JOURNAL OF VIROLOGY, Jan. 1996, p. 188–198 Vol. 70, No. 1
0022-538X/96/$04.00+0

The Promoter Activity of Long Terminal Repeats of the HERV-H Family of Human Retrovirus-Like Elements Is Critically Dependent on Sp1 Family Proteins Interacting with a GC/GT Box Located Immediately 3′ to the TATA Box

EVA SJØTTEM, SØLVI ANDERSSEN, AND TERJE JOHANSEN*

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

Received 26 June 1995/Accepted 26 September 1995

The HERV-H family of endogenous retrovirus-like elements is widely distributed in the human genome, with about 1,000 full-length elements and a similar number of solitary long terminal repeats (LTRs). HERV-H LTRs have been shown to direct the transcription of both HERV-H-encoded and adjacent cellular genes. Transcripts of HERV-H elements are especially abundant in placenta, teratocarcinoma cell lines, and cell lines derived from testicular and lung tumors. Here we report that only a subset of HERV-H LTRs display promoter activity in human cell lines and that these LTRs are characterized by the presence of a GC/GT box immediately downstream of the TATA box. This GC/GT box is required for promoter activity, while, surprisingly, the TATA box is dispensable. The ubiquitously expressed transcription factors Sp1 and Sp3 bound to this GC/GT box and stimulated transcription from the promoter-active LTRs in the teratocarcinoma cell line NTera2-D1. However, in HeLa and *Drosophila* SL-2 cells, Sp1 acted as a transcriptional activator of the LTRs, while Sp3 acted as a repressor of Sp1-mediated transcriptional activation. Cotransfection studies also revealed that the tissue-specific Sp1-related protein BTEB bound to this GC/GT box and stimulated transcription from the LTR promoters in NTera2-D1 cells. These results show that members of the Sp1 protein family are crucial determinants for transcriptional activation of HERV-H LTR promoters and suggest that these proteins may also be involved in determining the tissue-specific expression pattern of HERV-H elements.

During the past decade, several families of human endogenous retrovirus-like elements (HERVs) have been identified and partly characterized (for a recent review, see reference 61). HERV sequences are estimated to constitute at least 0.6% of the human genome. Their considerable number and random distribution suggest that they have been amplified primarily through repeated chromosomal insertions of reverse transcripts. Elements in the HERV-H (previously called RTVL-H) family constitute one of the most abundant HERV families, with about 1,000 full-length elements and a similar number of solitary long terminal repeats (LTRs) (45); the latter are most probably generated by homologous recombination between the LTRs (44). HERV-H-derived transcripts have been observed in a variety of cell lines, with the highest expression levels found in teratocarcinoma cell lines and other carcinoma cell lines derived from testicular and lung tumors (22, 59–61). In normal tissue, expression was found in placenta (29, 59) and at low levels in peripheral blood mononuclear cells (47). Most of the HERV-H elements examined so far contain nonfunctional genes, but a few elements contain open reading frames in both the *pol* and *env* genes (22, 60). So far, the presence of HERV-H-encoded proteins has not been reported, but reverse transcriptase activity and retrovirus-like antigens and particles have been detected in uninfected human tissues and cell lines, suggesting that HERV-H proteins and other HERV proteins may be expressed (5, 48, 58, 61). In addition, evidence indicating that HERVs encode some of the retrovirus-like particles (4, 38, 39) and retrovirus-related proteins (40, 52) found in uninfected cells is emerging. It has been speculated that such proteins may play a role in the initiation of autoimmune diseases (reviewed in reference 34) and in immunosuppression and defense against superinfection by related exogenous retroviruses (36). At least for some LTRs, transient transfections have revealed that their promoter sequences are able to drive the transcription of a reporter gene in different cell lines (13) and to influence the activity of a cellular promoter (13). In addition, transcription of cellular genes initiated in an HERV-H LTR has been reported (12, 14, 37). The evolution of LTRs into promoter or enhancer sequences for cellular genes has also been reported for other HERVs and mouse ERVs (61). In some cases these LTRs have contributed with regulatory elements to result in tissue-specific expression of the cellular genes. Also, the polyadenylation signal located in the R regions of the HERV-H LTRs has been found to direct 3′ end processing of both HERV-H transcripts and adjacent cellular genes (17, 29, 43). Thus, the regulatory sequences of endogenous LTRs do affect the expression of cellular genes, and rearrangement of these elements brought about either by homologous recombination (44) or by transposition (16) may influence the expression pattern of host genes. These findings stress the importance of studying various HERV-H LTRs for their performance as transcriptional promoters.

The constitutively expressed transcription factor Sp1 (9) binds to GC-rich sequences present in a variety of cellular and viral promoters and stimulates their transcriptional activity. Recently, three other ubiquitously expressed GC and/or GT box-binding factors homologous to Sp1, named Sp2, Sp3, and Sp4 (21, 33), and two more distantly related GC/GT box-binding factors, BTEB and BTEB2 (27, 55), have been cloned. Sp4 is reported to be a transcriptional activator comparable to

* Corresponding author. Phone: 47 776 44720. Fax: 47 776 45350. Electronic mail address: terjej@fagmed.uit.no.

Sp1, while Sp3 was found to act as a repressor of both Sp1-mediated transcriptional activation and the human immunodeficiency virus type 1 (HIV-1) LTR promoter (20, 46). Thus, Sp1 family proteins have the potential to regulate a wide variety of cellular and viral genes in different tissues.

To begin elucidating the transcriptional regulation of HERV-H LTR promoters, we have isolated different LTRs from human chromosome 18 and assayed them for promoter activity in human HeLa, JEG-3, and NTera2-D1 cells (1). Here we report that the HERV-H LTRs capable of driving the expression of a reporter gene contain either a GC or a GT box located immediately downstream of the TATA box. Mutational analysis showed that this GC/GT box, probably together with a putative Inr element at the transcription initiation site, was both necessary and sufficient for the transcriptional activity of the promoters. Surprisingly, point mutation of the TATA box did not significantly reduce the transcriptional activity of the LTR promoter, indicating that the LTR elements are able to act as TATA-independent promoters. We found that both Sp1 and the related Sp3 bound to the GC/GT box. As expected, Sp1 stimulated the transcription from the promoter-active LTRs in *Drosophila* SL-2 cells, HeLa cells, and the teratocarcinoma-derived cell line NTera2-D1. Sp3, however, repressed Sp1-mediated transcriptional activation in the *Drosophila* and HeLa cells, while it acted as a transcriptional activator comparable to Sp1 in the NTera2-D1 cells. These results show that members of the Sp1 protein family of transcription factors represent critical determinants for the promoter activity of HERV-H LTRs.

MATERIALS AND METHODS

Plasmid constructions. The LTRs were isolated from human chromosome 18 and inserted into pCR1000 (Invitrogen) as described elsewhere (1). The chloramphenicol acetyltransferase (CAT) reporter gene plasmids pBLCAT3 and pBLCAT2 (41) were used for assays of promoter and enhancer activities, respectively. For promoter activity analyses, LTRs 18102 and 18103 were released from pCR1000 by digestion with *Eco*RI and *Spe*I, the *Eco*RI end was made blunt with T4 DNA polymerase, and the LTR fragments were inserted into the *Hin*dIII (end-filled) and *Xba*I sites upstream of the promoterless CAT gene of pBLCAT3. LTRs 18107 and 18321 were cloned as *Eco*RI (end-filled)-*Hin*dIII fragments into *Sal*I (end-filled)-*Hin*dIII-cut pBLCAT3. Similarly, LTR 18316 was inserted as an *Eco*RI (end-filled)-*Spe*I fragment into *Xho*I (end-filled)-*Xba*I-cut pBLCAT3. For analyses of enhancer activity with the complete LTR sequences, *Hin*dIII-*Eco*RI fragments of the LTRs made blunt ended by T4 DNA polymerase were ligated into the *Sma*I site downstream of the CAT gene in pBLCAT2. For enhancer assays of the U3 region upstream of the TATA box, PCR fragments generated with primers P1 (5′-TGTCAGGCCTCTGAGC CCAAGC-3′) and P4 (5′-GATTTGGGTAGATAAAGAA-3′) (Fig. 1), 5′ phosphorylated with T4 polynucleotide kinase and made blunt ended with Klenow polymerase, were inserted into the *Sma*I site of pBLCAT2.

The p18321-Tm plasmid, containing LTR 18321 with a mutated TATA box inserted upstream of the CAT gene in pBLCAT3, and the p18321-Sp1m plasmid, containing LTR 18321 with a mutated Sp1 binding site, were generated by using the Altered Sites II in vitro mutagenesis system (Promega). The TATA box sequence (5′-TATAAAAT-3′) was changed to (5′-TCTAGAGT-3′), introducing an *Xba*I restriction site. The mutation was verified by both *Xba*I digestion and sequencing. The Sp1 binding sequence (5′-CCCCGCCCCT-3′) was changed to 5′-CCAAGCCCCT-3′, and the mutation was verified by sequencing.

The Sp1 expression vector pPac-Sp1 (7) was provided by A. Gègonne (15), after kind permission from R. Tjian. The mammalian expression vectors pRS VSp1 and pRSVBTEB were kindly provided by Y. Fujii-Kuriyama (27, 55), while the mammalian expression vectors pCMV-Sp1 and pCMV-Sp3 and the *Drosophila* expression vector pPac-Sp3 were generous gifts from G. Suske (20, 21).

Cell culture. NTera2-D1 cells, obtained from P. Andrews (2, 3), were cultured in Dulbecco's modified Eagle's medium with a high glucose concentration (4.5 g/liter) supplemented with fetal calf serum (10%), penicillin (100 U/ml), and streptomycin (100 μg/ml) in a CO_2 incubator (5% CO_2) at 37°C. To induce differentiation NTera2-D1 cells were grown for 4 days in the presence of 10 μM all-*trans*-retinoic acid (Sigma; catalog no. R2625). JEG-3 cells (ATCC HTB 36) were grown in Eagle's minimum essential medium supplemented with 10% fetal calf serum. HeLa cells (ATCC CCL 2) were grown in the same medium except for the addition of nonessential amino acids and 2 mM L-glutamine. CV-1 (ATCC CCL 70) and COS-1 (ATCC CRL 1650) cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum. *Drosophila* SL-2 cells, kindly provided by Anne Gègonne (15), were maintained in Schneider Insect Medium (Sigma; catalog no. S9895) supplemented with sodium bicarbonate (0.4 g/liter), calcium chloride (0.60 g/liter), L-glutamine (2 mM), streptomycin (100 μg/ml), penicillin (100 U/ml), and fetal calf serum (10%).

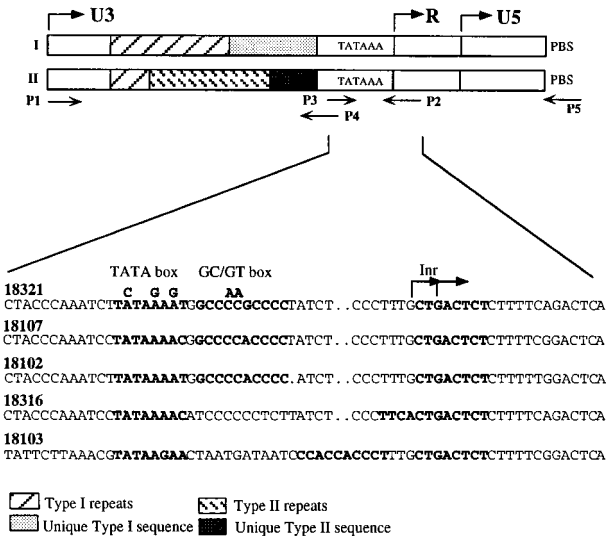

FIG. 1. Structures of HERV-H LTRs. At the top is a schematic drawing showing the structures of type I (I) and type II (II) LTRs. The 120- to 130-bp 5′ part of the U3 region is conserved between different HERV-H LTR subclasses. The remaining part of the U3 region upstream of the TATA box contains tandemly repeated sequences defining the different subclasses. The length of this region varies depending on the number of repeats. In the 3′ part of the U3 region, the TATA box is often immediately followed by a GC or GT box. The border between the U3 and R region defines the transcriptional initiation site and also the localization of a putative Inr element. For the five LTRs used in this study, the sequences of the core promoter region spanning the TATA box, the adjacent GC/GT box, and the transcription initiation site are shown aligned to each other below the drawing. Dots indicate gaps introduced to optimize the alignments. Only one of the LTRs, 18321, contains a consensus GC box. LTRs 18107, 18102, and 18103 contain a GT box, but for LTR 18103 the GT box is located 8 bp further downstream. LTR 18316 does not contain a GC or a GT box but instead contains a C-rich region at this position. The arrows labelled P1, P2, P3, P4, and P5 indicate the locations of primers used in PCR to generate probes for cloning, GMSA, and DNase I footprinting assays. The TATA boxes, GC/GT boxes, and putative Inr elements are indicated in boldface. Two arrows indicate the transcription initiation site. The 5′ one is the start site according to Feuchter and Mager (13), which also is within the first Inr elements for those LTRs containing two adjacent putative Inr elements, such as LTR 18316. The 3′ one is the expected start position if the putative Inr element is functional. Point mutations introduced to disrupt the TATA and GC boxes of the 18321 LTR are shown above the sequence. The extents of the U3, R, and U5 regions are denoted with arrows. PBS, primer binding site for $tRNA^{His}$.

Preparation of nuclear extracts. Nuclear extracts were prepared mainly as described by Lee et al. (35). Approximately 5×10^7 cells were harvested by scraping, centrifuged for 5 min at 790 × *g* in a microcentrifuge, and washed twice with phosphate-buffered saline (PBS). The cell pellet was resuspended in one packed-cell volume of hypotonic lysis buffer (10 mM Tris-HCl [pH 7.9], 10 mM KCl, 1.5 mM $MgCl_2$, 1 mM dithiothreitol [DTT], 0.2 mM phenylmethylsulfonyl fluoride [PMSF], 2 μg of aprotinin per ml, and 0.5 μg of leupeptin per ml) and allowed to swell for 15 min. The cells were lysed by using a 2-ml syringe to rapidly push them 5 to 10 times through a 25-gauge needle. The nuclei were pelleted by centrifugation for 20 s at 15,000 × *g* and 4°C in a microcentrifuge. The nuclear pellet was resuspended in two-thirds of a packed-cell volume of extraction buffer (50 mM Tris-HCl [pH 7.9], 0.42 M KCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 20% glycerol, 10% sucrose, 1 mM DTT, 0.2 mM PMSF, 2 μg of aprotinin per ml, and 0.5 μg of leupeptin per ml), and the nuclei were extracted for 30 min with continuous gentle mixing. The nuclear debris was pelleted by spinning for 5 min at 15,000 × *g* and the supernatant was dialyzed with 0.1 M TM (50 mM Tris-HCl [pH 7.9], 0.1 M KCl, 12.5 mM $MgCl_2$, 1 mM EDTA, 20% glycerol, 1 mM DTT, 0.2 mM PMSF, 2 μg of aprotinin per ml, and 0.5 μg of leupeptin per ml) for 2 to 3 h. The dialyzed extract was centrifuged for 5 min at 15,000 × *g* and 4°C and stored at −70°C. The protein concentrations in the extracts were determined by the method of Bradford (Bio-Rad) with bovine serum albumin (BSA) as the standard.

Gel mobility shift assays. (GMSA). The LTR probes spanning the U3 region and parts of the R region were generated by PCR with primers P1 and P2 (5′-TGAGTCC/TGAAAAGAGAGTCAG-3′). The primers were ^{32}P end labelled in a 15-μl reaction mixture (10 μM primers P1 and P2, 20 mM Tris-HCl [pH 7.75], 10 mM $MgCl_2$, 1 mg of BSA per ml, 10 mM 2-mercaptoethanol, 20 to 30 μCi of [γ-^{32}P]ATP, and 1 U of T4 polynucleotide kinase). After incubation at 37°C for 45 min, 5 μl of 10× PCR buffer (0.5 M KCl, 0.1 M Tris-HCl [pH 8.4], 15 mM $MgCl_2$, and 0.01% gelatin), 4 μl of deoxynucleoside triphosphate mix (10 mM), 500 ng of template, 2.5 U of *Taq* DNA polymerase, and distilled water to a final volume of 50 μl were added. PCR was carried out with an MJ Research MiniCycler with initial denaturation at 94°C for 1 min; 30 cycles with denaturation at 94°C for 20 s, annealing at 55°C for 20 s, and elongation at 72°C for 1 min; and a final extension at 72°C for 1 min. The probes were purified on a 6% polyacrylamide gel and eluted overnight in 400 μl of STE buffer (0.1 M NaCl, 20 mM Tris-HCl [pH 7.5], 1 mM EDTA).

The 64- to 65-bp LTR probes encompassing the TATA box and the proximal Sp1 binding site were prepared by the same procedure, except that primers P3 (5′-GATTTGGGTAGG/ATAAAGGAA-3′) and P2 were used for PCR. Protein-DNA interaction studies were performed with a total volume of 20 μl of 20 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid)-KOH (pH 7.5)–0.5 mM EDTA–0.5 mM DTT–0.05% Nonidet P-40–8% Ficoll–100 mM NaCl–2 μg of nonspecific competitor oligonucleotide. If not otherwise indicated, poly(dI-dC) was used as a nonspecific competitor. The binding reaction mixtures contained, if not otherwise indicated, 1 μg of nuclear extract and 10,000 to 20,000 cpm of labelled probed and were incubated on ice for 15 min. When specific competitor DNA or polyclonal Sp1 or Sp3 antibodies (20) were used, these reagents were added immediately prior to addition of the labelled probe. The following oligonucleotides containing consensus binding sites for specific transcription factors were used: Sp1 (5′-ATTCGATCGGGGCGGGGCGAG-3′), AP2 (5′-GATCGAACTGACCGCCCGCGGCCCGT-3′), AP3 (5′-GATGTGT GGAAAGTCCCAGTAG-3′), and NF-κB (5′-AGTTGAGGGGACTTTCCCA GGC-3′). DNA-protein complexes were resolved on a 4% polyacrylamide gel (39:1) in 1× TBE (89 mM Tris, 89 mM boric acid, and 2 mM EDTA) at 230 V and 4°C for 2 to 3.5 h, depending on the length of the probe. Gels were dried and subjected to autoradiography at −70°C with intensifying screens and/or analyzed on a PhosphorImager (Molecular Dynamics).

DNase I footprinting. Labelled probes for DNase I footprinting studies were generated by PCR as described for the gel shift assay, except that only one primer was ^{32}P labelled. Primers P1, P2, and P5 (5′-CGATCCGAGTCACGGCAC CAA-3′) were used. Labelled probes (8×10^4 cpm) were added to a 50-μl binding reaction mixture containing 10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 2.5 mM $MgCl_2$, 0.5 mM DTT, 4% glycerol, 2 μg of poly(dI-dC), and 8 to 32 μg of nuclear extract. Following a 20-min incubation at room temperature, $MgCl_2$ and $CaCl_2$ were added to final concentrations of 1 and 0.5 mM, respectively, together with 7 U of DNase I, and the probes were digested for 90 s. The DNase I digestion was stopped by adding 100 μl of phenol-chloroform-isoamyl alcohol (24:24:1), 50 μl of distilled water, and 1.5 μl of 0.5 M EDTA. The digested probes were extracted with phenol-chloroform (1:1), precipitated with ethanol with glycogen as a carrier, and finally resuspended in 3 μl of loading buffer (deionized formamide containing 10 mM EDTA, 0.3% bromophenol blue, and 0.3% xylene cyanol). The complete reaction mixture was run on a 5% polyacrylamide–7 M urea gel in 1× TBE. The gel was fixed in 10% acetic acid for 20 min, dried, and subjected to autoradiography for 3 to 5 days. The Maxam-Gilbert sequencing reaction was performed as described previously (51).

Transient-transfection assays. One day before transfection, *Drosophila melanogaster* Schneider SL-2 cells were seeded at 10^7 cells per 100-mm-diameter petri dish. The cells were transfected by the calcium phosphate coprecipitation method as described by Di Nocera and Dawid (8). The cells received 5 μg of CAT reporter plasmids and various amounts of Sp1 and Sp3 expression plasmid (10 ng to 5 μg). Salmon sperm DNA was used to bring the total DNA to 10 μg on each plate. After addition of DNA, the plates were left undisturbed until the time of harvest 40 to 48 h later. For preparation of nuclear extracts, eight 100-mm-diameter dishes of Schneider cells were transfected with 5 μg of Sp1 or Sp3 expression plasmid per plate. Forty to forty-eight hours later, the cells were harvested and nuclear extracts were prepared as described above. For CAT assays cells were harvested by scraping and suspended in 200 μl of buffer A (15 mM Tris-HCl [pH 8.0], 60 mM KCl, 15 mM NaCl, 2 mM EDTA, 0.15 mM spermine tetrahydrochloride, 1 mM DTT, 0.4 mM PMSF). The cells were lysed by three cycles of freezing and thawing, and cell debris was pelleted by centrifugation for 15 min at 15,000 × *g* in a microcentrifuge. The supernatant was heated to 70°C for 10 min, and denaturated proteins were pelleted by centrifugation as described above. The protein concentrations in the extracts were determined by the Bradford assay (Bio-Rad) with BSA as the standard. All transfections were carried out in triplicate in at least three independent experiments using different preparations of the plasmids. CAT assays were performed with 15 μg of protein extract, 25 μCi of *d-threo*-dichloroacetyl-[1,2-^{14}C]chloramphenicol per ml, 5 mg of *n*-butyryl coenzyme A (Sigma; catalog no. B-1508) per ml, and 0.25 M Tris-HCl (pH 8.0) in a 125-μl reaction volume. Following incubation for 1 h at 37°C, acetylated chloramphenicol was extracted in 300 μl of mixed xylene and washed twice with 100 μl of Tris-HCl (pH 8.0). The radioactivity in 200 μl of the xylene phase was quantitated by liquid scintillation counting (53). NTera2-D1 cells were seeded at a density of 5×10^6 cells per plate 2 days before transfection. The cells were transfected as described above except that a 70-s glycerol shock with 15% glycerol was performed 4 to 5 h after the addition of DNA. Five micrograms of the CAT reporter plasmid and, in the cotransfection analyses, various amounts of the pRSVSp1, pRSVBTEB, pCMV-Sp1, and pCMV-Sp3 expression vectors were added per 100-mm-diameter petri dish. Empty pRC-CMV vector or empty pRSV(XhoI) vector, in addition to salmon sperm DNA, was used to bring the total DNA to 12 μg on each plate. The cells were harvested by scraping 40 to 48 h after transfection, washed in PBS, and resuspended in 200 μl of buffer A. The cells were broken and the CAT assay was performed as described for the Schneider SL-2 cells. HeLa cells were transfected as described for NTera2-D1 cells, except that 2×10^5 cells per plate were seeded 2 days before transfections and the glycerol shock lasted for 90 s.

RESULTS

Different nuclear proteins bind to the U3 region of HERV-H LTRs. HERV-H LTRs can be grouped into three subtypes (1, 18). These subtypes are closely related in the first 120 to 130 bp of the U3 region and in the R and U5 regions. The remainder of the U3 region contains tandemly repeated sequences characteristic of each subtype, in addition to a subtype-specific region upstream of the TATA box (1, 18) (Fig. 1). We recently isolated eight different HERV-H LTRs from human chromosome 18 that belong to subtypes I (five elements) and II (three elements). In addition to the sequence diversity, we found the LTRs also to be functionally different with regard to promoter activity, as assayed by transfection experiments with the LTRs inserted upstream of the CAT gene in pBLCAT3 (1). Three of the chromosome 18 LTRs belonging to subtype I showed different, but relatively strong, promoter activities in undifferentiated and retinoic acid-stimulated NTera2-D1 cells and in the placental choriocarcinoma cell line JEG-3 and a weak promoter activity in HeLa cells. The remaining five chromosome 18 LTRs, however, showed very weak or no promoter activity in the cell lines tested. In order to determine which *cis*- and *trans*-acting factors were responsible for this differential functionality of the LTRs, we focused on the three promoter-active LTRs (18102, 18107, and 18321) and two of the promoter-inactive LTRs (18103 and 18316) (Fig. 1). GMSA performed with nuclear extracts from the human cell lines HeLa, NTera2-D1, and JEG-3, from retinoic acid-treated NTera2-D1 cells, and from the African green monkey cell lines CV-1 and COS-1 (data not shown) revealed two to four retarded bands for each LTR analyzed. The number and intensity of the protein-DNA complexes varied for the different LTRs. The promoter-active LTRs 18102 and 18321 displayed two to three high-intensity bands and some minor bands, while the third active LTR, 18107, showed only low-intensity bands. For the two promoter-inactive LTRs, 18103 and 18316, only one high-intensity band and some very weak protein-DNA complexes were detected (Fig. 2 and 3A).

Both Sp1 and Sp3 bind, with various affinities, to the three transcriptionally active LTRs. Computer analyses of the U3 region sequences of the HERV-H LTRs by using transcription factor data bases (11) identified consensus binding sites for several human transcription factors, among them, AP2, Sp1, GATA-1, -2, and -3, Myb, and PEA3. In an attempt to identify some of the LTR-binding proteins, GMSA with different oligonucleotide competitors containing the binding sites for the transcription factors AP1, AP2, AP3, NF-1, NF-κB, and Sp1 were performed. The oligonucleotide containing a consensus Sp1 binding site (GC box) was found to inhibit the formation of two to three protein-DNA complexes, while none of the other oligonucleotide competitors changed the band patterns (Fig. 2A and 3A). GMSA with probes spanning a 65-bp region from the TATA box to the transcription initiation site (Fig. 1) showed that the Sp1-related proteins bound to the 3′ half of the U3 region (Fig. 2B and 3B). GMSA with a polyclonal Sp1

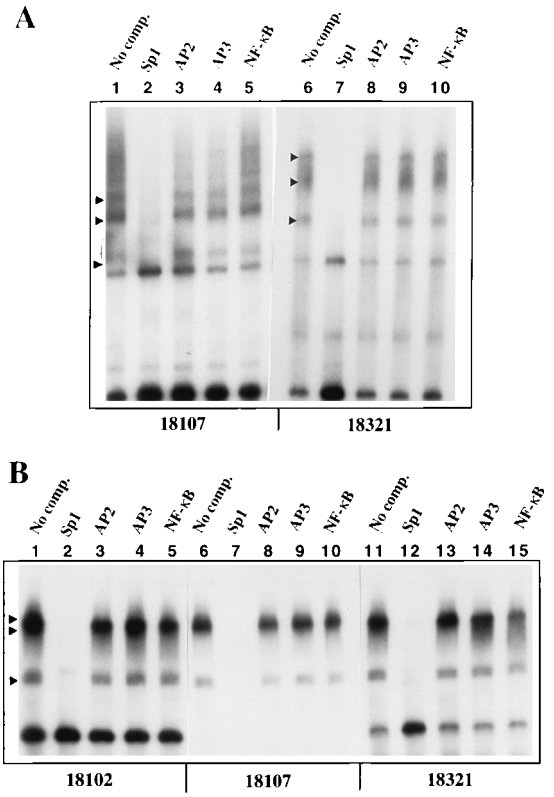

FIG. 2. Specific binding of GC box-binding proteins to the 3′ part of the U3 region of HERV-H LTRs. (A) NTera2-D1 nuclear extracts (1 μg) were incubated with end-labelled fragments generated by PCR with primers P1 and P2 of LTRs 18321 (lanes 1 to 5) and 18107 (lanes 6 to 10). Different specific oligonucleotide competitors containing consensus binding sites for the transcription factor AP2, AP3, NF-κB, or Sp1 were added as indicated. The complexes were separated on a 4% (39:1) polyacrylamide gel. The specific protein-DNA complexes inhibited by competition with the Sp1 oligonucleotide are indicated by arrowheads. (B) The three specific protein-DNA complexes inhibited by competition with the Sp1 oligonucleotide are due to proteins binding to the 3′ part of the LTR U3 region. NTera2-D1 nuclear extracts (2 μg) were incubated with end-labelled fragments generated by PCR with primers P2 and P3 of the promoter-active LTRs 18102 (lanes 1 to 5), 18107 (lanes 6 to 10), and 18321 (lanes 11 to 15). Seventy nanograms of different specific oligonucleotide competitors containing consensus binding sites for the transcription factor AP2, AP3, NF-κB, or Sp1 was added as indicated. The specific protein-DNA complexes inhibited by the Sp1 oligonucleotide are indicated by arrowheads. No comp., no competitor.

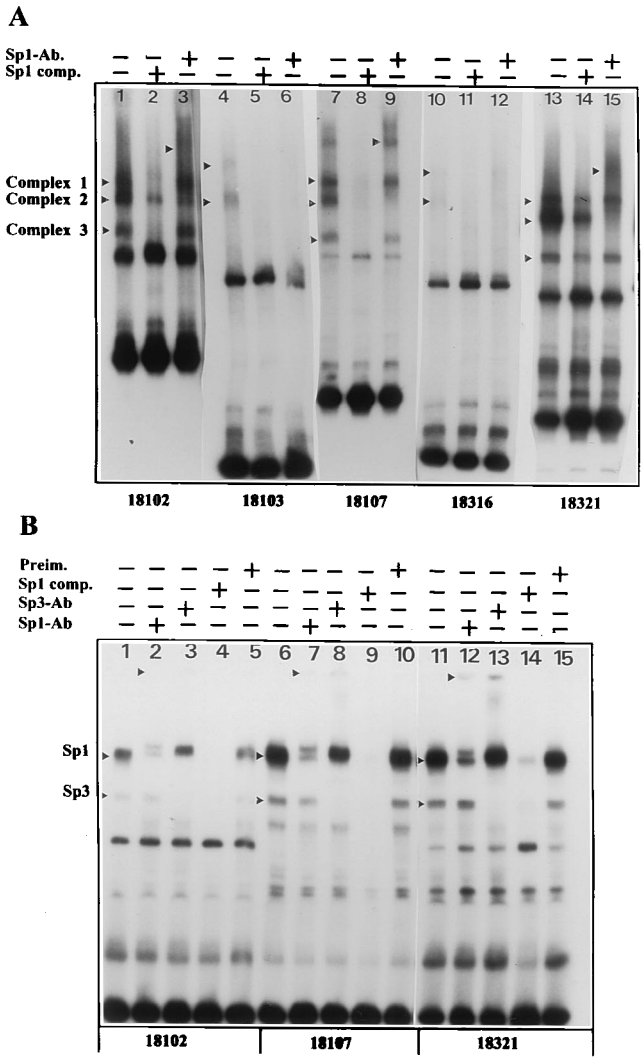

FIG. 3. Binding of Sp1 and Sp3 to HERV-H LTRs. (A) HeLa nuclear extracts (1 μg) were incubated with labelled fragments of the U3 regions of LTRs 18102, 18103, 18107, 18316, and 18321 generated by PCR with primers P1 and P2 (lanes 1, 4, 7, 10, and 13, respectively). In lanes 2, 5, 8, 11, and 14, 70 ng of a consensus Sp1 competitor (comp.) oligonucleotide was added to the binding reaction mixtures, while in lanes 3, 6, 9, 12, and 15, 2 μg of a polyclonal Sp1 antibody (Ab.) (Santa Cruz Biotechnology) was added. Arrowheads indicate protein-DNA complexes inhibited by competition with the consensus GC box oligonucleotide or supershifted by the Sp1 antibody. (B) Both Sp1 and Sp3 bind to the promoter-active LTRs. NTera2-D1 nuclear extracts (2 μg) were incubated with the 65-bp labelled fragments of LTRs 18102, 18107, and 18321 generated by PCR with primers P2 and P3. The Sp1 and Sp3 complexes as well as the supershifted complexes are indicated by arrowheads. The Sp1 antibody was added in lanes 2, 7, and 12; the Sp3 antibody was added in lanes 3, 8, and 13; and 70 ng of the Sp1 oligonucleotide competitor was added in lanes 4, 9, and 14. As a control, preimmune serum (Preim.) was added in lanes 5, 10, and 15.

antibody showed that only the high-intensity band inhibited by competition with the Sp1 oligonucleotide was supershifted (Fig. 3). This strongly suggested that the inhibited high-intensity band represented Sp1 binding to the LTRs, while the two minor bands could represent other proteins binding to the GC-rich region. Several different proteins binding to GC and GT boxes have been reported (21, 27, 30, 33, 55, 56, 62). A comparison of our band patterns with those reported by others indicated that complex 3 could represent binding of the Sp1-related protein Sp3 (20, 21, 33). GMSA with polyclonal Sp1- and Sp3-specific antibodies (20) confirmed that complex 2 was due to binding of Sp1 and that complex 3 was due to binding of Sp3 (Fig. 3B). Complex 1 is inhibited by competition with the Sp1 oligonucleotide but not supershifted by the Sp1 or Sp3 antibodies, and it may represent binding of another member of the Sp1 protein family or an unrelated GC box-binding protein. Thus, our results suggest that both the transcriptional activator Sp1 and the repressor of Sp1-mediated transcriptional activation Sp3 bind to the 3′ half of the HERV-H LTR's U3 region.

The results in Fig. 3B could indicate that the various LTRs exhibit different affinities for the three Sp1 family proteins and that the Sp1-DNA complex is dominant. Confirming this assumption, competition analyses with increasing amounts of the Sp1 competitor oligonucleotide (data not shown) revealed that LTR 18321, which contains a GC box, bound the Sp1 family proteins with the highest affinity, while LTRs 18102 and 18107, both of which contain a GT box in this position (Fig. 1), bound the proteins with a lower affinity. LTR 18103, which contains a GT box located 8 bp further downstream relative to the TATA

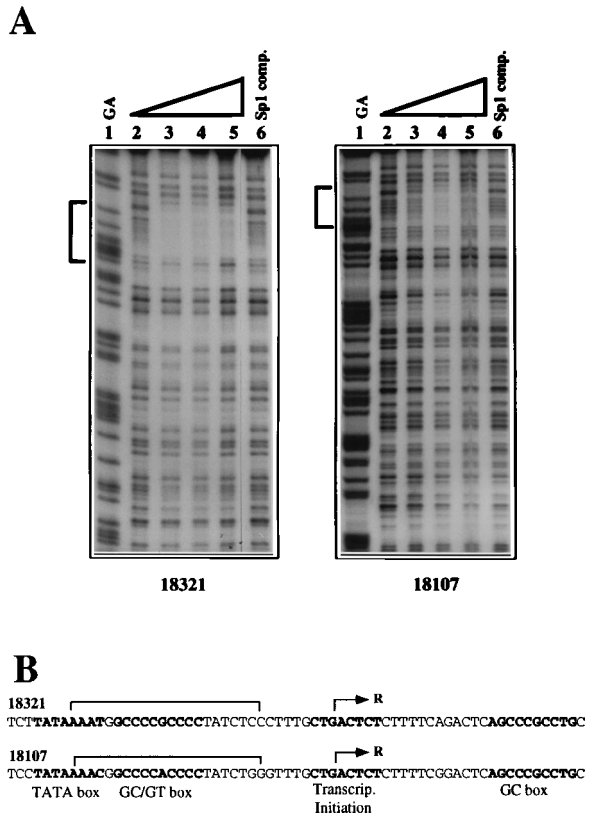

FIG. 4. Sp1 protects a region encompassing the putative Sp1 binding site downstream of the TATA box. (A) A 3′ end-labelled fragment of LTR 18321 generated by PCR with primers P1 and P5 or of LTR 18107 generated with primers P1 and P2 was incubated with 8, 16, and 32 μg of NTera2-D1 nuclear extracts (lanes 3, 4, and 5, respectively). Lanes 2 are controls without nuclear proteins, while in lanes 6, 140 ng of the Sp1 oligonucleotide competitor (comp.) was included in the binding reaction mixture containing 16 μg of protein. Lanes 1 contain the Maxam-Gilbert GA sequencing reaction mixtures. The Sp1 footprints are indicated by brackets. Because of increased unspecific binding, the footprint of LTR 18107 partly disappeared when the amount of crude nuclear extracts was increased from 16 to 32 μg. (B) Sequences of LTRs 18321 and 18107, respectively, with the protected regions indicated by brackets. Arrows denote the transcription (Transcrip.) initiation sites. The TATA box, the GC/GT box, the putative Inr element, and the distal GC box are shown in boldface type.

box, and LTR 18316, which has only a C-rich sequence at this position, bound Sp1 very weakly, and no binding of Sp3 or complex 1 was detected (Fig. 3A). Both LTRs 18102 and 18316 contain a consensus Sp1 binding motif about 150 bp 5′ to the TATA box. However, this site binds Sp1 very weakly (data not shown).

In order to further confirm that members of the Sp1 protein family bind to the putative Sp1 binding site downstream of the TATA box, DNase I footprinting assays were performed with probes spanning the U3 regions of LTRs 18107 and 18321. Figure 4A shows that for both LTRs there is a protected area encompassing the putative Sp1 binding site 3′ to the TATA box. The protection disappeared when the Sp1 competitor oligonucleotide was added to the binding reaction (Fig. 4A). The protected areas of 22 bp are centered on the central G or A in the GC or GT boxes, respectively, and include the 3′ half of the TATA box (Fig. 4B). Strikingly, footprinting with a probe spanning both the U3 and R regions of LTR 18321 demonstrated that a consensus Sp1 binding site located in the R region of LTR 18321 was not protected (data not shown). Also, GMSA performed with probes spanning different regions of the LTRs indicated that Sp1 mainly bound to only one site in the LTRs (data not shown). Taken together, our results clearly suggest that for these LTRs the Sp1 protein family seems to prefer the binding site located 3′ to the TATA box.

Both Sp1 and Sp3 stimulate transcription from the HERV-H LTR promoters in NTera2-D1 cells. To evaluate the functional importance of Sp1 and Sp3 binding relative to the observed promoter activities of HERV-H LTRs, we performed cotransfection analyses with *Drosophila* Schneider SL-2 cells by using the expression plasmids pPac-Sp1 (7) and pPac-Sp3 (20). Since *Drosophila* SL-2 cells do not express any endogenous Sp1, this system is well suited for evaluating the importance of Sp1 for promoter activity (7). We found that the LTR 18102, 18107, and 18321 promoters were strongly activated by Sp1 in this system (Fig. 5A). Cotransfection with 100 ng of pPac-Sp1 resulted in a 10- to 14-fold induction, which is slightly stronger than the Sp1 induction of the herpes simplex virus thymidine kinase (TK) promoter in pBLCAT2. The TK promoter contains two Sp1 binding sites, one distal, which is important for Sp1 induction, and one proximal, which is less important (7). The LTR 18103 promoter displayed only a three- to fourfold activation, while LTR 18316 was not activated at all (Fig. 5A). Thus, these results correlate very well with the binding affinities that the LTRs exhibit for Sp1. Nuclear extracts from SL-2 cells transfected with the pPac-Sp1 plasmid were used in GMSA with the five LTRs to verify that Sp1 was expressed and bound to the 3′ part of the LTRs (Fig. 5B). The relative efficiency of steady-state Sp1 binding corresponded completely with the binding affinity we observed for Sp1 in human cell lines: LTR 18321 bound Sp1 with high affinity, LTRs 18107 and 18102 bound with medium affinity, LTR 18103 bound with very low affinity, and the completely inactive LTR 18316 did not bind Sp1 at all. Interestingly, LTRs 18107 and 18102, both of which contain a weaker Sp1 binding site than LTR 18321, showed an Sp1 induction close to and even stronger than that of LTR 18321. This may be due to a weak Sp1 binding site in the U3 region upstream of the TATA box, which is not present in LTR 18321 (data not shown). It has previously been shown that Sp1 acts synergistically and that high-level activation by Sp1 requires multiple Sp1 binding sites (7). Our results show that Sp1 is clearly capable of activating transcription from several different HERV-H LTR promoters. This activation corresponds well with the promoter activity of the LTRs in human cells, indicating that Sp1 may be important for the in vivo activity of the LTRs.

Sp3 was recently shown to act as a repressor of Sp1-mediated transcriptional activation (20) and of the HIV-1 LTR (46). We therefore wanted to determine if Sp3 could repress the transcriptional activity of the HERV-H LTRs or at least repress the Sp1-mediated transcriptional activity. Cotransfections with pPac-Sp3 in SL-2 cells resulted in very low-level or nearly no activation of the transcriptionally active LTRs compared with Sp1 (Fig. 6A). GMSA with nuclear extract from SL-2 cells transfected with pPac-Sp3 verified that Sp3 was expressed and bound to the LTRs (Fig. 6C). Coexpression of Sp3 repressed the Sp1-mediated stimulatory effect in a dose-dependent manner (Fig. 6B), but this repression was not as potent as the repression of the HIV-1 LTR promoter reported by Majello and coworkers (46). A 10-fold excess of pPac-Sp3 resulted in a 50% reduction of the Sp1-mediated activity, while a complete repression was seen for the HIV-1 LTR. This difference may be due to the weaker binding of Sp3 than Sp1 to the HERV-H LTRs, compared with the HIV-1 LTR. GMSA with nuclear extract from SL-2 cells transfected with pPac-Sp3 or pPac-Sp1 indicated that the Sp3 binding to the HERV-H LTRs was 60 to 70% of the Sp1 binding (data not

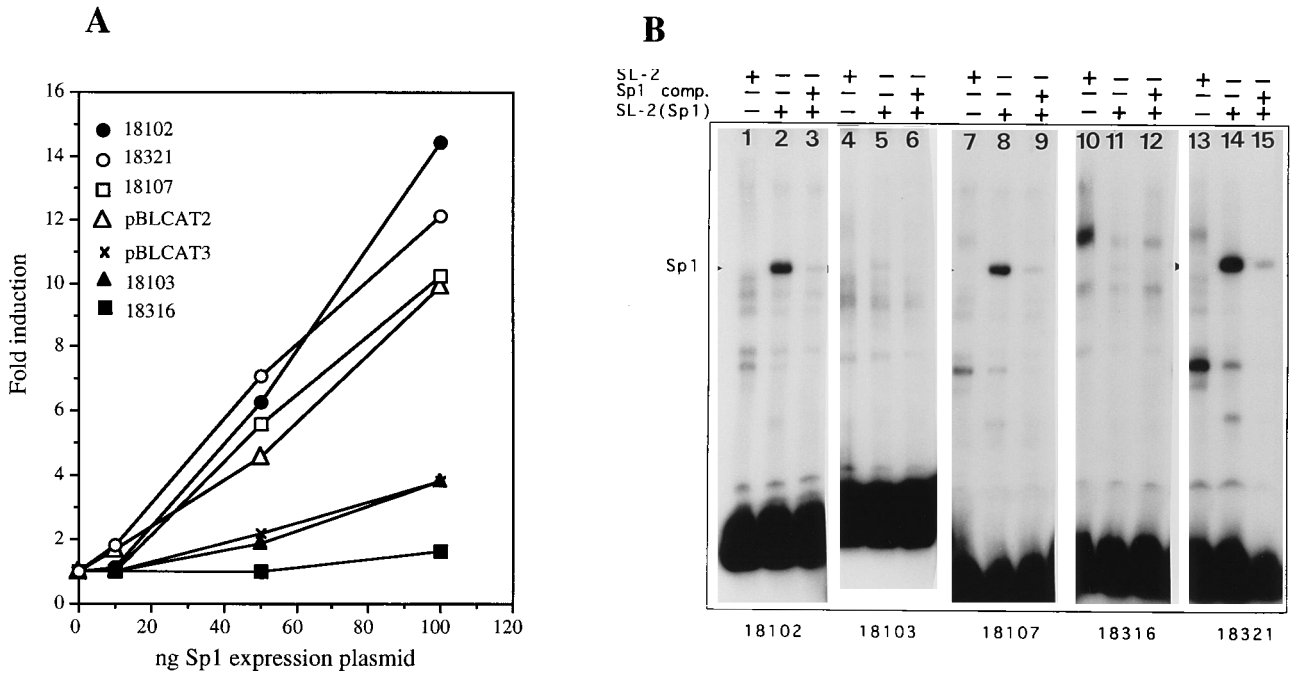

FIG. 5. The transcriptional activity of the promoter-active HERV-H LTRs in SL-2 cells is strongly induced by Sp1. (A) Reporter vectors containing different HERV-H LTRs inserted upstream of the CAT gene in pBLCAT3 were cotransfected with increasing amounts of pPac-Sp1 expression plasmid in *Drosophila* SL-2 cells. Fold induction is plotted versus the amount (nanograms) of cotransfected Sp1 expression plasmid. pBLCAT2 (containing the herpes simplex virus TK promoter in front of the CAT gene) and pBLCAT3 (promoterless) were used as positive and negative controls, respectively. LTRs 18102, 18107, and 18321 show higher fold inductions than pBLCAT2. The fold induction of the TK promoter found here is very similar to that reported earlier (7). The data represent the means from three independent experiments performed in triplicate. (B) Nuclear extracts from SL-2 cells transfected with the Sp1 expression vector were incubated with labelled fragments spanning the 3′ halves of the U3 regions of LTRs 18102 (lanes 2 and 3), 18103 (lanes 5 and 6), 18107 (lanes 8 and 9), 18316 (lanes 11 and 12), and 18321 (lanes 14 and 15). As controls, the same fragments were incubated with nuclear extracts from untransfected SL-2 cells in lanes 1, 4, 7, 10, and 13, respectively, while 70 ng of the Sp1 oligonucleotide competitor (comp.) was added to the binding reactions in lanes 3, 6, 9, 12, and 15, respectively. The Sp1-DNA complexes are indicated by arrowheads.

shown). The HIV-1 LTR was reported to bind Sp1 and Sp3 with similar affinities.

To explore the potential activation of the LTRs by Sp1 and Sp3 in mammalian cells more directly, NTera2-D1 cells were cotransfected with the GC box-containing LTR 18321 or the GT box-containing LTR 18107 and various amounts of the mammalian expression vectors pCMV-Sp1 and pCMV-Sp3 (20). Both LTRs 18321 and 18107 were clearly stimulated by Sp1, with a 2.5- to 4.5-fold induction when 2 μg of Sp1 expression plasmid was used (Fig. 7A). Larger amounts of cotransfected Sp1 expression plasmid (from 5 to 8 μg, depending on the LTR), led to repression, most probably because of squelching (50), since Sp1 interacts with both TFIIB and several members of the TFIID complex (6, 10). Surprisingly, cotransfection of pCMV-Sp3 in NTera2-D1 cells resulted in a two- to fourfold stimulation of the promoter activities of both of the LTRs and the TK promoter of pBLCAT2 (Fig. 7A). Also, the third promoter-active LTR, 18102, was stimulated by Sp3 in a similar manner (data not shown). Cotransfections with a mixture of pCMV-Sp1 and pCMV-Sp3 did not indicate any repression of Sp1-mediated activity by Sp3. Thus, in NTera2-D1 cells, Sp3 seems to be a transcriptional activator similar to Sp1 and not a repressor of Sp1-mediated transcriptional activation as found in *Drosophila* SL-2 cells and the mammalian cell lines HeLa, Ishikawa, and CV-1 (20, 46). These intriguing results prompted us to perform cotransfections in HeLa cells. We found that Sp1 was an even more potent activator of the HERV-H LTR promoters in HeLa cells than in NTera2-D1 cells, while Sp3 did not stimulate transcription from the LTR promoters at all (Fig. 7B). Similar results were obtained for the TK promoter in pBLCAT2, where Sp3 in addition seemed to be a weak repressor (data not shown). These results suggest that Sp3, although being ubiquitously expressed, may be involved in cell-specific regulation of transcription, perhaps via cell-specific coactivators.

The small GC box-binding protein BTEB is also able to activate the HERV-H LTR promoters. The Sp1-related low-molecular-weight GC/GT box-binding protein BTEB was recently found to be a transcriptional activator of the HIV-1 LTR promoter (25) and of other promoters containing more than one GC/GT box (27). Cotransfections of NTera2-D1 cells with the BTEB expression plasmid pRSVBTEB stimulated the transcriptional activity of the LTR promoters two- to fourfold, similar to Sp1 and Sp3. However, for the TK promoter in pBLCAT2, BTEB was a more potent activator than Sp1 and Sp3 (Fig. 7A). GMSA with nuclear extracts from NTera2-D1 cells transfected with the pRSVBTEB expression vector confirmed that BTEB was expressed and bound to the LTRs (data not shown), but the BTEB complex did not comigrate with any protein-DNA complexes of untransfected NTera2-D1 cells. This indicates that BTEB is not expressed in NTera2-D1 cells, which is in agreement with a recent report showing human BTEB to be transcribed ubiquitously but to be translated only in nerve cells in the brain and sperm in the testis (26).

Stimulation of transcription by Sp1 is not dependent on the TATA box. The proximity of the Sp1 binding site to the TATA box (Fig. 1) raised the question whether Sp1 may sterically interfere with binding of the general transcription factor complex TFIID to the TATA box. If so, Sp1 may direct the transcriptional stimulation independently of the TATA box, per-

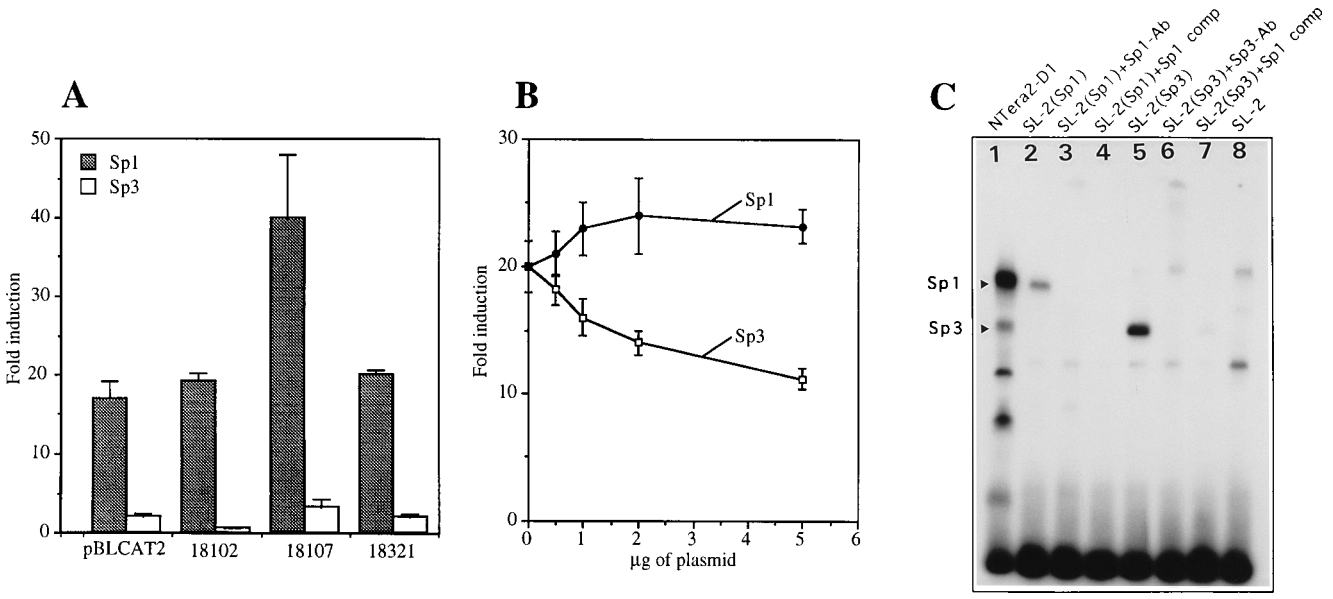

FIG. 6. Sp3 is unable to activate transcription in SL-2 cells and functions instead as a repressor of Sp1-mediated transcription. (A) Sp3 does not stimulate the transcriptional activity of the HERV-H LTRs in SL-2 cells. Reporter vectors containing the promoter-active HERV-H LTRs 18102, 18107, and 18321 inserted upstream of the CAT gene in pBLCAT3 were cotransfected with 0.5 μg of the Sp1 expression plasmid pPac-Sp1 or 0.5 μg of the Sp3 expression plasmid pPac-Sp3. pBLCAT2 was used as a positive control. The CAT activity of extracts from cells transfected only with the reporter plasmids was set at 1.0. The data are the means from two independent experiments performed in triplicate. Error bars indicate standard errors of the means. (B) Sp3 acts as a repressor of Sp1-mediated transcriptional activation in SL-2 cells. LTR 18321 inserted upstream of the CAT gene in pBLCAT3 was used as a reporter plasmid and was cotransfected with a constant amount of the Sp1 expression plasmid pPac-Sp1 (0.5 μg) in the presence of increasing amounts of the Sp3 expression plasmid pPac-Sp3 as indicated. Cotransfection with increasing amounts of pPac-Sp1 was used as a control to ensure that squelching did not occur. The CAT activity of extracts from cells cotransfected with only the reporter plasmid was set at 1.0. The data represent the means from two independent experiments performed in triplicate. Error bars indicate standard errors of the means. (C) Sp1 and Sp3 are expressed in *Drosophila* SL-2 cells transfected with pPac-Sp1 and pPac-Sp3, respectively. Nuclear extracts from *Drosophila* SL-2 cells transfected with pPac-Sp1 (lanes 2 to 4) or pPac-Sp3 (lanes 5 to 7) were incubated with the 65-bp labelled fragment of LTR 18321 generated by PCR with primers P2 and P3. As controls, nuclear extracts from NTera2-D1 cells were used in lane 1, Sp1 antibody (Ab) was added in lane 3, Sp3 antibody was added in lane 6, 70 ng of the Sp1 oligonucleotide competitor (comp.) was added in lanes 4 and 7, and nuclear extracts from untransfected SL-2 cells were used in lane 8.

haps via an Inr element further downstream. Inspection of the sequence of the LTR region where the transcriptional initiation site has been found to be located (13) revealed that there may be an Inr element at the transcriptional initiation site. This element is pyrimidine rich and contains the critical A in the +1 position and T in the +3 position (28) (Fig. 1). The LTRs contain a G in the −1 position instead of the C found to be optimal for the Inr activity. However, Inr elements with a G in the −1 position have also been found to be functional (28, 54). The presence of the G generally shifts the transcriptional initiation site from the +1 A to the −1 G (54). To determine the importance of the TATA box and the Sp1 binding site for the transcriptional activity, an LTR 18321 with a mutated TATA box, LTR 18321-Tm, and another with a mutated Sp1 binding site, LTR 18321-Sp1m, were generated (Fig. 1). The mutated LTRs were inserted upstream of the CAT gene in pBLCAT3. Transfections of these constructs into NTera2-D1 cells showed that mutation of the TATA box only slightly reduced the transcriptional activity of LTR 18321, while mutation of the Sp1 binding site reduced the transcriptional activity to background levels (Fig. 8A). Also, cotransfections with both Sp1 and Sp3 stimulated the transcription from the TATA-mutated LTR, while the Sp1 binding site-mutated LTR was unaffected (Fig. 8B). These results clearly suggest that the Sp1 binding site is critical for the transcriptional activities of these LTR promoters and that these activities are seemingly independent of the adjacent TATA box. Thus, the Sp1 binding site enables the LTR promoters to act reasonably well also as TATA-less promoters, probably via the putative Inr element located at the transcriptional initiation site. GMSA with NTera2-D1 nuclear extracts verified that the Sp1 family proteins bound to LTR 18321 when the TATA box was mutated but bound not at all or very weakly when the Sp1 binding site was mutated (Fig. 8C). Interestingly, the amount of a retarded protein-DNA complex, which was not found to be Sp1 related, was strongly reduced when the TATA box was mutated (Fig. 8C).

In order to determine whether the promoter activity also is dependent on sequences upstream of the TATA box or in the R and U5 regions of the LTRs, a minimal promoter element of 65 bp containing only the TATA box, the Sp1 binding site, and the putative Inr element (generated by PCR with primers P2 and P3) was inserted upstream of the CAT gene and transfected into NTera2-D1 cells alone and together with the Sp1 and Sp3 expression plasmids. The results of these transfections showed that this minimal promoter contains a transcriptional activity close to that of the full-length LTR (Fig. 8A) and that its transcriptional activity is stimulated by Sp1 and Sp3 similarly to the full-length LTR (Fig. 8B). Taken together, these results show that the TATA box-proximal Sp1 binding sequence, probably together with a putative Inr element, is necessary and sufficient to drive the transcriptional activity of the LTR promoter.

Sp1 is not important for enhancer activity of the HERV-H LTRs. HERV-H LTRs isolated by others have been found to contain enhancer activity (13). In order to determine if these five LTRs could stimulate transcription from the TK promoter in a distance- and orientation-independent manner as expected for enhancers, they were inserted in both orientations downstream of the CAT gene in pBLCAT2. Transient-transfection assays showed that three of the five LTRs displayed enhancer

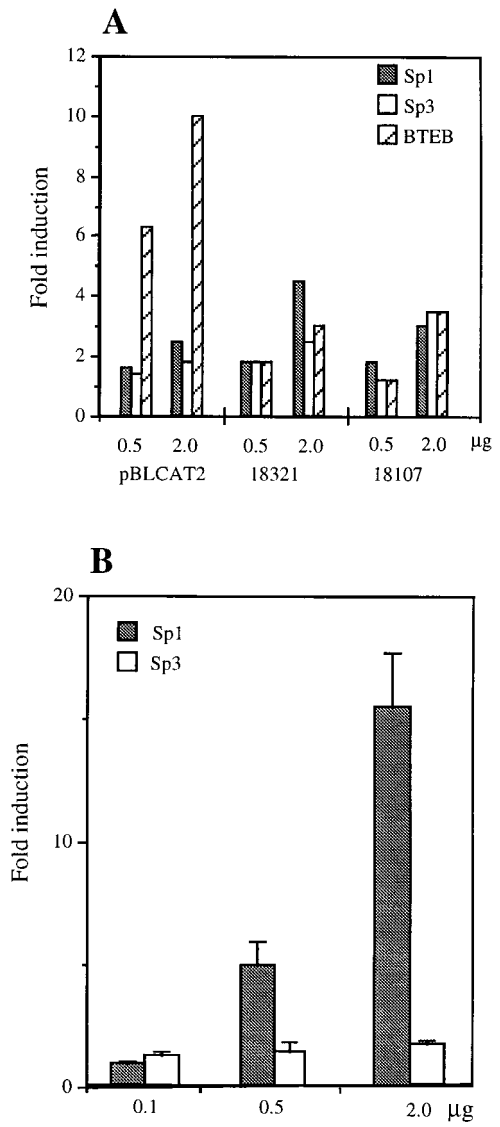

FIG. 7. The three GC/GT box-binding proteins Sp1, Sp3, and BTEB stimulate the transcriptional activities of the promoter-active LTRs in NTera2-D1 cells. (A) Reporter vectors for the GC box containing LTR 181321 and the GT box containing LTR 18107 were cotransfected with 0.5 and 2.0 μg of the Sp1 expression plasmids pRSVSp1 and pCMV-Sp1, the Sp3 expression plasmid pCMV-Sp3, and the BTEB expression plasmid pRSVBTEB. pBLCAT2 was used as a positive control. The CAT activity of extracts from cells transfected with the reporter plasmids alone was assigned a value of 1.0. Other CAT activities were normalized accordingly. The data represent the means from at least four independent experiments using different plasmid preparations. The standard error varied between 5 and 25%. (B) Sp3 is not a transcriptional activator of the LTR promoter in HeLa cells. The LTR 18321 reporter vector was cotransfected with 0.1, 0.5, and 2.0 μg of the Sp1 expression plasmid pCMV-Sp1 and the Sp3 expression plasmid pCMV-Sp3. The CAT activity of extracts from cells transfected with LTR 18321 and 2 μg of empty pRc-CMV expression vector was set at 1.0. The values represent the means from three independent experiments performed in triplicate with different preparations of the plasmids. Error bars indicate standard errors of the means.

activity in NTera2-D1 cells (Fig. 9). LTR 18102 proved to be the strongest enhancer, with a 15-fold induction of the TK promoter, while LTRs 18107 and 18321 showed 6- to 8-fold inductions. Sp1 has been shown to act from both adjacent and distant sites, and we therefore wanted to see if Sp1 also could be important for the enhancer capacity of the LTRs. Cotransfections in *Drosophila* SL-2 cells with various amounts of the Sp1 expression plasmid were performed, but none of the LTRs demonstrated any stimulation of the TK promoter activity even when 2 μg of pPac-Sp1 was cotransfected (data not shown). Thus, Sp1 seems not to be important for the enhancer activity of these LTRs. This notion was further confirmed by analyses of the enhancer activity of the HERV-H U3 region upstream of the TATA and GC/GT boxes (Fig. 1). This part of the U3 region was inserted in both orientations downstream of the CAT gene in pBLCAT2. For LTRs 18102 and 18107, both orientations of the complete LTRs and the U3 regions were tested in NTera2-D1 cells, and the induction of the TK promoter was shown to be orientation independent (data not shown). Importantly, the U3 regions alone exhibited enhancer activities close to those of the complete LTRs (Fig. 9). Thus, the LTR enhancer activity seems to be dependent on proteins binding upstream of the TATA and GC/GT boxes.

DISCUSSION

The results presented in this paper show that members of the Sp1 protein family are critically involved in the regulation of the transcriptional activity of the HERV-H LTR promoters. We have found that transcriptionally active LTRs contain a GC/GT box immediately downstream of the TATA box and that three different members of the Sp1 protein family can bind to this site and modulate the promoter activity. First, Sp1, which is a well-known activator of transcription of both cellular and viral genes, binds to this GC/GT box and stimulates the promoter activity in all cell lines tested. Second, Sp3, which has been reported to act as a repressor of Sp1-mediated transcriptional activation and of the HIV-1 LTR promoter (20, 21, 33, 46), binds to the GC/GT box. As expected, we found that Sp3 repressed the Sp1-mediated transcriptional activation of the HERV-H LTRs in *Drosophila* SL-2 cells and had no stimulatory effect on the transcription from the LTR promoters and the TK promoter in HeLa cells and *Drosophila* SL-2 cells. However, in the embryonal carcinoma cell line NTera2-D1, Sp3 stimulated the transcriptional activities of both the TK and the HERV-H LTR promoters. These results suggest that Sp3 may be dependent on cell-specific cofactors to act as a transcriptional activator and thus may be involved in cell-specific regulation of both the HERV-H LTRs and other GC/GT box-containing promoters. Interestingly, a B-cell-specific coactivator converting the ubiquitously expressed transcription factor Oct-1 to a cell-type-specific activator has recently been characterized (19). The Sp3-mediated repression of Sp1-directed transcriptional activation is most probably due to competitive binding between these two proteins. Thus, various levels of the ubiquitously expressed Sp1 and Sp3 proteins in different cell lines may result in significant variation of the transcriptional activity from the LTR promoters, which together with the cell-specific activation potential of Sp3 may contribute to the different expression levels of HERV-H transcripts in different cell lines and tissues (22, 59). Third, competition with the Sp1 oligonucleotide showed that a third protein also bound to the GC/GT box of the LTRs. This protein could be the Sp1-related Sp4 or another, not-yet-identified, member of the Sp1 protein family or an unrelated GC/GT box-binding protein. Fourth, the more distantly Sp1-related GC/GT box-binding protein BTEB also bound to and stimulated the promoter activities of the HERV-H LTRs. However, gel shift assays indicated that BTEB is not expressed in NTera2-D1 and HeLa cells, which is in agreement with the fact that BTEB is specifically translated in brain and testis (26). Thus, tissue-specific members of the Sp1 protein family also can bind to and activate the LTR

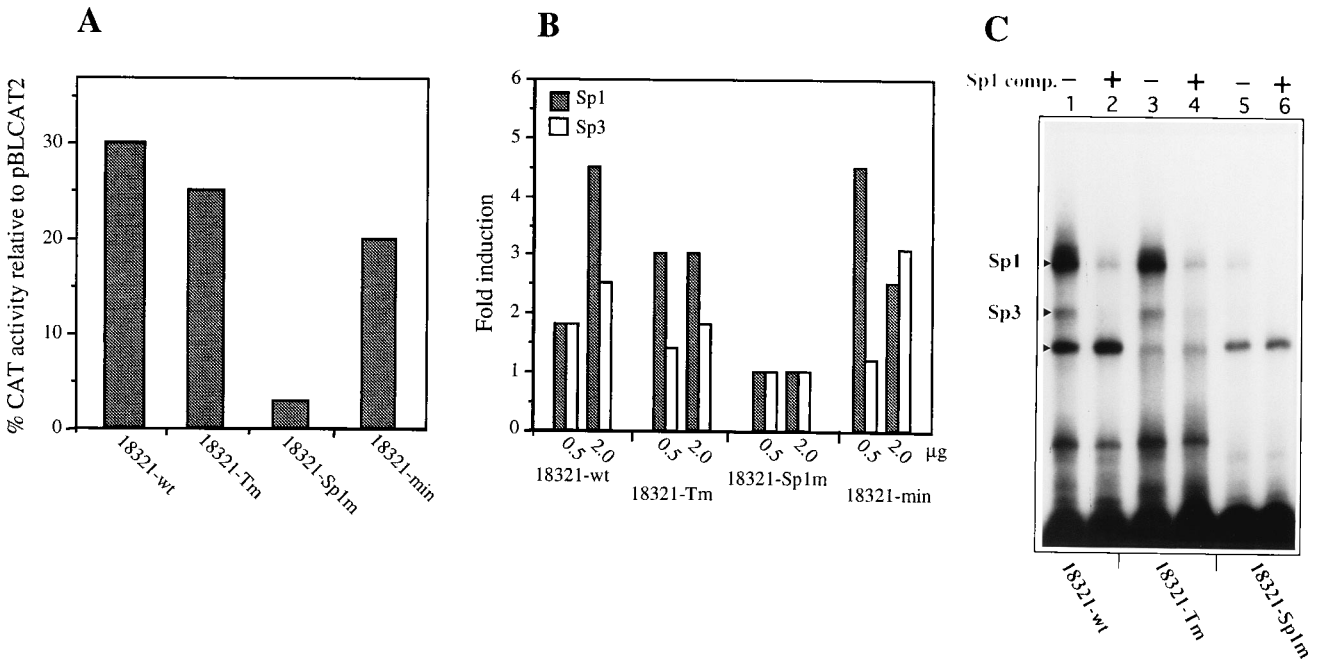

FIG. 8. The Sp1 binding sequence, but not the TATA box, is required for the promoter activity of LTR 18321. (A) Reporter vectors containing wild-type LTR 18321 (18321-wt), LTR 18321 with a mutated TATA box (18321-Tm), LTR 18321 with a mutated Sp1 binding site (18321-Sp1m), and the LTR 18321 minimal promoter (a 65-bp fragment of LTR 18321 generated by PCR with primers P2 and P3) (18321-min) inserted upstream of the CAT gene in pBLCAT3 were transfected into NTera2-D1 cells. The CAT activity of pBLCAT2 was set at 100%. The values represent the means from at least four independent experiments using different preparations of the plasmids. The standard error varied between 5 and 25%. (B) The four reporter plasmids described for panel A were cotransfected with 0.5 and 2.0 μg of the Sp1 expression plasmid pCMV-Sp1 or the Sp3 expression plasmid pCMV-Sp3. The CAT activities of extracts from cells transfected with the reporter plasmids alone were assigned a value of 1.0. The values represent the means from at three independent experiments using different preparations of the plasmids. The standard error varied between 5 and 25%. (C) Sp1 and Sp3 bind to LTR 18321 with a mutated TATA box but not to LTR 18321 with a mutated Sp1 binding site. Nuclear extracts (2 μg) from NTera2-D1 cells were incubated with labelled fragments (generated by PCR with primers P3 and P5) of wild-type LTR 18321 (lanes 1 and 2), LTR 18321 with a mutated TATA box (lanes 3 and 4), and LTR 18321 with a mutated GC box (lanes 5 and 6). The specific Sp1 competitor (comp.) oligonucleotide (70 ng) was added in lanes 2, 4, and 6. The Sp1- and Sp3-DNA complexes are indicated, as is an unidentified TATA box-binding protein.

promoters and thereby induce tissue-specific expression of these elements.

Transient transfections with three different constructs of the transcriptionally active LTR 18321, one with a mutated TATA box, one with a mutated GC box, and one with only the 65-bp minimal promoter element, showed that the transcriptional activity of the LTR promoter is highly dependent on the TATA-proximal GC box. This result is supported by the fact that no subtype II LTRs (18), which lack this GC/GT box (Fig. 1), have been found to contain promoter activity in human cell lines (1, 13). In addition, the promoter-inactive subtype I LTRs do not contain a consensus GC/GT box in this position. Surprisingly, mutation of the TATA box did not significantly reduce the transcriptional activity of the promoter, suggesting that these promoters are able to act independently of the TATA box. Thus, this GC/GT box is required and, perhaps together with the putative Inr element, is also sufficient for transcriptional activity of the HERV-H LTR promoters.

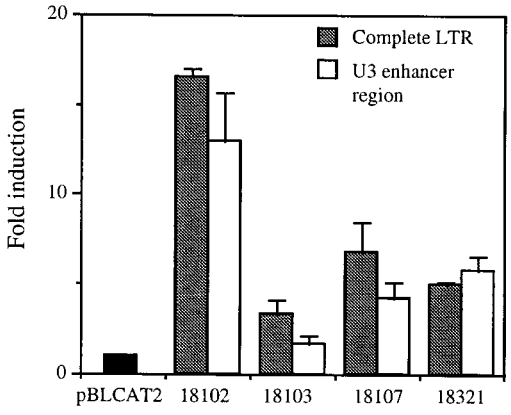

FIG. 9. The enhancer activity of HERV-H LTRs is mainly due to sequences in the U3 region upstream of the TATA box. Reporter vectors containing either the complete LTR or the U3 region upstream of the TATA box (Fig. 1) of LTRs 18102, 18103, 18107, and 18321 inserted downstream of the CAT gene in pBLCAT2 were transfected into NTera2-D1 cells. The data represent the means from three independent experiments performed in triplicate and are shown as fold induction relative to the activity of pBLCAT2 without any LTR insert. Error bars indicate standard errors of the means.

The proximity of the GC/GT box to the TATA box (Fig. 1) raised the question whether Sp1 family proteins could bind to this site simultaneously with the general transcription factor complex TFIID binding to the TATA box. Recently it has been reported that the footprint of the TFIID complex bound to the TATA box extends 42 bp downstream of the transcription initiation site (57). However, the crystal structure of the TATA box-binding protein (TBP) binding to the TATA box shows that TBP alone does not contact any bases downstream of the TATA box (31, 32), indicating that members of the Sp1 protein family and TBP may bind simultaneously to the HERV-H LTR promoters during initiation of transcription. However, a mutual exclusion between these proteins is very likely, since Sp1 binding to GC boxes separated by 2 or 3 bp is reported to be mutually exclusive (23). Our results showed that the promoter activity of the HERV-H LTRs is nearly independent of the

TATA box, suggesting that the GC/GT box, probably together with a putative Inr element, may direct binding of the general transcription complex to another position within the promoter.

Interestingly, some of the HERV-H LTRs contain consensus GC/GT boxes upstream of the TATA box exhibiting very low affinity for Sp1 and related proteins (data not shown). Thus, the Sp1 protein family seems to prefer the binding site located next to the AT-rich TATA box. Both Sp1 (24) and Sp3 (our unpublished results) distort the DNA structure upon binding. The preference for a binding site next to the TATA box may therefore be due to facilitated bending/unwinding of an AT-rich region, as has been reported for the yeast GC box-binding zinc finger protein MIG1 (42). Such distortion of the DNA structure will probably stimulate the formation of a transcription initiation complex in this region, since prebending of a promoter sequence has been shown to enhance the affinity for TBP and the TFIID complex (49).

Apart from the Sp1 family proteins, other nuclear proteins also bound to the HERV-H LTRs. One of these was a protein-DNA complex not found to be Sp1 related, which was strongly reduced when the TATA box was mutated. Thus, this complex may represent a TATA box-binding protein and could be TBP or another protein belonging to the TFIID complex. The intensity of this complex was enhanced when binding of Sp1 was inhibited with the Sp1 competitor oligonucleotide, indicating that Sp1 may sterically interfere with its binding.

Nuclear proteins were also found to bind to the U3 region upstream of the TATA box, but none of these proteins were identified. However, we found that these proteins are probably important for the LTRs' enhancer activity. Thus, further analysis of *cis*- and *trans*-acting factors acting upstream of the TATA box will be the subject of a future study. The high copy numbers and dispersed distributions of HERV-H elements and solitary LTRs in the human genome strongly suggest that nuclear proteins binding to these elements may influence the expression of nearby cellular genes. Also, reorganization of HERV-H LTRs by homologous recombination or transposition events (16) may change the expression pattern of cellular genes. Finally, the possible involvement of endogenous retroviral proteins in the development of autoimmunity and immunodeficiency diseases underlines the importance of understanding how these promoters are regulated in vivo.

ACKNOWLEDGMENTS

We are very grateful to A. Gègonne for SL-2 cells and pPac-Sp1, P. Andrews for NTera2-D1 cells, Y. Fujii-Kuryiama for pRSVSp1 and pRSVBTEB, and G. Suske for pPac-Sp3, pCMV-Sp1, pCMV-Sp3, and antibodies against Sp1 and Sp3. We also thank U. Moens for oligonucleotides.

This work was funded by grants from the Norwegian Cancer Society and the Aakre Foundation to T.J. E.S. is a fellow of the Norwegian Cancer Society.

REFERENCES

1. **Anderssen, S., E. Sjøttem, G. Svineng, and T. Johansen.** Unpublished data.
2. **Andrews, P. W.** 1984. Retinoic acid induced neuronal differentiation of a cloned human embryonal carcinoma cell line *in vitro*. Dev. Biol. **103:**285–293.
3. **Andrews, P. W., I. Damjanov, D. Simon, G. Banting, C. Carlin, N. Dracopoli, and J. Fogh.** 1984. Pluripotent embryonal carcinoma clones derived from the human teratocarcinoma cell line TERA-2. Lab. Invest. **50:**147–162.
4. **Boller, K., H. Konig, M. Souter, N. Mueller-Lantzsch, R. Löwer, J. Löwer, and R. Kurth.** 1993. Evidence that HERV-K is the endogenous retrovirus sequence that codes for the human teratocarcinoma derived retrovirus HTDV. Virology **196:**349–353.
5. **Brodsky, I., B. Foley, and D. Gillespie.** 1993. Expression of human endogenous retrovirus (HERV-K) in chronic myeloid leukemia. Leuk. Lymphoma **11:**119–123.
6. **Chiang, C. M., and R. G. Roeder.** 1995. Cloning of an intrinsic human TFIID subunit that interacts with multiple transcriptional activators. Science **267:** 531–536.
7. **Courey, J. A., and R. Tjian.** 1988. Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. Cell **55:**887–898.
8. **Di Nocera, P. P., and I. B. Dawid.** 1983. Transient expression of genes introduced into cultured cells of *Drosophila*. Proc. Natl. Acad. Sci. USA **80:**7095–7098.
9. **Dynan, W. S., and R., Tjian.** 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell **35:**79–87.
10. **Emili, A., J. Greenblatt, and C. J. Ingles.** 1994. Species-specific interaction of the glutamine-rich activation domains of Sp1 with the TATA box-binding protein. Mol. Cell. Biol. **14:**1582–1593.
11. **Faisst, S., and S. Meyer.** 1992. Compilation of vertebrate-encoded transcription factors. Nucleic Acids Res. **20:**3–26.
12. **Feuchter, A., J. D. Freeman, and D. L. Mager.** 1992. Strategy for detecting cellular transcripts promoted by human endogenous long terminal repeats: identification of a novel gene (CDCL4L) with homology to yeast CDC4. Genomics **13:**1237–1246.
13. **Feuchter, A., and D. L. Mager.** 1990. Functional heterogeneity of a large family of human LTR-like promoters and enhancers. Nucleic Acids Res. **18:**1261–1270.
14. **Feuchter-Murthy, A. E., J. D. Freeman, and D. L. Mager.** 1993. Splicing of a human endogenous retrovirus to a novel phospholipase A2 related gene. Nucleic Acids Res. **21:**135–143.
15. **Gègonne, A., R. Bosselut, R. A. Bailly, and J. Ghysdael.** 1993. Synergistic activation of the HTLV1 LTR Ets-responsive region by transcription factors Ets1 and Sp1. EMBO J. **12:**1169–1178.
16. **Goodchild, N. L., J. D. Freeman, and D. L. Mager.** 1995. Spliced HERV-H endogenous retroviral sequences in human genomic DNA: evidence for amplification via retrotransposition. Virology **206:**164–173.
17. **Goodchild, N. L., D. A. Wilkinson, and D. L. Mager.** 1992. A human endogenous long terminal repeat provides a polyadenylation signal to a novel, alternatively spliced transcript in normal placenta. Gene **121:**287–294.
18. **Goodchild, N. L., D. A. Wilkinson, and D. L. Mager.** 1993. Recent evolutionary expansion of a subfamily of HERV-H human endogenous retrovirus-like elements. Virology **196:**778–788.
19. **Gstaiger, M., L. Knoepfel, O. Georgiev, W. Schaffner, and C. M. Hovens.** 1995. A B-cell coactivator of octamer-binding transcription factors. Nature (London) **373:**360–362.
20. **Hagen, G., S. Mueller, M. Beato, and G. Suske.** 1994. Sp1-mediated transcriptional activation is repressed by Sp3. EMBO J. **13:**3843–3851.
21. **Hagen, G., S. Müller, M. Beato, and G. Suske.** 1992. Cloning by recognition site screening of two novel GT box binding proteins: a family of Sp1 related genes. Nucleic Acids Res. **20:**5519–5525.
22. **Hirose, Y., M. Takamatsu, and F. Harada.** 1993. Presence of *env* genes in members of of the HERV-H family of human endogenous retrovirus-like elements. Virology **192:**52–61.
23. **Ikeda, K., K. Nagano, and K. Kawakami.** 1993. Anomalous interaction of Sp1 and specific binding of an E-box-binding protein with the regulatory elements of the Na,K-ATP-ase $\alpha 2$ subunit gene promoter. FEBS Lett. **218:** 195–204.
24. **Ikeda, K., K. Nagano, and K. Kawakami.** 1993. Possible implications of Sp1-induced bending of DNA on synergistic activation of transcription. Gene **136:**341–343.
25. **Imataka, H., A. Mizuno, Y. Fujii-Kuriyama, and M. Hayami.** 1993. Activation of the human immunodeficiency virus type 1 long terminal repeat by BTEB, a GC box-binding transcription factor. AIDS Res. Hum. Retroviruses **9:**825–831.
26. **Imataka, H., K. Nakayama, K. I. Yasumoto, A. Mizuno, Y. Fujii-Kuriyama, and M. Hayami.** 1994. Cell-specific translational control of transcription factor BTEB expression. J. Biol. Chem. **269:**20668–20673.
27. **Imataka, H., K. Sogawa, K. Yasumoto, Y. Kikuchi, K. Sasano, A. Kobayashi, M. Hayami, and Y. Fujii-Kuriyama.** 1992. Two regulatory proteins that bind to the basic transcription element (BTE), a GC box sequence in the promoter region of the rat P-4501A1 gene. EMBO J. **11:**3663–3671.
28. **Javahery, R., A. Khachi, K. Lo, B. Zenzie-Gregory, and S. T. Smale.** 1994. DNA sequence requirements for transcriptional initiator activity in mammalian cells. Mol. Cell. Biol. **14:**116–127.
29. **Johansen, T., T. Holm, and E. Bjørklid.** 1989. Members of the HERV-H family of human endogenous retrovirus-like elements are expressed in placenta. Gene **79:**259–267.
30. **Kageyama, R., and I. Pastan.** 1989. Molecular cloning and characterization of a human DNA binding factor that represses transcription. Cell **59:**815–825.
31. **Kim, J. K., D. B. Nikolov, and S. K. Burley.** 1993. Co-crystal structure of TBP recognizing the minor groove of a TATA-element. Nature (London) **365:** 520–527.
32. **Kim, Y., J. H. Geiger, S. Hahn, and P. B. Sigler.** 1993. Crystal structure of a yeast TBP/TATA-box complex. Nature (London) **365:**512–520.
33. **Kingsley, C., and A. Winoto.** 1992. Cloning of GT box-binding proteins: a novel Sp1 multigene family regulating T-cell receptor gene expression. Mol. Cell. Biol. **12:**4251–4261.

34. **Krieg, A. M., M. F. Gourley, and A. Perl.** 1992. Endogenous retroviruses: potential etiologic agents in autoimmunity. FASEB J. **6:**2537–2544.
35. **Lee, K. A. W., A. Bindereif, and M. R. Green.** 1988. A small-scale procedure for preparation of nuclear extracts that support efficient transcription and pre-mRNA splicing. Gene Anal. Tech. **5:**22–31.
36. **Leib-Mösch, C., M. Bachmann, R. Brack-Werner, T. Werner, V. Erfle, and R. Hehlmann.** 1992. Expression and biological significance of human endogenous retroviral sequences. Leukemia **6:**72s–76s.
37. **Liu, A. Y., and B. A. Abraham.** 1991. Subtractive cloning of a hybrid human endogenous retrovirus and calbindin gene in the prostate cell line PC3. Cancer Res. **51:**4107–4110.
38. **Löwer, R., K. Boller, B. Hasenmaier, C. Korbmacher, N. Müller-Lantzsch, J. Löwer, and R. Kurth.** 1993. Identification of human endogenous retroviruses with complex mRNA expression and particle formation. Proc. Natl. Acad. Sci. USA **90:**4480–4484.
39. **Löwer, R., J. Löwer, C. Tondera-Koch, and R. Kurth.** 1993b. A general method for the identification of transcribed retrovirus sequences (R-U5 PCR) reveals the expression of the human endogenous retrovirus loci HERV-H and HERV-K in teratocarcinoma cells. Virology **192:**501–511.
40. **Löwer, R., R. R. Tönjes, C. Korbmacher, R. Kurth, and J. Löwer.** 1995. Identification of a Rev-related protein by analysis of spliced transcripts of the human endogenous retroviruses HTDV/HERV-K. J. Virol. **69:**141–149.
41. **Luckow, B., and G. Schütz.** 1987. CAT constructions with multiple unique restriction sites for the functional analysis of eukaryotic promoters and regulatory elements. Nucleic Acids Res. **15:**5490.
42. **Lundin, M., J. O. Nehlin, and H. Ronne.** 1994. Importance of a flanking AT-rich region in target site recognition by the GC box-binding zinc finger protein MIG1. Mol. Cell. Biol. **14:**1979–1985.
43. **Mager, D. L.** 1989. Polyadenylation function and sequence variability of the long terminal repeats of the human endogenous retrovirus-like family HERV-H. Virology **173:**591–599.
44. **Mager, D. L., and N. L. Goodchild.** 1989. Homologous recombination between the LTRs of a retrovirus-like element causes a 5-kb deletion in two siblings. Am. J. Hum. Genet. **45:**848–854.
45. **Mager, D. L., and P. S. Henthorn.** 1984. Identification of a retrovirus-like repetitive element in human DNA. Proc. Natl. Acad. Sci. USA **81:**7510–7514.
46. **Majello, B., P. De Luca, G. Hagen, G. Suske, and L. Lania.** 1994. Different members of the Sp1 multigene family exert opposite transcriptional regulation of the long terminal repeat of HIV-1. Nucleic Acids Res. **22:**4914–4921.
47. **Medstrand, P., M. Lindeskog, and J. Blomberg.** 1992. Expression of human endogenous retroviral sequences in peripheral blood mononuclear cells of healthy individuals. J. Gen. Virol. **73:**2463–2466.
48. **Mueller-Lantzsch, N., M. Sauter, A. Weiskircher, K. Kramer, B. Best, M. Buck, and F. Grässer.** 1993. Human endogenous retroviral element K10 (HERV-K10) encodes a full-length gag homologous 73-kDa protein and a functional protease. AIDS Res. Hum. Retroviruses **9:**343–350.
49. **Parvin, J. D., R. J. McCormick, P. A. Sharp, and D. E. Fisher.** 1995. Prebending of a promoter sequence enhances affinity for the TATA-binding factor. Nature (London) **373:**724–727.
50. **Ptashne, M.** 1988. How eukaryotic transcriptional activators work. Nature (London) **335:**683–689.
51. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
52. **Sauter, M., S. Schommer, E. Kremmer, K. Remberger, G. Dölken, I. Lemm, M. Buck, B. Best, D. Neumann-Häfelin, and N. Mueller-Lantzsch.** 1995. Human endogenous retrovirus K10: expression of gag protein and detection of antibodies in patients with seminomas. J. Virol. **69:**414–421.
53. **Seed, M. R., and J. Y. Sheen.** 1988. A simple phase-extraction assay for chloramphenicol acyltransferase activity. Gene **67:**271–277.
54. **Smale, S. T., M. C. Schmidt, A. J. Berk, and D. Baltimore.** 1990. Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. Proc. Natl. Acad. Sci. USA **87:**4509–4513.
55. **Sogawa, K., H. Imataka, Y. Yamasaki, H. Kusume, H. Abe, and Y. Fujii-Kuriyama.** 1993. cDNA cloning and transcriptional properties of a novel GC box-binding protein, BTEB2. Nucleic Acids Res. **21:**1527–1532.
56. **St.-Arnaud, R., and J. M. Moir.** 1993. *Wnt*-1-inducing factor-1: a novel G/C box-binding transcription factor regulating the expression of *Wnt*-1 during neuroectodermal differentiation. Mol. Cell. Biol. **13:**1590–1598.
57. **Sypes, M. A., and D. S. Gilmour.** 1994. Protein/DNA crosslinking of a TFIID complex reveals novel interactions downstream of the transcription start. Nucleic Acids Res. **22:**807–814.
58. **Vogetseder, W., A. Dumfahrt, P. Mayersbach, D. Schönitzer, and M. P. Dierich.** 1993. Antibodies in human sera recognizing a recombinant outer membrane protein encoded by the envelope gene of the human endogenous retrovirus K. AIDS Res. Hum. Retroviruses **9:**687–694.
59. **Wilkinson, D. A., J. D. Freeman, N. L. Goodchild, C. A. Kelleher, and D. L. Mager.** 1990. Autonomous expression of HERV-H endogenous retrovirus-like elements in human cells. J. Virol. **64:**2157–2167.
60. **Wilkinson, D. A., N. L. Goodchild, T. M. Saxton, S. Wood, and D. L. Mager.** 1993. Evidence for a functional subclass of the HERV-H family of human endogenous retrovirus-like sequences. J. Virol. **67:**2981–1989.
61. **Wilkinson, D. A., D. L. Mager, and J. C. Leong.** 1994. Endogenous human retroviruses, p. 465–535. *In* J. A. Levy (ed.), The Retroviridae, vol. 3. Plenum Press, New York.
62. **Williams, T., and R. Tjian.** 1991. Analysis of the DNA-binding and activation properties of the human transcription factor AP-2. Genes Dev. **5:**670–682.