Capture of a Cellular Transcriptional Unit by a Retrovirus: Mode of Provirus Activation in Embryonal Carcinoma Cells

CLAIRE BONNEROT, EDITH LEGOUY,[†] ANDRÉ CHOULIKA, AND JEAN-FRANÇOIS NICOLAS*

Unité de Biologie moléculaire du Développement, Institut Pasteur, Unité Associée 1148 du Centre National de la Recherche Scientifique, 25 rue du Dr. Roux, 75724 Paris Cedex 15, France

Received 8 January 1992/Accepted 6 May 1992

The expression of murine leukemia provirus in embryonal carcinoma (EC) cells is blocked by a mechanism still incompletely understood. The blockage is not overcome by deleting a large portion of the enhancer region (in U3) in recombinant retroviruses (M-MuLVneo Δ Enh). This confirms the presence of negative elements outside the viral 82-bp repeats. However, a few sites in the genomes of EC cells permit M-MuLVneo Δ Enh proviral expression. One such site, identified in PCC4, PCC3, and LT, was studied. The complete analysis of the mechanism of activation by Northern (RNA) blotting, cloning, and sequencing of partial cDNA copies of the viral transcript and of the site of integration establishes that viral transcripts are initiated from an upstream host-cell promoter and are spliced from a host donor to a cryptic viral acceptor at position 542 in the Moloney murine leukemia virus (M-MuLV) genome. In consequence, the mature transcripts are host cell-virus fusion transcripts from which M-MuLV sequences, including the *cis*-active negative elements of the 5' long terminal repeat-containing region, are absent. The provirus integrates apparently randomly into any of the three most proximal introns of the transcriptional unit. The host cell promoter contains a TATA box and 14 potential SpI binding sites included in a 1.0-kb GC-rich island. These elements promote gene expression of recombinant vectors in EC and differentiated cells. The mechanism described points to a mechanism by which retroviruses can be transcribed from upstream nonviral elements and can acquire host genes by 5' annexation of exons.

Retroviruses are very versatile genetic elements which account for as much as 1 to 3% of the genome (41, 44). The viral forms integrate into the genomes of infected cells by intercalation of a DNA copy of their own genome into the host DNA (14, 44). Their presence in the genome can cause alteration of the host genes by insertion, mutation, and activation in *cis* of adjacent genes (including proto-oncogenes). The proviral forms of retroviruses can also transpose by a conservative mechanism via an intermediary RNA (12). Their excision from the genome by homologous recombination (7, 42) may lead to structural rearrangements, usually deletions, at the integration site. Finally, the genome of the virus mutates at a high rate and is susceptible to rearrangement (37). One of these rearrangements with important consequences is the capture of sequences downstream of the provirus. This mechanism has been suggested as being central to the generation of defective oncogenic retroviruses (3).

The analysis of defective oncogenic retroviruses has emphasized their potential as naturally occurring genetic vectors and has served as a guide to their exploitation as engineered vehicles for the introduction of genes into cells (6, 23). Retroviruses have been transduced into both somatic and germinal cells to mutagenize the genome (11) or to serve as genetic cell markers (35). Although neither of these uses necessitates expression of the integrated DNA, other applications require it. In somatic cells, expression is generally readily obtained either from the viral long terminal repeat (LTR) or from internal promoters. In embryonic cells, however, expression cannot be obtained from the viral LTR because of the absence of transcription from the U3 region

(17). During the last 4 years, the complexity of this negative regulation has become evident. Negative control elements within the 82-bp repeats (1, 8, 9, 17) and others mapped 5' to the 82-bp repeats (39) have been described. A silencer has also been discovered in the leader sequence of the virus (18, 26).

Despite the multiplicity of negative controls of viral transcription in embryonal carcinoma (EC) cells, rare cases in which viral expression occurs have been detected when selective methods with neomycin-based recombinant retroviruses have been used (2, 29, 36). In the nullipotent F9 EC cell line, Barklis et al. have demonstrated that resistant clones fall into two distinct classes (2). In one, viral expression correlates with a mutation in the silencer in the leader sequence of the virus. In the other, the virus is wild type and the expression is due to a chromosomal position effect. Only a few sites in the genome of F9 cells permit viral expression. On the basis of the sequence of the integration site, Peckham et al. have demonstrated that in two cases, the recombinant retrovirus integrated into the first introns of cellular transcription units in close proximity to an active promoter (25). They hypothesized that viral transcription is initiated in the 5'-flanking region.

We report the analysis of the mechanism which in PCC4, a multipotential EC cell line (21), activates an enhancer-minus M-MuLVneo recombinant retrovirus (22). The analysis reveals the presence of a region capable of activating the provirus. A cDNA corresponding to a cell-virus transcript has been cloned from the resistant clones. The comparison of its 5' sequence with the sequence of the site of integration establishes that the mechanism of activation is by transcription initiation in a 5'-flanking promoter, readthrough of the 5' viral LTR, and splicing from a cellular donor to a cryptic acceptor in Moloney murine leukemia virus (M-MuLV) sequences. These results demonstrate that subgenomic viral sequences can be expressed from a cellular promoter. This

* Corresponding author.

[†] Present address: Unité de Biologie du Développement-INRA, 78352 Jouy-En-Josas Cedex, France.

mechanism, therefore, is reciprocal to cellular oncogene activation by viral LTRs. These results also suggest the possibility of the capture of an upstream transcriptional unit by retrovirus, therefore increasing their potential to acquire cellular genes for transduction.

MATERIALS AND METHODS

Plasmid DNA. pM-MuLVneo Δ Enh is similar to pM-MuLVneo (29) except that the 3' LTR is deleted of sequences from nucleotide 7929 to nucleotide 8113 (44).

pABnslacZ and pBnslacZ contain, respectively, from 5' to 3', the AB fragment or the B fragment of the 1.7-kb *Sph*I region (Fig. 8); the *Sal*I-to-*Bam*HI fragment from pM-MuLVnslacZ Δ Enh (4) containing the nslacZ reporter gene; and the *Bcl*I-to-*Bam*HI fragment (nucleotides 2753 to 2716) of simian virus 40 containing the polyadenylation signal (38). Bacterial sequences are from pTZ18-1R (gift of J. Cébrian).

pLTRA contains, from 5' to 3', sequences from M-MuLV (that is, the LTR from nucleotide 8113 and the leader sequences up to nucleotide 563 [44]), sequences of the nslacZ reporter gene, and the *Bcl*I-to-*Bam*HI fragment of simian virus 40 containing the polyadenylation signal (38). Bacterial sequences are from pTZ18-1R. pGB contains, from 5' to 3', the BCD fragment of the 1.7-kb *Sph*I region (see Fig. 8) combined with the M-MuLV and nslacZ sequences as in pLTRA.

pBSGA Δ LTR was constructed as follows. A three-piece ligation was performed with the *Hind*III-*Pst*I fragment of pBSKSII+, the 2.1-kb *Hind*III-*Sal*I fragment of pAB containing the regions A and B, and the 5.6-kb *Sal*I-*Pst*I fragment of pM-MuLVnslacZ Δ Enh containing nslacZ and a polyadenylation signal (4).

Production of retrovirus and infections. pM-MuLVneo Δ Enh (10 μ g) was precipitated with calcium phosphate and used to transfect ψ 2 cells by the method of Graham and Van der Eb (10) except that the precipitate was directly added to the culture medium. pM-MuLVneo Δ Enh carries the gene encoding resistance to aminoglycoside G418. Stable ψ 2 transformants were selected as previously described (29).

Clones were tested for viral production as follow. Culture supernatants were prepared and added to 3T3-WOP cells in the presence of 10 μ g of Polybrene per ml. The 3T3-WOP cells were then tested for resistance to G418. Mouse EC cell lines were infected with 1 or 10 ml of virus stock added to 10^6 cells in 10-cm plates. The samples were incubated at 37°C for 18 h in the presence of 5 μ g of Polybrene per ml, and then the medium was replaced with fresh medium. One day after viral infection, 5×10^5 cells were transferred to 10-cm plates. One day after this transfer, the medium was replaced with medium containing 500 μ g of G418 per ml and was changed every 2 days until only G418-resistant cells were present. The G418-resistant colonies were counted 10 days after the beginning of the G418 selection. Individual clones were obtained by sucking them up into a 2-ml pipette and transferring them to 1-cm culture plates.

Cell culture. PCC3, PCC4, LT, PCC7-S (16), and 3T3-WOP fibroblasts were grown in Dulbecco's modified Eagle's medium containing high concentrations of glucose (4.5 g/liter) and glutamine (0.584 g/liter) and 10% fetal calf serum. Cultures were incubated at 37°C in a humidified air-CO₂ atmosphere. Except for 3T3-WOP fibroblasts and LT cells, which were detached by a mixture of EDTA and trypsin, cells were replated by pipetting. For other details, see reference 16.

Southern and Northern (RNA) hybridization analyses. Ge-

nomic DNA for Southern blot hybridization was prepared from cell lines by the method referred to in reference 16. High-molecular-weight DNA (10 μ g) was digested with 10 U of various restriction endonucleases and electrophoresed in 0.8% agarose gels. The DNA was transferred to nylon membranes. Total RNA for Northern blot hybridization was prepared by standard methods (31). RNA was fractionated in a formaldehyde-agarose gel and transferred to nylon membranes. Probes for hybridizations were radiolabelled by the random primer method. DNA-DNA and DNA-RNA hybridizations were carried out according to the method of reference 5.

Transfections and β -galactosidase assays. The day before transfection, cells were seeded at a density of 10^5 for 3T3-WOP and 3×10^4 for PCC4, LT, and PCC7-S per 35-mm dish. Calcium phosphate DNA precipitates were prepared by the standard procedure (46). Precipitate (200 μ l) containing 2 μ g of nslacZ reporter plasmid to be tested, 1 μ g of tkLuciferase plasmid as a control of the efficiency of transfection, and 7 μ g of pTZ18-1R plasmid as a DNA carrier was added directly to the culture medium of cells. After overnight incubation, the calcium phosphate DNA precipitate was removed and fresh medium was added to the cells, which were incubated for an additional 24 h. Cells were then harvested, and crude extracts were prepared and assayed for β -galactosidase and luciferase activity. The *O*-nitrophenyl- β -D-galactopyranoside (ONPG) assay of protein extracts was performed as described previously (30).

Luciferase analysis. One Bradford unit of protein extract was diluted in luciferase buffer (25 mM triphosphate [pH 7.8], 8 mM MgCl₂, 1 mM EDTA, 2 mM dithiothreitol, 0.3 mM ATP, 0.1 mM luciferin [p-luciferin; 4,5 dihydro-2,6-hydroxy-2-benzothiazolyl-1,4-thiazolecarboxylic acid] [Sigma]). Luminescence was counted on a Lumat LB 9501 (Berthold) for 10 s.

Sequencing strategy. The sequences of both strands of pTZ- or KS⁺-derived plasmids were determined. Sequencing procedures with commercial sequencing kits were done according to the manufacturers' instructions. Several primers were synthesized to complete sequences on both strands to obtain sequences overlapping all restriction cloning sites.

Lambda cloning. High-molecular-weight DNA was prepared from clone PCC4-10-5 by using the standard procedure (31). The DNA was partially digested with *Mbo*I and size fractionated on an agarose gel. DNA fragments ranging from 15 to 25 kb were recovered from the gel and ligated to EMBL3cos DNA (45) digested with *Bam*HI. The library was amplified through *Escherichia coli* NM646. This modified cytosine restriction (Mcr⁻) *E. coli* strain increased the yield of recombinant phage 10-fold. The library was screened with a *neo* probe, and genomic probes surrounding the retrovirus were subsequently used to isolate the unrearranged allele (from the same library).

cDNA preparation and amplification by PCR. Total RNA was extracted from PCC4-10-5. It was used for synthesis of cDNA by using reverse transcriptase from avian myeloblastosis virus (Gibco BRL) and a primer from *neo*, 5'-AAGCG GCCGAGAACC-3', named oligo a (Fig. 1). After a passage through a Sepharose CL6B spin column, the cDNA was tailed by using the terminal transferase in the presence of dGTP. The tailed cDNA was then amplified by polymerase chain reaction (PCR) with a second primer from *neo*, 5'-GCGGATCCTCATCCTGTCTCTTG-3', named oligo b (Fig. 1), and a poly(C) primer, 5'-CGGGATCCCCCCCC-3', both connected to a *Bam*HI adapter. Thirty cycles of amplification were performed, with denaturation at 95°C for

FIG. 1. Structure of the recombinant retrovirus. The shaded box represents the *neo* gene sequences, open boxes represent LTR sequences (U3, R, and U5), the triangle above the 3' LTR indicates the U3 deletion, and numbers refer to the nucleotide sequences in M-MuLV (44). PBS, tRNA primer binding site; PPT, polypurine tract; Ψ , packaging sequences; D, donor splice site for the *env* RNA; I, initiation codon in *neo* RNA; T, termination codon in *neo* RNA.

1 min, annealing at 60°C for 2 min, and elongation at 72°C for 3 min. The products were digested with *Bam*HI restriction enzyme and subcloned into the Bluescript vector pBSK-SII+. DNAs from bacterial clones transformed with the resulting product were transferred to nylon membranes. Hybridization with oligonucleotide probes from *neo*, 5'-CAGATCTCGACCTGCAGCAGAC-3', named oligo c (Fig. 1), was used to identify positive clones. DNA sequencing was carried out directly on double-stranded plasmid clones by the dideoxy chain termination method (31).

Amplification of genomic sequences. The genomic DNAs of various PCC4 clones resistant to G418 were amplified by PCR with the oligo b primer from *neo* and the following primer from the first exon in the 1.7-kb *Sph*I region: 5'-GGGCTCGAGGAGCTGTGCGGCATTCTG-3' (see Fig. 4). Thirty cycles of amplification were performed, with denaturation at 95°C for 1 min, and annealing and elongation at 72°C for 3 min 30 s. The products were electrophoresed on a 1% agarose gel. After being transferred onto positive membranes (from Appligene), they were hybridized to an oligonucleotide (5'-CATAAACCTTCATCCTCC-3') from the third exon and to oligo c from *neo*.

Microinjection in fertilized eggs. Fertilized eggs were obtained from the oviducts of (C57BL/6J $\times$ DBA/2) F_1 females mated with F_1 males of the same strain. Manipulation of eggs and microinjection were carried out as previously described (13). Eggs were microinjected with 2,000 copies of the appropriate construct at 12 to 14 h postfertilization. Expression was analyzed 30 h after microinjection by histochemical staining with 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-Gal) as a substrate of β -galactosidase. The construct tested for activity was an insert deleted of all plasmid DNA sequences. pBSGA Δ LTR was digested with *Not*I and *Xho*I restriction enzymes. The resulting 8.1-kb DNA fragment was purified on glass beads before microinjection.

RESULTS

The recombinant retroviruses. As described above, it has been established that the absence of M-MuLV expression in EC cells is due to several elements: two or more *cis* sequences within and outside the 82-bp repeats in U3, which are potential targets for repressors, and a silencer in the leader sequences. To minimize the number of conditions to be overcome to obtain proviral expression in EC cells, the enhancer regions of M-MuLV promoter (17) were deleted to give a vector named pM-MuLVneo Δ Enh, which is replication defective.

The pM-MuLVneo Δ Enh vector was constructed from a

Moloney murine leukemia provirus (15) by replacing the *Pst*II (base 563 in M-MuLV)-*Bam*HI (base 6537 in M-MuLV) fragment with a neomycin resistance gene and by deleting the two 82-bp tandem repeats of U3 in the 3' LTR (bases 7929 to 8113), which contain enhancer sequences (17). The vector (Fig. 1) therefore transfers an enhancerless neomycin resistance gene. Other viral elements in the provirus are, from 5' to 3', the leader sequences (including the R region), U5, the primer-binding site (bases 146 to 163), the splice donor site for *env* gene mRNA (nucleotide 206 in M-MuLV), and the 5' noncoding region up to the *Pst*II site (base 563). Downstream of the *Pst*II site are bacterial sequences coding for the neomycin resistance gene followed by part of the *env* region and the enhancerless 3' LTR.

Virus-producing cell lines were generated by transfecting pM-MuLVneo Δ Enh or pM-MuLV-neo, a vector with unmodified U3 (29), into the ψ 2 packaging cell line (19). Titers from virus-producing cell lines were estimated by dot blotting viral nucleic acids in the cell supernatants (data not shown). One transfected cell line (each producing a similar quantity of virus) was retained for each vector. Virus was prepared from the culture medium of each cell line, and the preparations were titrated by infecting multipotential EC cells (PCC3, PCC4, and PCC7-S) and fibroblastic cells (3T3-WOP) (Table 1). The titer of M-MuLVneo in 3T3-WOP cells was five times higher than that of M-MuLVneo Δ Enh. The deletion of the enhancers thus appears to be easily complemented by cellular sequences (43) and/or to decrease the expression of the *neo* gene but to levels still sufficient for resistance of the cells to G418. In contrast, in PCC3 and PCC4 cells, the titers of both viruses were similarly low. In PCC7-S, the titers of both viruses were intermediate (Table 1). Therefore, we conclude that the deletion in the viral regulatory sequences does not completely inactivate the inhibition of viral expression in EC cells (a conclusion also

TABLE 1. Titers of virus produced by ψ 2 cell lines^a

Cell line	CFU of G418 ^r virus/ml	
	M-MuLVneo Δ Enh	M-MuLVneo
PCC3	0.1	5
PCC4	2-4	1
PCC7-S	2×10^2	7×10^2
3T3-WOP	4×10^3	2×10^4

^a Virus stocks were prepared by incubating 10 ml of medium with 2×10^6 ψ 2-producing cells for 18 h. Infection and selection were as described in Material and Methods. Resistant clones were counted after 10 days. Titers are for 1 ml of medium.

supported by the mapping of negative control elements outside the viral enhancers [18, 26, 39]).

In PCC4, one G418-resistant clone was obtained for every 3×10^5 cells infected at a virus-to-cell multiplicity of 2 (see the discussion of Southern blot hybridization experiments below). Assuming that all regions of the host genome are equally accessible to the virus, this corresponds to one virus for every 8 kb of the genome (genome size $[4.6 \times 10^6 \text{ kb}]$ divided by the number of integration sites required to obtain one G418-resistant clone $[6 \times 10^5]$). This low efficiency makes it possible to classify the resistant clones into two categories: those integrating into distinct regions of the host genome where the provirus is active presumably because of mutations in its viral sequences and those integrating into the same region of the host genome (preferential expression sites) where the provirus is active presumably because of complementation by cellular sequences. Therefore, we analyzed several independent integration events in PCC4 to test for the presence of preferential expression sites in EC cells.

A preferential site of expression common to three EC lines: PCC3, PCC4, and LT. Structures of the integrated M-MuLV neo Δ Enh in 17 independent geneticin-resistant PCC4 clones were analyzed by Southern blot hybridization. The recombinant retrovirus with a complete 5' LTR contains a 2.9-kb *EcoRV* fragment (Fig. 2B). Total DNA of each of the 17 clones was therefore digested with *EcoRV* and probed with *neo* (Table 2). There was no 2.9-kb fragment in any of the samples. Therefore, the proviruses had all successfully transferred the modified 3' U3. The number of *EcoRV* fragments indicates the number of independent integration events. Each of these fragments corresponds to one *EcoRV* site in the provirus and one *EcoRV* site in the upstream flanking chromosomal DNA, and thus the size of such a fragment varies according to the position of the nearest chromosomal *EcoRV* site. Six clones had one, eight clones had two, and three clones had more than three additional fragments. This distribution frequency indicates a multiplicity of infection of two viruses per cell, assuming that all cells in the population are equally infectable. However, in view of the low frequency of existing clones, it is probable that only one copy of the virus is active in each cell. The conclusions concerning transfer of the deleted 3' U3 and the number of provirus were confirmed by a similar analysis of *PvuII* digest (Table 2 and map in Fig. 2B). In addition, these experiments indicate that there were no major sequence rearrangements in the recombinant retroviruses.

To test for the presence of a family of integration sites in the genome, *EcoRI* endonuclease, which does not cut within the viral sequence, was used to digest cellular DNAs, and the digests were probed with the *neo* probe. In 12 clones of 17, the cleavage fragments generated were of identical sizes (6 kb; Table 2 and examples in Fig. 2A; note the presence of fragments of variable sizes in clones infected with more than one virus), indicating that the virus had integrated into the same region of the host genome in all 12 clones. For confirmation, *SphI* endonuclease, which has a single site in the provirus (in *neo* sequences; Fig. 2B), was used. Thus, if proviruses integrated between the same two cellular *SphI* recognition sites, they should generate two fragments with the same total length. Indeed, the 12 clones which gave an *EcoRI* fragment of 6 kb appeared to have integrated into a *SphI* region 1.7 kb long (the sum of the two *SphI* fragments is 5.3 kb, and M-MuLVneo Δ Enh provirus is 3.6 kb long) (Table 2 and examples in Fig. 2A). No other family of integration sites was detected.

A similar analysis with the DNA of G418-resistant clones

FIG. 2. Southern blot analysis of Neo^r EC cell lines and physical map of the 1.7-kb *SphI*-containing region. (A) Structure of integrated provirus in PCC4 and PCC3 EC cell lines. Total cellular DNA (approximately 10 μ g per lane) digested with *EcoRI* (a noncutter in proviral DNA) or *SphI* (a single cutter in proviral DNA) was fractionated in agarose gels, blotted onto nylon filters, and hybridized to a DNA probe from within *neo* (Fig. 1, probe *neo*). Clones PCC4-B5 and PCC3-B11 are from infection by M-MuLVneo; other clones are from infection by M-MuLVneo Δ Enh. The *EcoRI* fragment common to clones from infection by M-MuLVneo Δ Enh is 6 kb, and the *EcoRI* fragment common to clones from infection by M-MuLVneo is 6.4 kb (arrows at left). For DNA digested with *SphI*, note that the sum of the sizes of the two fragments indicated by arrowheads is exactly 5.3 kb for clones infected by M-MuLVneo Δ Enh and 5.7 kb for clones infected by M-MuLVneo. (B) Scheme of the M-MuLVneo Δ Enh provirus, illustrating the positions of the two LTRs and pertinent restriction enzyme sites. The sizes of the *PvuII* fragments (1.1 and 0.76 kb) and the *EcoRV* fragment (2.9 kb) are indicated. The positions of the viral enhancers are indicated by a shaded box. (C) Restriction map of flanking DNA. Arrows indicate the sites of integration of the provirus. In panels B and C, the vertical bars represent endonuclease restriction sites: B, *BamHI*; E, *EcoRV*; H, *HindIII*; P, *PvuII*; R, *EcoRI*; S, *SphI*; X, *XbaI*.

of LT infected with M-MuLVneo Δ Enh indicates that the frequency of integrants in the 1.7-kb *SphI* region was significantly lower (2 of 30) as determined from Southern blots of cellular DNA hybridized with radiolabelled DNA of the integration site (see below). This suggests that there are other families of sites of preferential expression in this cell line. Alternatively, the blockage of proviral expression in LT may not involve all the elements used in PCC4, and the

TABLE 2. Restriction enzyme analysis of G418^r EC cell lines^a

G418 ^r clone ^b	n ^c	Size(s) (kb) of fragment(s) produced by ^d :							
		<i>Pvu</i> II	<i>Eco</i> RV	<i>Eco</i> RI	<i>Sph</i> I ^e	<i>Bam</i> HI	<i>Bam</i> HI- <i>Eco</i> RI	<i>Hind</i> III	<i>Xba</i> I
PCC4-10-2	1	1, 0.76	6.4	6	<u>1.95</u> , <u>3.05</u>	9.6	4	8.6	3.2
PCC4-10-5	1	1.1, 0.76	6.4	6	<u>2.2</u> , <u>3.1</u>	9.6	4.2	8.6	3.2
PCC4-10-13	1	3.4, 0.76	6	6	<u>2</u> , <u>3.4</u>	10	3.9	9	3.2
PCC4-10-8	1	>15, 0.76	9.2	6		6.5	3.5	6.5	3.2
PCC4-10-12	1	3.5, 0.76	7.4	>10		3.2	3.2	>10	>10
PCC4-10-18	1		5.4	>10		1.9	2.2	>10	3.2
PCC4-10-3	2	3.3, 5, 0.76	4.7, 6	6, >10	3.3, 2.1	9.6, 4.7	3.9, 4.9	8.6, 5.2	3.2
PCC4-10-4	2	1.1, 3, 0.76	6.4, 8	6, 7.5	<u>2.2</u> , <u>3.2</u>	9.6, 3.6	3.6, 4.2	8.6, 10	3.2
PCC4-10-6	2	1.7, 3.1, 0.76	6, 10	6, 5.6	<u>1.9</u> , <u>3.5</u>	9.6, 6.4	3.9, 3.5		3.2
PCC4-10-9	2		6.4, 4.7	6, >10	<u>2.1</u> , <u>3</u>	9.8, 4.8	4, 5	9.4, 5.2	3.2
PCC4-10-19	2		6, >10	6, >10	<u>2.1</u> , <u>3</u>	5.5, 10	4.3, 5.6	9.8, 9.2	3.2
PCC4-10-21	2		6, >10	6, >10	<u>2.1</u> , <u>3</u>	4, 10	4.3, 4	9.8, 9.2	3.2
PCC4-10-7	2	2.8, 0.76*	8.6, >10	10, >10		14*	2.9, 7.6	>10, >10	3.2
PCC4-10-10	2	9, 0.76*	6, 8.3	7.6, 9.6		6.2, 3.2	2.8, 3.2	6.5, >10	3.2
PCC4-10-1	3		4.3, 6, >10	6, 7.5*	<u>2.6</u> , <u>2.8</u>	9.6, 3.8*	4.5, 4.2, 3.8	4.9, 8.6, 10	3.2
PCC4-10-17	3	6.9, 0.76*	3.3, 6.6, >10	6, >10, >10		3.8, 12*	2.3, 4.5, 10	5.2, 8.6, 10	3.2
PCC4-10-16	5		5.8, >10	6, 6.5, 7.5	<u>1.6</u> , <u>3.5</u>	9.4, 8, 7.2, 4.2, 3.2	4.5, 3.9, 3.4*	5, 5.8, 6.8, 8, 8.6	3.2
			>10*	10*					
PCC4-B3	1			8		7	6	>10	
PCC4-B4	1			4.5		3.8	2.7	4.5	
PCC4-B5	1			6.4	<u>2.5</u> , <u>3.2</u>	10	4.5	9.6	
PCC3-B11	1			6.4	<u>2.4</u> , <u>3.3</u>	10	4.4	9.6	
PCC3-B12	1			>10		2.5	2.6	>10	
PCC3-B13	1			>10		13	10	5.6	

^a Total cellular DNA (10 µg) digested with the indicated restriction enzyme(s) was fractionated in agarose gels, blotted onto nylon filters, and hybridized to a DNA probe from within *neo* (Fig. 1, probe *neo*). *Bam*HI, *Eco*RV, *Pvu*II, *Sph*I, and *Xba*I cut proviral DNA. *Eco*RI and *Hind*III do not cut proviral DNA (Fig. 2B). The restriction map of the flanking DNA derived from this analysis is in Fig. 2C.

^b Clones with presumably no proviral integration in the 1.7-kb *Sph*I region are in italics.

^c Number of proviral copies.

^d Stars indicate missing fragments.

^e For DNA digested with *Sph*I, the sum of the two underlined fragments is 5.3 kb from infection by M-MuLVneoΔEnh and 5.7 kb from infection by M-MuLVneo. Therefore, the size of the *Sph*I region where the viruses appeared to have integrated is 1.7 kb (M-MuLVneoΔEnh is 3.6 kb long). Only the relevant fragments of the clones integrated in the 1.7-kb *Sph*I region are indicated.

requirement for viral mutations capable of overcoming this blockage may be, in consequence, less stringent.

To test whether the deletion of the 82-bp repeats in U3 in M-MuLVneoΔEnh was a requirement for provirus activation in the 1.7-kb *Sph*I region, we searched for provirus integrated in the 1.7-kb *Sph*I region in PCC4 and PCC3 infected with M-MuLVneo. With both cell lines, one integrant of the three mapped in the 1.7-kb *Sph*I region (Fig. 2A and Table 2), indicating that the deletion in the 5' LTR is not a requirement for provirus activation in the 1.7-kb *Sph*I region.

Figure 2C summarizes the restriction mapping of the sites of integration of 11 resistant PCC4 clones infected with M-MuLVneoΔEnh and 2 resistant EC clones infected with M-MuLVneo as deduced from analyses of the lengths of the *Sph*I fragments in Fig. 2A and of additional digestions (Table 2). All proviruses integrated in the same orientation, and there does not appear to be a hot spot of integration in the region.

Mechanism of activation of the provirus. The transcription of provirus sequences in three G418-resistant EC clones, PCC4-10-5, PCC4-B5, and PCC3-B11 (Table 2) integrated into the 1.7-kb *Sph*I region was analyzed by Northern blot hybridization (Fig. 3). The 3T3-WOP 10-3 clone infected with M-MuLVneoΔEnh expressed the expected 3.3-kb provirus transcripts, whereas PCC4-10-5 expressed transcripts of 3 kb, about 300 nucleotides shorter than the viral genome. Similarly, 3T3-WOP B1, a clone infected with M-MuLVneo, expressed the expected transcripts of 3.5 kb, whereas

PCC4-B5 and PCC3-B11 from infections with M-MuLVneo expressed transcripts of 3.2 kb, again about 300 nucleotides shorter than the virus. No additional transcripts hybridizing to *neo* were detected even after prolonged exposure. There-

FIG. 3. Northern blot analysis of viral transcripts. Total RNA was fractionated in 1% formaldehyde-agarose gels, transferred to nylon filters, and hybridized to a *neo*-specific DNA probe. Lane 1, 3T3-WOP B1 (5 µg) from M-MuLVneo infection; lane 2, 3T3-WOP 10-3 (10 µg) from M-MuLVneoΔEnh infection; lane 3, PCC4-10-5 (25 µg) from M-MuLVneoΔEnh infection; lane 4, PCC4-B5 (25 µg) from M-MuLVneo infection; lane 5, PCC3-B11 (25 µg) from M-MuLVneo infection. The 3.5-kb marker is indicated by the arrowhead.

FIG. 4. Sequence comparison between the cDNA copy of the hybrid transcript and the DNA of the preintegration target site. (A) The PCR product of reverse-transcribed RNA from PCC4-10-5 amplified by using a primer in *neo*. The black box represents the *neo* sequence, and the grey box represents the M-MuLV sequence (nucleotides 542 to 563). The three nucleotides in M-MuLV preceding the junction are indicated. Open boxes 1 to 3 represent nonviral sequences. (B) Map of the insertion. The line represents 1.5 kb of genomic DNA upstream of proviral sequences. The positions of host sequences within the PCR product are indicated as open boxes. The position of a TATA box is indicated by a vertical bar. (C) Sequences at the boundaries of boxes 1 to 3. Note the concordance with conserved 5' donor (AG/GT) and 3' acceptor (YAG/G) splice signals.

fore, PCC4-10-5, PCC4-B5, and PCC3-B11 do not express transcripts of genomic size.

To clarify this observation, total RNA was extracted from PCC4-10-5, and reverse transcriptase was used to prepare cDNA from a primer hybridizing to *neo* sequences on the sense transcript (oligo a, Fig. 1). The cDNA samples were treated with terminal transferase to add a poly(G) tail. Then, they were amplified by using two primers, one that bound to *neo* sequences (oligo b, Fig. 1) and one that bound to the poly(G) tail. The products were cloned and then sequenced by the chain termination technique.

We examined the nucleotide sequences of several clones obtained from the PCR experiments. Nucleotide sequence identity between the PCR product and the provirus is represented in Fig. 4A. The PCR product contained two viral DNA segments that contained, from 3' to 5' *neo* sequences, M-MuLV sequences from base 563 (*Pst*I site) to base 542, and nonviral sequences (115 nucleotides). The viral sequences are interrupted 3' at an AG site. This suggests RNA splicing from a host cell donor to a viral acceptor and presumably transcription from a cellular promoter.

To determine whether the nonviral sequences originate from the integration site, genomic sequences flanking the provirus were subcloned from a partial genomic library of PCC4-10-5 (see Materials and Methods), and 3 kb were sequenced. The nonviral sequences in the PCR products correspond to three putative exons of 30, 53, and 32 bp separated by introns of 561 and 175 bp, respectively, in the genome (Fig. 4B). The exon-to-exon junctions strictly respect the GT-AG rule, and the cellular acceptor and donor splice sites conform to the consensus sequences (34) (Fig. 4C). In addition, a typical UACUAAC branching sequence (34) is located 18 to 25 nucleotides from the 3' splice site of the first intron.

Examination of the cellular sequence revealed additional features (Fig. 5). A TATA box is located 26 nucleotides upstream of the 5'-most exon (see also Fig. 4B), but no CAAT box is present in this upstream region. The upstream region is over 80% G+C, with CpG approximately as frequent as GpC (CpG to GpC ratio is 0.61), meeting the criteria for a CpG island. There are 14 putative SpI binding sites regularly spaced. Octamer motifs are not present.

In addition, there were no mutations in the provirus in PCC4-10-5. The sequence of the 5' LTR confirms the transfer of the U3ΔEnh from 3' to 5'. The mechanism of retroviral integration (40) appears to have been respected. In particular, the invariant structural features of host-provirus DNA junctions are observed: the terminal 2 bp (ApA and TpT) from each end of the precursor are lost, and 4 bp of host DNA (ApC-pApT) flanking the integrated provirus are duplicated (data not shown). Sequencing of the preintegration site shows that the target sequences are otherwise not modified (not shown).

Restriction mapping suggested that the clones integrated in the 1.7-kb *Sph*I region are scattered at many sites in the region (Fig. 2C). To determine conclusively whether they were distributed indifferently in the three introns in the 1.7-kb *Sph*I region, the proviral integration points were precisely mapped in each of 12 PCC4 clones. The distance from the first exon to the provirus was measured by using PCR. PCR amplification of the DNA fragment polymerized from an oligonucleotide complementary to sequences in the first exon and an oligonucleotide complementary to *neo* sequences (oligo b, Fig. 1) indicates random integration (Fig. 6) in the first three introns.

We conclude that the provirus is activated because of its position in one of the first three introns downstream from a

```

GAGCCCCGCGCAAGGAGTTGCGGACAGGTAAGGCGCGGACCGGGAAGGCGCGGATCGGTCTAAGG
CGGACTGCGCGGCGGTGTGTCTGGGTGTGGGCGCGCGCGGGTCTGAAGGAGGGGTCTGCCAAGGGGGG
CGGCGACTGCGAGCGCCGCAAGCCAAAGCGGGAAGGTCTGCGAGTCTCGGAGGCTCGGGCAGCACTGAGCC
CCGAGACTGCTGCAACCCAGACCGCGGCGGACGACGAGTACGCTAACGCGCGGGGACAGGGAACGAGGCCAGTC
AAGGAGCACCAGCACCAGCGGCGGGAATGGGCGGAACTGGGCGCGGACGAGCTTCCCGAAGTGAAACCCCGG
CTCTTCGCGAGCGAGCGAGCTGCGGGAAGGAATCAGTACCCCTCCCGCGGCTCCCGGCGAGCTTTCTTGCAGAGCC
CCCTCCCTCAGCGCTGACTCTCTCAGAGGCCCAAGCGGCTATGGTGCCGCGGAACTCGGCTCCGGCTCTCGGGGC
GTGGCCAGAGGGAAGTTTGTGGGCTGAAGAAGGTGGGCTTGAAGGAGAGTCTGAGGCGGTGTGAGCGGCGGTTT
AAGTGGACGAGCAAGCAGCTTTGCGTCTTGACGTACACTCAAGAGCCCTATATCTCTGACTCAAGCCGCGATCTCAG
CTTTTGGAGCTGTGCGGCACTCTGAGCAGATGCAATTTCTGAGTATGTTATGGCTT
+1

```

FIG. 5. Sequence of the promoter in the 1.7-kb *Sph*I region. The most 5' nucleotide of the cell-virus cDNA was designated +1. The putative TATA box and potential SpI binding sites are underlined. The sequence in the open box represents the first exon.

FIG. 6. Location of the integration points of M-MuLVneoΔEnh (LTRΔEnh) and M-MuLVneo (LTR) in PCC4 clones. PCR products from amplification with an oligonucleotide complementary to sequences in the first exon of the cellular gene and an oligonucleotide complementary to sequences in the proviral *neo* gene (see Materials and Methods) were transferred to a nylon membrane and hybridized with an oligonucleotide complementary to sequences in the third exon (oligo 3) or to sequences in the *neo* gene (oligo c; Fig. 1). (A) Autoradiograms. The sizes indicated on both sides of the gels are the sizes of the smaller (1.2 kb) and larger (2.3 kb) PCR products. (B) Sites of integration of the 12 proviruses in the 1.7-kb *SphI* region. Black boxes 1 to 3 indicate positions of the exons.

cellular promoter. The provirus uses the cellular promoter to transcribe a chimeric RNA. By splicing, cellular exons are joined to a viral exon and thus remove major viral control sequences.

Targeted gene is autosomal. Probes from cellular DNA flanking the provirus in PCC4-10-5 were used to determine the copy number of the gene activating the integrated provirus and serving as the target of viral integration in the genome. Two probes free of repetitive sequences (1 and 2, Fig. 7B) were used to probe DNA from PCC4-10-5 and from PCC4 digested by *HindIII* restriction enzyme in Southern blot analysis (Fig. 7A). Two bands of approximately 5.0 and 8.0 kb in PCC4-10-5 and one band of 5.0 kb in PCC4 were detected. The 8.0-kb band was shown to correspond to the region with the proviral insertion by hybridization with *neo* probes. The 5.0-kb band detected corresponds to the uninterrupted fragment in the allelic gene. Therefore, the 1.7-kb *SphI* region marks a single locus in the haploid genome. As PCC4 is karyotypically XO (21), the region is autosomal.

Promoter of the identified gene is active in EC cells, various differentiated cells, and early embryos. To assay directly for promoter function of the genomic DNA, a 2-kb *HindIII*-*NcoI* fragment (AB) and a 1.2-kb *EcoRI*-*NcoI* fragment (B) were inserted upstream of the *nlsLacZ* gene (Fig. 8, I). The resulting recombinants were named pABnlsLacZ and pBnlsLacZ. pLTRΔnlsLacZ, in which an enhancer-minus LTR was linked to the reporter gene, and ptknlsLacZ, in which the promoter of the thymidine kinase gene of the herpes simplex virus drives *nlsLacZ*, were used as controls. Transient expression in EC and differentiated cells was performed by following the protocol detailed in Materials and Methods. The results were normalized for variations of efficiency of DNA uptake and of expression by reference to

FIG. 7. Southern blot analysis of cellular DNA. (A) Southern transfer of DNA. Cellular DNA was prepared from PCC4 cells and PCC4-10-5. The DNA was subjected to Southern analysis after digestion with *HindIII*. (B) Localization of probes 1 and 2 relative to the proviral insertion. Restriction endonuclease sites are abbreviated as follows: H, *HindIII*; N, *NcoI*; B, *BamHI*; R, *EcoRI*. The proviral insertion does not contain *HindIII* recognition sequences.

the expression of a cotransfection reference plasmid, ptkLuciferase. In Fig. 8, the structures of the plasmids and the levels of β -galactosidase activities in PCC4 are presented.

Plasmid pABnlsLacZ promoted the transcription of the *lacZ* gene (Fig. 8, compare AB with LTRΔ). To compare the respective promoter functions of this plasmid and other LacZ plasmids, we calculated their β -galactosidase activities relative to that of the enhancerless pLTRΔLacZ plasmid. In PCC4, the AB fragment activated *lacZ* gene transcription (in pABnlsLacZ) to a 1,148-fold increase. Roughly similar levels of increased activation (862-fold) were obtained with the 1.2-kb B subfragment in pBnlsLacZ (Fig. 8). We conclude that promoter elements are contained in the B region, and this is in agreement with the sequence data.

To assess the specificity of the promoter elements, the recombinant plasmids pABnlsLacZ and pBnlsLacZ were transfected into another EC cell line and various differentiated cell lines. In EC cells (LT), fibroblasts (3T3), myoblasts (PCD1), osteogenic cells (KDP1 E6), and neuroblastoma cells (NS20Y and N18) (for references, see reference 32), the *lacZ* gene was transcribed at a high level. Finally, a recombinant ABnlsLacZ DNA fragment (from pBSGAΔLTR; see Materials and Methods) was microinjected into fertilized oocytes. β -Galactosidase activity was assessed by X-Gal staining of late two-cell embryos. Six of six of the eggs (100%) expressed the fragment. Therefore, the region can

FIG. 8. Functional assays for promoter and enhancer activities at the integration site. (I) The top line is a map of the 5' portion of the integration site. Restriction endonuclease sites are abbreviated as follows: H, *Hind*III; E, *Eco*RI; N, *Nco*I; S, *Sca*I. (II) Structures of *lacZ* fusion constructs. The reporter gene is nlsLacZ (shaded box) followed by the polyadenylation signal of simian virus 40. Sequences present in the vector are represented by open rectangles; those absent from the vector are represented by dashed lines. In pABnlsLacZ (map AB) and pBnlsLacZ (map B), host genomic sequences are directly fused to reporter sequences. It tests for promoter activity. In the LTRΔ control plasmid, viral sequences present in M-MuLVneoΔ drive nlsLacZ. pGBnlsLacZ (map GB) is the integration site plus proviral sequences combined with nlsLacZ. In the tknlsLacZ (map tk) control plasmid, the promoter of thymidine kinase of the herpes simplex virus (nucleotides -200 to +50 [20]) drives nlsLacZ. (III) The constructs were introduced into EC cells (PCC4) to assess their capacities for expression. The levels of β-galactosidase activity normalized for transfection efficiency of pABnlsLacZ, pBnlsLacZ, pLTRΔnlsLacZ, and pGBnlsLacZ are indicated.

function as a promoter but does not confer cell type specificity on the reporter gene.

To test the efficiency of the M-MuLV 5' LTR as a transcriptional terminator for RNA initiated from an upstream distant nonviral promoter, an additional plasmid was constructed. pGBnlsLacZ contains part of the viral transcriptional unit, from the 5' U3 to the *Pst*I restriction site (base 563 in M-MuLV; Fig. 1), inserted between nlsLacZ and the BCD fragment (Fig. 8, II). The region upstream of nlsLacZ is therefore identical to the region upstream of *neo* in PCC4-10-5. Data from transfection experiments using this clone show that pGBnlsLacZ produces higher amounts of β-galactosidase in PCC4, i.e., about 171-fold more than pLTRΔnlsLacZ (compare GB and LTRΔ in Fig. 8). Thus, with the relatively small difference in β-galactosidase activity between pGBnlsLacZ and pBnlsLacZ, the data indicate that viral sequences have poor transcriptional terminator functions in EC cells.

DISCUSSION

Mechanism of viral activation. This study establishes that provirus activation in PCC4 cells can occur through a very specific mechanism: the capture of a transcriptional unit. The model is as follows. The provirus integrates into an intron of an active gene. Next, it utilizes an upstream cellular promoter to synthesize a transcript which passes through the polyadenylation signal in the R region of the 5' LTR of the provirus. The transcript is spliced from a cellular donor to a viral acceptor (Fig. 9). This is the first demonstration of this model, which has previously been postulated by Barklis et al. to explain activation of provirus by the site of integration in F9 cells (2, 25).

There is one central prediction of the model, which is the presence of host cell-virus fusion transcripts in the infected resistant cells. Their production necessitates the presence of a functional promoter upstream from the provirus and poor efficiency of the 5' enhancer-minus LTR as a transcriptional terminator. The central prediction is validated by the cloning and analysis of the sequence of a hybrid transcript isolated from PCC4 EC cells infected by an enhancer-minus M-MuLVneo recombinant retrovirus (PCC4-10-5). The cel-

lular sequences in the hybrid transcript are made up of three discontinuous fragments in the host chromosome. The fragments are linked in the RNA by splicing with consensus acceptor, branch point, and donor sequences. The three exons are fused to viral sequences whose 5' boundaries are defined by a cryptic splice acceptor in M-MuLV sequences.

The proposed mechanism of viral activation is probably not restricted to PCC4, as the same region was identified in PCC3 and LT. Also, it does not require the deletion of the enhancers in U3, as in PCC3 and PCC4 the region can complement M-MuLVneo recombinant retrovirus.

Conditions for activation. The capture of a cellular transcriptional unit by a provirus requires several conditions. First, as the viral transcript is initiated from an upstream cellular position, proviral integration must be in an active gene on the 3' side of its promoter. This is validated, in this work, by the analysis of the sequence of the genomic DNA 5' from the integration point of the retrovirus and by biological tests for promoter function of the DNA. In the

FIG. 9. Mechanism of activation of provirus in EC cells. (A) Proviral DNA integrates on the 3' side of a gene in the same transcriptional orientation. (B) Readthrough transcription from the cellular promoter (P) generates a hybrid RNA which terminates at the polyadenylation site in R of the 3' LTR. Splicing from a cellular donor to a viral acceptor signal fuses host exons to viral sequences.

genomic DNA, there is a TATA box 26 bp from the putative initiation point of transcription and a GC-rich region containing numerous binding sites for the ubiquitous transcriptional factor Sp1. This region combined with a reporter gene promotes its expression in cells. These observations raise the question of whether the cellular promoter is active before the integration of the recombinant retrovirus. In PCC4, transcripts hybridizing with exonic sequences in the 1.7-kb *SphI* region are detected (unpublished data). Similarly, in F9, transcripts of the three regions identified by an M-MuLV recombinant retrovirus are detected (27). It indicates that usually the genes tagged by M-MuLVneo Δ Enh or M-MuLVneo are active before provirus integration. There are other demonstrations that retrovirus can integrate into active regions, and certain data suggest that active regions may even constitute preferred targets for integration (28). Second, transcription through the 5' LTR, which contains a polyadenylation signal in the R region, must be possible. It has been experimentally tested. We found that the efficiency of the 5' enhancer-minus LTR as a transcriptional terminator for transcripts initiated from a distant nonviral promoter is low (at least within the context sequence of the 1.7-kb *SphI* region and in EC cells). We do not know whether this is restricted to the 1.7-kb *SphI* region and if transcription through the 5' LTR requires reduced or null transcriptional activity from the LTR. The analysis of specially designed recombinant retroviruses would help answer these questions.

In cells with a provirus integrated at a position which fulfills these two conditions, hybrid transcripts could be produced. However, according to the termination-reinitiation model for translation of mammalian-cell mRNAs (24), the translation of the *neo* gene from these transcripts would occur only if the termination of translation initiated at the first AUG in cellular sequences is close to the AUG initiator of the *neo* product. In the LTRs of M-MuLV with or without enhancer sequences, terminators for the three open reading frames are close to the 5' end of U3, and the next AUG initiators are more than 300 nucleotides distant. This separation may constitute a barrier to the expression of genes internal to the virus. The selection of clones of PCC4 and F9 in which the *neo* gene is expressed may therefore select those in which splicing links the chromosomal donor to an acceptor close to the *neo* gene, thus removing a number of initiation codons. Consequently, those genes which have small 5' exons dispersed over several hundreds of base pairs are more easily detected as preferential expression regions.

This third condition necessitates a splice acceptor in viral sequences. In M-MuLVneo Δ Enh provirus, the acceptor splice sequence revealed by these experiments is AAG/C at position 542 in M-MuLV. The sequence fits the 3' splice site consensus (34) and is preceded by a stretch of pyrimidines located 30 nucleotides upstream adjacent to a possible branch point sequence, GUCUGAA, that is consistent with the YNYURAX branch point consensus (34). These elements are not known to be functional in M-MuLV and therefore are probably cryptic in other circumstances. If the mechanism proposed for proviral activation in PCC4-10-5 also applies to clones integrated in the first and second introns, then there is no incompatibility between this cryptic acceptor splice sequence and the different 5' donor splice regions of presumably different affinities.

Trapping of a preferential region of expression in PCC4. The results presented in this article do not shed light on the mechanism of blockage of M-MuLV in EC cells. Other studies have established that the promoter of M-MuLV does not function in EC cells because of the presence of repressor

sequences within the 82-bp repeats and 5' to them (at -345) and because of the lack of positive activation. An EC-specific repression binding site has been mapped to nucleotides +147 to +174. Clearly, in the 1.7-kb *SphI* region, none of these elements act on the upstream cellular promoter sufficiently to inhibit expression. Reciprocally, the elements of the upstream cellular promoter do not seem to activate the viral LTR, as no RNA corresponding to the viral genome size was detected. This may be because of the absence of enhancers in the cellular promoter or the dominance of the negative controls from the virus on cellular enhancers.

Another question is why only a few regions able to complement M-MuLVneo Δ Enh were identified in PCC4. Presumably, it is because the mechanism for the activation of the provirus requires that a number of local conditions, which have already been discussed, be fulfilled (integration into an active transcriptional unit, production of readthrough transcripts, a structure of the transcripts compatible with the translation of an active *neo* product). In addition, there may be less-obvious conditions resulting from the mechanism of the negative regulation of the LTR in multipotential cells such as, for example, a negative action of the silencer at the RNA level. Finally, although the 1.7-kb *SphI* region does not show the features characteristic of a known preferred integration site (in particular, independent integration events did not occur at exactly the same base [33]), it remains possible that this region is one of the group of as-yet-undefined preferred integration sites.

On the other hand, in differentiated cells, retrovirus can be activated by a different mechanism based on complementation by cellular enhancers. A large number of positions in the genome can complement expression of M-MuLVneo Δ Enh by this mechanism (43). The intermediate values of complementing positions found in PCC7-S (Table 1) could be due to two mechanisms: (i) the repression of M-MuLV expression is different in LT and PCC7-S compared with PCC4, allowing activation to occur by complementation by cellular enhancers as described for differentiated cells, and (ii) the number of active regions enabling the mechanism described in PCC4 is increased.

Finally, it is unclear whether this phenomenon is unique to M-MuLV. It would be interesting to analyze whether other retroviruses, and in particular bovine leukemia virus, human T-cell leukemia virus, and human immunodeficiency virus, can be activated by host cell promoters. Indeed, this mechanism could be the basis of expression, in cells harboring an otherwise silent proviral copy, of viral products, including oncogenes and molecules with potential for controlling the genes of the host cell.

ACKNOWLEDGMENTS

E.L. gratefully acknowledges C. Bishop for the gift of Embl3cos vector and for helpful suggestions. We acknowledge C. Castelle for performing experiments on expression of plasmids in cell lines.

This work was supported by the Centre National de la Recherche Scientifique (Unité Associée 1148), the Institut National de la Santé et de la Recherche Médicale (grant 871011), the Association pour la Recherche sur le Cancer (grants 6622 and 6744), and the Ligue contre le Cancer. C.B. and J.-F.N. are from INSERM.

REFERENCES

1. Akgün, E., M. Ziegler, and M. Grez. 1991. Determinants of retrovirus gene expression in embryonal carcinoma cells. *J. Virol.* 65:382-388.
2. Barklis, E., R. C. Mulligan, and R. Jaenisch. 1986. Chromosomal position or virus mutation permits retrovirus expression in embryonal carcinoma cells. *Cell* 47:391-399.

3. Bishop, J. M., and H. E. Varmus. 1982. Functions and origins of retroviral transforming genes, p. 246–356. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), *RNA tumor viruses*, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
4. Bonnerot, C., M. Vernet, G. Grumber, P. Briand, and J. F. Nicolas. 1991. Transcriptional selectivity in early mouse embryos—a qualitative study. *Nucleic Acids Res.* **19**:7251–7257.
5. Church, G. M., and W. Gilbert. 1984. Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **81**:1991–1995.
6. Coffin, J. 1985. *RNA tumor viruses: molecular biology of tumor viruses*, p. 17–75. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
7. Copeland, N. G., K. W. Hutchison, and N. A. Jenkins. 1983. Excision of the DBA ecotropic provirus in dilute coat-color revertants of mice occurs by homologous recombination involving the viral LTRs. *Cell* **33**:379–387.
8. Feuer, G., M. Taketo, R. C. Hanecak, and H. Fan. 1989. Two blocks in Moloney murine leukemia virus expression in undifferentiated F9 embryonal carcinoma cells as determined by transient expression assays. *J. Virol.* **63**:2317–2324.
9. Gorman, C. M., P. W. J. Rigby, and D. P. Lane. 1985. Negative regulation of viral enhancers in undifferentiated embryonic stem cells. *Cell* **42**:519–526.
10. Graham, F. L., and A. J. van der Eb. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* **52**:456–467.
11. Hartung, S., R. Jaenisch, et M. Breindl. 1986. Retrovirus insertion inactivates mouse alpha 1 (I) collagen gene by blocking initiation of transcription. *Nature (London)* **320**:365–366.
12. Heidmann, T., O. Heidmann, and J. F. Nicolas. 1988. An indicator gene to demonstrate intracellular transposition of defective retroviruses. *Proc. Natl. Acad. Sci. USA* **85**:2219–2223.
13. Hogan, B., F. Costantini, and E. Lacy. 1986. *Manipulating the mouse embryo. A laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
14. Hughes, S. H., P. R. Shank, D. H. Spector, H. J. Kung, J. M. Bishop, H. E. Varmus, P. K. Vogt, and M. L. Breitman. 1978. Proviruses of avian sarcoma virus are terminally redundant, co-extensive with unintegrated linear DNA and integrated at many sites. *Cell* **15**:1397–1410.
15. Jaenisch, R., H. Fan, and B. Crocker. 1975. Infection of preimplantation mouse embryos and newborn mice with leukemia virus: tissue distribution of viral DNA and RNA and leukemogenesis in the adult animal. *Proc. Natl. Acad. Sci. USA* **72**:4008–4012.
16. Jakob, H., and J. F. Nicolas. 1987. Mouse teratocarcinoma cells. *Methods Enzymol.* **151**:66–81.
17. Linney, E., B. Davis, J. Overhauser, E. Chao, and H. Fan. 1984. Non-function of a Moloney murine leukaemia virus regulatory sequence in F9 embryonal carcinoma cells. *Nature (London)* **308**:470–472.
18. Loh, T. P., L. L. Sievert, and R. W. Scott. 1990. Evidence for a stem cell-specific repressor of Moloney murine leukemia virus expression in embryonal carcinoma cells. *Mol. Cell. Biol.* **10**:4045–4057.
19. Mann, R., R. C. Mulligan, and D. Baltimore. 1983. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. *Cell* **33**:153–160.
20. McKnight, S., and R. Tjian. 1986. Transcriptional selectivity of viral genes in mammalian cells. *Cell* **46**:795–805.
21. Nicolas, J. F., P. Avner, J. Gaillard, J. L. Guénet, H. Jakob, and F. Jacob. 1976. Cell lines derived from teratocarcinomas. *Cancer Res.* **36**:4224–4231.
22. Nicolas, J. F., C. Bonnerot, C. Kress, H. Jouin, P. Briand, P. Grumber, and M. Vernet. 1989. Visualization by nlsLacZ of gene activity during mouse embryogenesis. *In* *Vectors as tools for the study of normal and abnormal growth and differentiation*. NATO ASI Ser. Ser. H **34**:33–45.
23. Nicolas, J. F., and J. Rubenstein. 1987. Retroviral vectors, p. 493–513. *In* J. E. Davies (ed.), *Vectors: a survey of molecular cloning vectors and their uses*. Butterworths, Boston.
24. Peabody, D. S., and P. Berg. 1986. Termination-reinitiation occurs in the translation of mammalian cell mRNAs. *Mol. Cell. Biol.* **6**:2695–2703.
25. Peckham, I., S. Sobel, J. Comer, R. Jaenisch, and E. Barklis. 1989. Retrovirus activation in embryonal carcinoma cells by cellular promoters. *Genes Dev.* **3**:2062–2071.
26. Petersen, R., G. Kempler, and E. Barklis. 1991. A stem cell-specific silencer in the primer-binding site of a retrovirus. *Mol. Cell. Biol.* **11**:1214–1221.
27. Petersen, R., S. Sobel, C.-T. Wang, R. Jaenisch, and E. Barklis. 1991. Cellular transcripts encoded at a locus which permits retrovirus expression in mouse embryonic cells. *Gene* **101**:177–183.
28. Rohdewohld, H., H. Weiher, W. Reik, R. Jaenisch, and M. Breindl. 1987. Retrovirus integration and chromatin structure: moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *J. Virol.* **61**:336–343.
29. Rubenstein, J., J. F. Nicolas, and F. Jacob. 1984. Construction of a retrovirus capable of transducing and expressing genes in multipotential embryonic cells. *Proc. Natl. Acad. Sci. USA* **81**:7137–7140.
30. Rubenstein, J. L. R., J. F. Nicolas, and F. Jacob. 1984. L'ARN non sens (nsARN): un outil pour inactiver spécifiquement l'expression d'un gène donné in vivo. *C.R. Acad. Sci. Paris Ser. III* **299**:271–274.
31. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
32. Sanes, J., J. Rubenstein, and J. F. Nicolas. 1986. Use of a recombinant retrovirus to study post-implantation cell lineage in mouse embryos. *EMBO J.* **5**:3133–3142.
33. Shih, C. C., J. P. Stoye, and J. M. Coffin. 1988. Highly preferred targets for retrovirus integration. *Cell* **53**:531–537.
34. Smith, C. W. J., J. G. Patton, and B. Nadal-Ginard. 1989. Alternative splicing in the control of gene expression. *Annu. Rev. Genet.* **23**:527–577.
35. Soriano, P., and R. Jaenisch. 1986. Retroviruses as probes for mammalian development: allocation of cells to the somatic and germ cell lineages. *Cell* **46**:19–29.
36. Taketo, M., E. Gilboa, and M. I. Sherman. 1985. Isolation of embryonal carcinoma cell lines that express integrated recombinant genes flanked by the Moloney murine leukemia virus long terminal repeat. *Proc. Natl. Acad. Sci. USA* **82**:2422–2426.
37. Temin, H. M. 1989. Evolution of retroviruses, p. 189–196. *In* A. L. Notkins and M. B. A. Oldstone (ed.), *Concepts in viral pathogenesis*. III. Springer-Verlag, New-York.
38. Toozé, J. 1980. *DNA tumor viruses. Molecular biology of tumor viruses*, 2nd ed., part 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
39. Tsukiyama, T., O. Niwa, and K. Yokoro. 1991. Analysis of the binding proteins and activity of the long terminal repeat of Moloney murine leukemia virus during differentiation of mouse embryonal carcinoma cells. *J. Virol.* **65**:2979–2986.
40. Varmus, H. 1988. Regulation of HIV and HTLV gene expression. *Genes Dev.* **2**:1055–1062.
41. Varmus, H., and P. Brown. 1989. Retroviruses, p. 53–108. *In* D. E. Berg and M. M. Howe (ed.), *Mobile DNA*. American Society for Microbiology, Washington, D.C.
42. Varmus, H. E., N. Quintrell, and S. Ortiz. 1981. Retroviruses as mutagens: insertion and excision of a nontransforming provirus alter expression of a resident transforming provirus. *Cell* **25**:23–36.
43. von Melchner, H., and H. E. Ruley. 1989. Identification of cellular promoters by using a retrovirus promoter trap. *J. Virol.* **63**:3227–3233.
44. Weiss, R., N. Teich, H. Varmus, and J. Coffin. 1985. *RNA tumor viruses. Molecular biology of tumor viruses*, 2nd ed., part 2. Supplements and appendixes. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
45. Whittaker, P. A., J. B. Campbell, E. M. Southern, and N. E. Murray. 1988. Enhanced recovery and restriction mapping of DNA fragments cloned in a new λ vector. *Nucleic Acids Res.* **16**:6725–6737.
46. Wigler, M., S. Silverstein, L.-S. Lee, A. Pellicer, Y. Cheng, and R. Axel. 1977. Transfer of purified herpes virus thymidine kinase gene to cultured mouse cells. *Cell* **11**:223–233.