus LTR between positions 130–220. This region contains several single nucleotide substitutions which are unique to H6, as well as two unique short deletions. Another difference of note in H6 occurs at positions 245–254. In this region, two nucleotide substitutions, which again are unique to H6, create a potential SP-1 binding site, CCCCGCCCC (31).

Unlike the 5'R2, and N10-14 LTRs, which also showed a more restricted range of activity, the PB-3 LTR was capable of relatively high levels of promoter activity in COS-1 cells. It is thus interesting to note that PB-3 is quite similar in sequence to H6 in a region where H6 differs from the other LTRs (see positions 132–160 in figure 3b). Downstream of this region, however, the PB-3 sequence differs completely from the other LTRs (see positions 170–350 in figure 3b). PB-3 has been designated a 'type II' RTVL-H LTR based on these extensive sequence differences (23).

Bidirectional Promoter Activity

In order to determine whether RTVL-H LTRs could promote transcription in the opposite orientation, we compared the activities of the PB-3 and H6 LTR^{r}CAT constructs with those of the LTR^{f}CAT(X) constructs in 293 and COS-1 cells. Figure 4 illustrates the results of one of these experiments. The H6^{r}CAT(X) construct functioned at moderate levels in both cell types. Furthermore, this construct also displayed significant activities in NTera2D1 cells but not in Ltk$^-$ cells (results not shown). Although the H6 'reverse promoter' activity showed similar cell specificity to that of the forward promoter, the relative strength of the reverse promoter was not always related to that of the forward promoter. The PB-3^{r}CAT(X) construct showed activity in COS-1 cells but not in 293 or in any other of the cell lines tested. Again, the reverse promoter had the same pattern

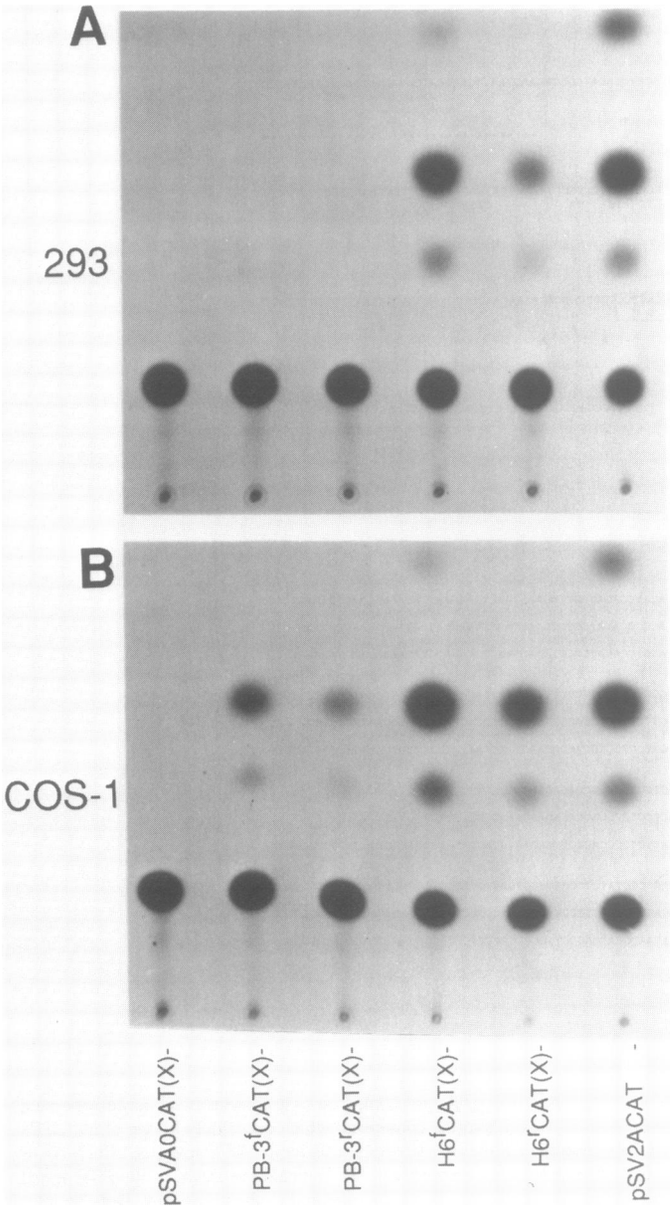

Figure 4. Bidirectional promoter activity of RTVL-H LTRs. CAT constructs containing the PB-3 or H6 LTRs in the forward (f) or reverse (r) orientation were transfected into the cell lines shown and assayed for CAT activity 44 h later as described in figure 2 and in Materials and Methods.

of cell specificity as the forward promoter. None of the other LTRs tested promoted measurable levels of CAT activity in the reverse orientation.

RNA Mapping Analysis

To determine whether the CAT transcripts were initiated within the LTR, primer extension experiments were performed to define transcriptional start sites. RNA from 293 cells, either mock transfected or transfected with the H6^{f}CAT(X) or H6^{r}CAT(X) constructs, was isolated and primer extension analysis was performed using an oligonucleotide primer complementary to RNA sequences downstream of the LTR/CAT junction. Results of this analysis are shown in Figure 5a. One major extended product was consistently observed using RNA from cells transfected with the H6^{f}CAT(X) construct. It maps to the expected initiation site in the LTR, 23−24 nucleotides downstream of the TATAA box (see figure 5b). Smaller bands, such as the one shown in figure 5a, were occasionally observed but were not consistent in size between experiments.

Several RNA start sites were found using the H6^{r}CAT(X) transfected RNA. These transcripts were mapped to a region between 30−50 nucleotides downstream of a GC rich segment which is preceded by a TTAA (figure 5b). The exact positions of these transcriptional start sites varied between experiments but were consistently located in the same region. This heterogeneity of start sites is not unexpected as the function of the TATAA box is primarily to correctly position the start of transcription (for review, see ref. 32). In cases where a TATAA box is altered or absent, decreases in initiation frequency and heterogeneity of start sites have been observed (33). Nevertheless, in both the forward and reverse H6-CAT(X) constructs, transcription has been initiated within the LTR.

RTVL-H LTRs Can Be Activated by SV40 Enhancer Sequences

As most of the LTRs tested displayed relatively weak promoter activity, we next decided to test whether transcription directed from RTVL-H LTRs could be augmented by heterologous enhancer sequences. The H6 and PB-3 LTRs were chosen for analysis in this respect because they differed significantly in both their sequence (see figure 3) and in their pattern of promoter function. These LTRs were inserted, in both orientations, into the vector pSVA0CAT(LR) (figure 1b), which contains no promoter sequences, but has an SV40 early enhancer region inserted downstream of the CAT gene. The promoter activity of these LTRs under the influence of the SV40 enhancer was compared to their activities in the absence of an enhancer in NTera2D1 and COS-1 cells (figure 6). The addition of SV40 enhancer sequences increased the activity of the H6 LTR in both orientations and in both cell types. Expression from the PB-3 LTR was significantly increased in COS-1 cells (figure 6b) but increased only to barely detectable levels in the NTera2D1 cells (figure 6a), or CV-1 cells (results not shown). Thus, the extent of induction of the PB-3 LTR promoter was dependent upon the cell in which it was tested. To determine whether this apparent restriction held for the other LTRs, we inserted the 5′R2 and N10-14 LTRs into pSVA0CAT(LR) and tested them in NTera2D1 and CV-1 cells. Interestingly, these LTRs could be induced to function at low levels in both orientations by the addition of the (LR) enhancer fragment (results not shown).

The H6 LTR Contains Enhancer Sequences

To determine whether RTVL-H LTRs contain a distinct enhancer activity, the H6 LTR was inserted, in both orientations, downstream of a human β-globin promoted CAT gene in the expression vector pSVAβGCAT(X) (figure 7a). As the human β-globin promoter is strongly enhancer dependent (23), any activity observed can be directly attributed to RTVL-H enhancer sequences. The vector pSVAβGCAT(LR) (figure 7a) which contains the SV40 early enhancer fragment, LR, was used as a positive control. The results of one experiment are shown in figure 7b and c. The enhancerless pSVAβGCAT(X) construct promoted only very low levels of CAT expression in NTera2D1 or 3T3 cells. Addition of H6 LTR sequences in either orientation, downstream of the CAT gene, activated the β-globin promoter to significant levels in both cell types. The SV40 enhanced control, pSVAβGCAT(LR), is shown for comparison.

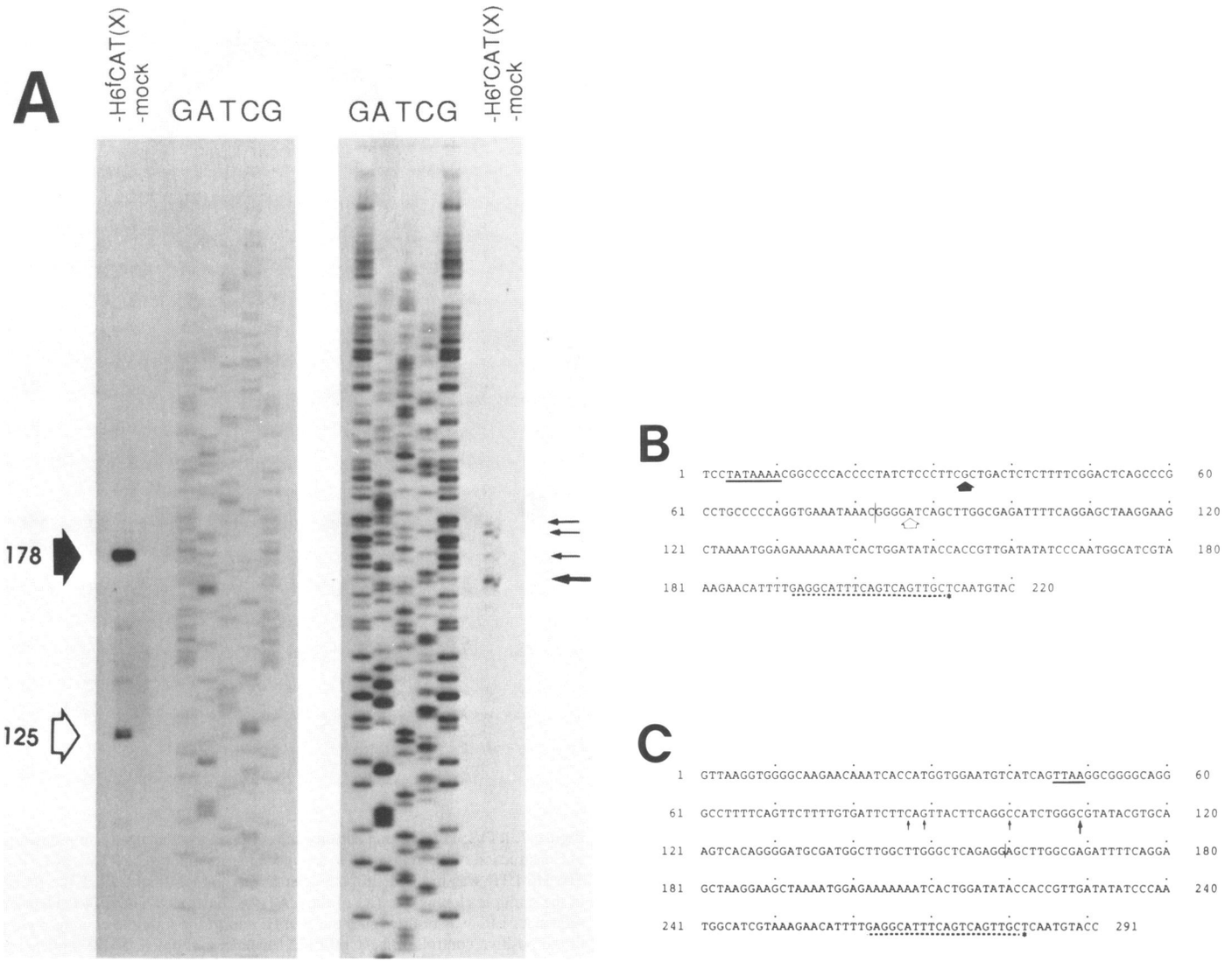

Figure 5. Primer extension analysis. (a) 293 cells, either transfected with H6^{f}CAT(X) or H6^{r}CAT(X), or mock transfected were hybridized with a 5′ end labeled 21 bp oligonucleotide primer derived from the 5′ end of the CAT gene. The hybrids were extended with reverse transcriptase and the cDNA products were electrophoresed on a sequencing gel alongside a sequencing reaction using the same primer. In the forward orientation, a major band (dark arrow) of 178 bp was observed in all experiments. Secondary products (e.g. white arrow, 125 bp) were occasionally observed, but their size varied between experiments. Several extended products were found in the H6^{r}CAT(X) transfected RNA and are indicated by arrows. The size of the arrow indicates the relative abundance of the transcript as determined by band intensity. (b,c) Mapping of transcription start sites. Results of the primer extension analysis were used to map the transcription start sites in the forward (b) and reverse (c) orientations. Start sites have been localized to the regions indicated by the arrows. The region homologous to the primer is shown by a dashed underline with an asterisk indicating the 5′ end label. A vertical line seperates LTR sequences from linker and CAT sequences.

DISCUSSION

Diversity of RTVL-H Promoters

In a parallel study on RTVL-H expression, our laboratory has shown that RTVL-H transcripts initiating in the 5′ LTR and terminating in the 3′ LTR are abundant in NTera2D1 cells and are also found in Hep2 cells and some other cell lines (Wilkinson et al., submitted). Of the primary tissues examined, little or no expression was detected in adult blood or fibroblast samples while amnion and chorion membranes from normal placenta had significant levels of expression (Wilkinson et al., submitted). These results indicate that endogenous tissue-specific expression of these sequences occurs and may be biologically relevant.

In this study we have examined the functional capacities of individual LTRs in an attempt to determine the range of promoter/enhancer functions that this large family of sequences may possess. Of the 5 LTRs tested, the H6 LTR derived from a Hep2 transcript promoted high levels of CAT gene expression in a wide variety of cell types. Figure 3 shows that the H6 LTR contains several nucleotide differences and two length differences that are not found in the other LTRs tested. Southern blot analysis using a specific subregion probe (spanning positions 139−219 in figure 3) indicates that LTRs closely related to H6 are present in 40−60 copies per haploid genome (unpublished observations). This finding indicates that a distinct subpopulation of RTVL-H LTRs with potentially strong promoter acitivity exists in the genome. Two of the nucleotide differences unique to H6 create a potential SP-1 binding site (boxed in figure 3a). An SP-1 binding

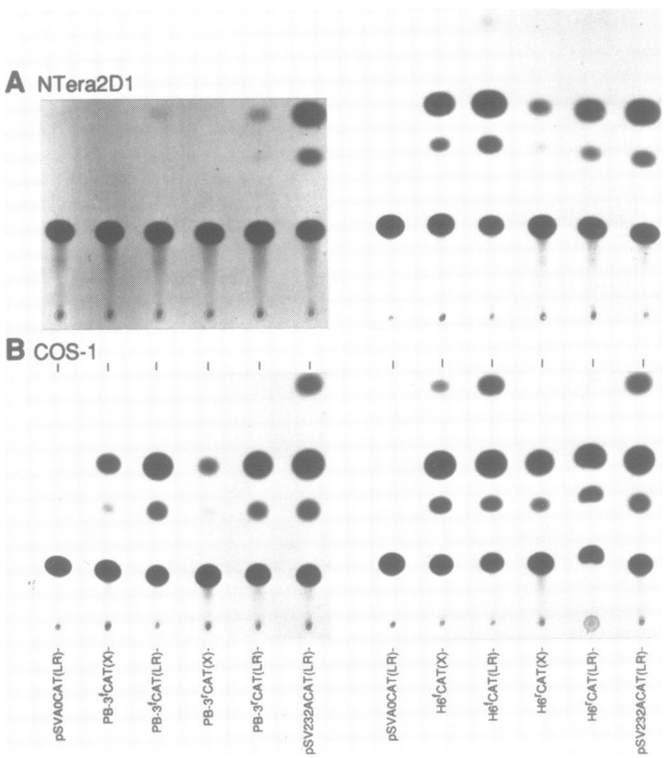

Figure 6. Activation of RTVL-H LTRs by the addition of a downstream enhancer. The activity of the PB-3 and H6 LTRs, in both orientations was compared to their activity in the presence of an SV40 enhancer in (a) NTera2D1 and (b) COS-1 cells.

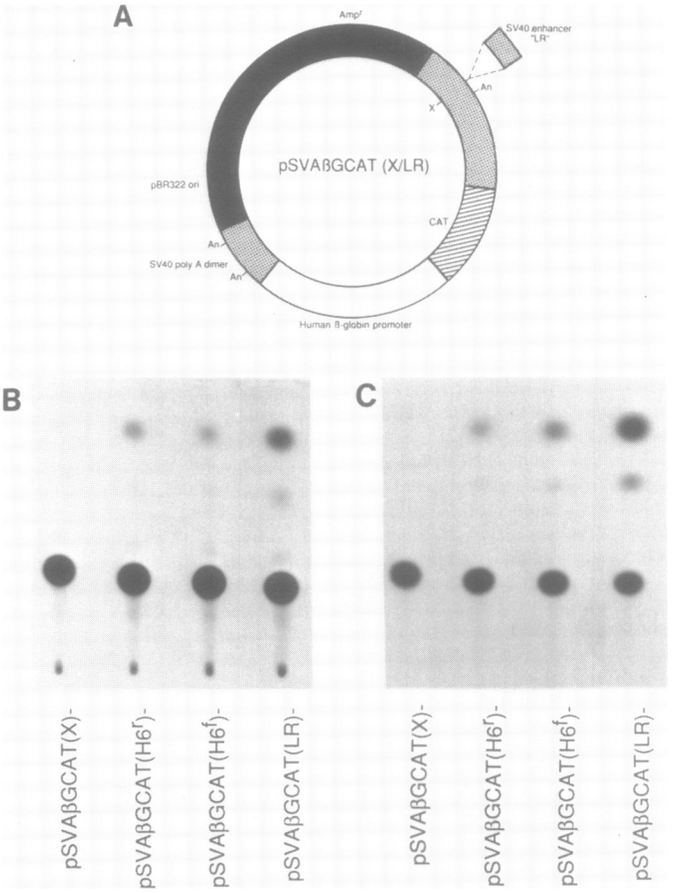

Figure 7. RTVL-H LTRs can enhance transcription from an upstream promoter. a) Construction of plasmids used to test the enhancer abilities of RTVL-H LTRs. The H6 LTR was inserted, in both orientations, into a unique XbaI site found in the multiple cloning site (X) of the CAT(X) constructs. The SV40 enhancer sequence, LR, was inserted into the BamHI site of the multiple cloning region in the positive controls. Black, pBR322 sequences; stippled, SV40 sequences; striped, CAT sequences; white, human β-globin promoter. Test constructs containing the H6 LTR, in both orientations were transfected into (b) NTera2D1 cells and (c) NIH 3T3 cells and assayed for CAT activity.

site is found in a similar location in the myelproliferative sarcoma virus LTR and may be important for expression of this retrovirus in the embryonal carcinoma line F9 (34). The possible importance of this and other H6-specific sequence differences in conferring increased promoter strength or range of activity to this LTR is currently under investigation.

Over the course of these investigations, we observed correlations between the cDNA library of origin and the cell types in which a particular LTR was active. For example, the N10-14 LTR was derived from an NTera2D1 library and was most active in these cells. The H6 LTR was derived from a Hep2 cDNA library and was the only LTR tested to function to a detectable level in this cell line. In addition, the two LTRs not derived from transcripts, 5′R2 and 3′R1, functioned very weakly if at all in the cell lines tested. Thus, although there could be factors affecting endogenous expression of these sequences which are not detectable using transient assays, the results presented here do appear to parallel the endogenous capabilities of different RTVL-H LTRs.

We have also observed that several of the LTRs tested were capable of bidirectional promoter activity, either alone (see figure 4) or when activated by heterologous enhancer sequences (figure 6). Interestingly, bidirectional promoter activity has also been reported for the LTRs of murine retrovirus-like intracisternal A particle (IAP) elements (35). The 'reverse' and forward RTVL-H LTR promoters showed similar cell specificities. This shared specificity may occur because both promoters share the same upstream regulatory elements. This finding suggests that some RTVL-H LTRs are able to promote the expression of upstream as well as downstream genes.

RTVL-H LTR Sequences Can Interact with Cellular Enhancers and Promoters

In this study, we have also found that expression of the CAT gene from the PB-3 and H6 LTR promoters can be augmented by the presence of SV40 early enhancer sequences. However, the degree of enhancement observed depended on the basal level of promoter activity. For example, in the presence of the SV40 early enhancer, the expression from the H6 LTR was raised from a level of 38% of the SV40 early promoter to a level approximately equal to the SV40 early promoter in NTera2D1 cells. Although expression from PB-3 was also increased by the addition of the LR enhancer fragment, it could not be increased to a level comparable to the SV40 early promoter. The 5′R2 and N10-14 LTRs could also be enhanced to low levels of activity in cells in which their basal activities were not detectable. This finding raises the possibility that genomic rearrangements which place a heterologous enhancer in the proximity of an RTVL-H LTR may activate the LTR. The extent of this activation would depend upon the particular LTR involved. The PB-3 and H6 LTRs both contain a TATAA box, however, the sequences surrounding this element (positions 170−325 and 344−355, in

figure 3b) are not closely related. Therefore, it may be that some as yet unrecognized elements which can affect promoter activity are involved.

We have also observed that the H6 LTR can enhance expression from the human β-globin promoter. Enhancers are typically modular elements made up of short, often repeated units. It has been shown that combining different units can result in new tissue specificities conferred by the enhancer (for example, see ref 36). Therefore, it is possible that if we had tested smaller regions of the LTR we may have observed differences in cell type restrictions. As we have not fragmented the H6 LTR it is likely that our results reflect the effect that an intact RTVL-H LTR of this type may have on a nearby cellular promoter.

Significance of These Findings

Several retrovirus-like families in mice have been implicated in both the inactivation and activation of cellular genes. For example, comparison of the κ light chain genes of two hybridoma cell lines defective in κ light chain gene synthesis with the wild type genes revealed that the mutant genes had IAP element insertions in their introns, suggesting that the insertions caused gene inactivation (5). IAP elements have also been implicated in gene activation events. An interesting example occurs in the myelomonocytic murine leukemia cell line WEHI-3B. This line constitutively produces the growth factor Interleukin-3 due to an IAP insertion 5′ of the gene (9) and expresses the Hox2.4 homeobox gene due to an IAP insertion in the first exon of this gene (10, 37). In both cases, transcription has been shown to originate within the LTR. Therefore, insertions of endogenous retrovirus-like elements have altered the pattern of expression of these genes and may have contributed to the transformation of this cell line.

While H6 promoted high levels of CAT expression, the other LTRs tested in this study displayed a much more limited range of activity and much weaker levels of promoter strength. That most of the LTRs tested contained relatively weak promoter sequences is not unexpected. The human genome contains between 2000–3000 RTVL-H LTRs in addition to several thousand LTR- like sequences belonging to different families (16, 38, 39). If many of these sequences had strong promoter capabilities they would probably require cellular regulation so as not to interfere with normal cellular processes. It may, however, confer a selective advantage to an organism to have a variety of weak promoters in the genome which could be activated by rearrangements or other genetic changes to impose new patterns of regulation on cellular genes. Support for this hypothesis comes from the identification of two rodent genomic LTR insertions which have evolved into a regulatory role. Transcription of the mouse sex-limited protein gene is androgen dependent due to the enhancer donation of an upstream endogenous proviral insertion (40). It has also been shown that the oncomodulin gene of the rat is promoted by a solitary LTR related to IAP elements (41). We are currently screening the human genome for similar events involving RTVL-H LTRs.

Here, we have demonstrated that RTVL-H LTRs are functionally diverse and can promote the expression of a linked gene as well as interact with cellular promoters and enhancers. Although transposition of RTVL-H elements has not yet been demonstrated, a rearrangement involving an existing RTVL-H proviral element has been detected (42). These findings raise the possibility that rearrangements or activation of these endogenous sequences may contribute to alterations in human gene expression.

ACKNOWLEDGEMENTS

We thank Doug Freeman for invaluable assistance in constructing the LTR plasmids and for DNA sequencing. We also thank Dr. Keith Humphries for advice and helpful discussions during the course of this work and Dr. Paula Henthorn for providing the plasmid vectors used here. We are grateful to Steve Bain, Ferguson Neville, and Gregg Richardson for artwork and photography. This work was supported by a grant from the British Columbia Health Care Research Foundation. A.F. is a Research Student of the National Cancer Institute of Canada. D.M. is a Research Scholar of the Medical Research Council of Canada.

REFERENCES

1. Cohen, M., and Larsson, E. (1988) *Bioessays* **9**, 191–196.
2. Steele, P.E., Rabson, A.B., Bryan, T., and Martin, M.A. (1984) *Science* **225**, 943–947.
3. O'Connell, C., and Cohen, M. (1984) *Science* **226**, 1204–1206.
4. Ono, M., Yasunaga, T., Miyata, T., and Ushikubo, H. (1986) *J. Virol.* **60**, 589–598.
5. Hawley, R.G., Shulman, M.J., Murialdo, H., Gibson, D.M., and Hozumi, N. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7425–7429.
6. Min Man, Y., Delius, H., and Leader, D.P. (1987) *Nucleic Acids Res.* **15**, 3291–3304.
7. Cohen, J.B., Ungar, T., Rechavi, G., Canaani, E., and Givol, D. (1983) *Nature* **306**, 797–799.
8. Gattoni-Celli, S., Hsiao, W-W., and Weinstein, I.B. (1983) *Nature* **306**, 795–796.
9. Ymer, S., Tucker, W.Q.J., Sanderson, C.J., Hapel, A.J., Campbell, H.D., and Young, I.G. (1985) *Nature* **317**, 255–258.
10. Kongsuwan, K., Allen, J., and Adams, J.M. (1989) *Nucleic Acids Res.* **17**, 1881–1892.
11. Temin, H.M. (1981) *Cell* **27**, 1–3.
12. Varmus, H.E. (1982) *Science* **216**, 812–820.
13. Chen, H.R., and Barker, W.C. (1984) *Nucleic Acids Res.* **12**, 1767–1778.
14. Kingsman, A.J., and Kingsman, S.M. (1988) *Cell* **53**, 333–335.
15. Finnegan, D.J. (1985) *Int. Rev. Cytol.* **93**, 281–326.
16. Mager, D.L., and Henthorn, P.S. (1984) *Proc. Natl. Acad. Sci.* USA **81**, 7510–7514.
17. Mager, D.L., and Freeman, J.D. (1987) *J. Virol.* **61**, 4060–4066.
18. Adams, S.E., Rathjen, P.D., Stanway, C.A., Fulton, S.M., Malim, M.H., Wilson, W., Ogden, J., King, L., Kingsman, S.M., and Kingsman, A.J. (1988) *Mol. Cell. Biol.* **8**, 2989–2998.
19. Johansen, T., Holm, T., and Bjorklid, E. (1989) *Gene* **79**, 259–267.
20. Lavappa, K.S. (1978) *In Vitro* **14**, 469–475.
21. Andrews, P.W., Damjanov, I., Simon, D., Banting, G.S., Carlin, C., Dracopoli, N.C., and Fogh, J. (1984) *Lab. Invest.* **50**, 147–162.
22. Graham, F., and van der Eb, A. (1973) *Virology* **52**, 456–457.
23. Mager, D.L. (1989) *Virology* **173**, 591–599.
24. Kadesch, T., and Berg, P. (1986) *Mol. Cell. Biol.* **6**, 2593–2601.
25. Kadesch, T., Zervos, P., and Ruezinsky, D. (1986) *Nucleic Acids Res.* **14**, 8209–8221.
26. Henthorn, P., Zervos, P., Raducha, M., Harris, H., and Kadesch, T. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6342–6346.
27. Gorman, C.M., Moffat, L.F., and Howard, B.H. (1982) *Mol. Cell. Biol.* **2**, 1044–1051.
28. Davis, L.G., Dibner, M.D., and Battey, J.F. (1986) In: Basic Methods in Molecular Biology. New York, New York: Elsevier.
29. Paulson, K.E., Matera, A.G., Deka, N., and Schmid, C.W. (1987) *Nucleic Acids Res.* **15**, 5199–5215.
30. Tabor, S., and Richardson, C.C. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4767–4771.
31. Dynan, W.S., and Tjian, R. (1983) *Cell* **35**, 79–87.
32. Maniatis, T., Goodbourn, S., and Fischer, J.A. (1987) *Science* **236**, 1237–1244.
33. Grosveld, G.C., de Boer, E., Shewmaker, C.K., and Flavell, R.A. (1982) *Nature* **295**, 120–126.
34. Hilberg, F., Stocking, C., Ostertag, W., and Grez, M. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 5232–5236.
35. Christy, R.J., and Huang, R.C.C. (1988) *Mol. Cell. Biol.* **8**, 1093–1102.
36. Rochford, R., Campbell, B.A., and Villarreal, L.P. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 449–453.

37. Blatt, C., Aberdam, D., Schwartz, R., and Sachs, L. (1988) *EMBO J.* **7**, 4283–4290.
38. Paulson, K.E., Deka, N., Schmid, C.W., Misra, R., Schindler, C.W., Rush, M.G., Kadyk, L., and Leinwand, L. (1985) *Nature* **316**, 359–361.
39. Larsson, E., Kato, N., and Cohen, M. (1989) *Curr. Top. Microbiol. Immunol.* **148**, 115–132.
40. Stavenhagen, J.B., and Robbins, D.M. (1988) *Cell* **55**, 247–254.
41. Banville, D., and Boie, Y. (1989) *J. Mol. Biol.* **207**, 481–490.
42. Mager, D.L., and Goodchild, N.L. (1989) *Am. J. Hum. Genet.* **45**, 848–854.