JOURNAL OF VIROLOGY, Aug. 1989, p. 3227–3233
0022-538X/89/083227-07$02.00/0

Vol. 63, No. 8

Identification of Cellular Promoters by Using a Retrovirus Promoter Trap

HARALD VON MELCHNER* AND H. EARL RULEY

Center for Cancer Research, Massachusetts Institute of Technology, 40 Ames Street, Cambridge, Massachusetts 02139

Received 2 February 1989/Accepted 11 April 1989

A retrovirus vector has been developed that selects for instances in which the provirus integrates in close proximity to cellular promoters. Coding sequences for a selectable marker (histidinol dehydrogenase) were inserted into the 3′ long terminal repeat (LTR) of an enhancerless Moloney murine leukemia virus. Although 1.8 kilobase pairs in size, the elongated LTR did not appear to interfere with virus replication or integration. Thus, when the virus was passaged, the elongated LTRs efficiently duplicated, placing histidinol dehydrogenase-coding sequences in the 5′ LTR just 30 nucleotides from the flanking cellular DNA. Selection for histidinol expression generated cell clones in which histidinol gene sequences in the 5′ LTR were invariably expressed on transcripts initiating in nearby cellular sequences. The efficiency of transducing histidinol resistance was 2,500-fold lower than the efficiency of transducing neomycin resistance when the neomycin phosphotransferase gene was located within the retrovirus and expressed from an independent promoter. By tagging transcriptionally active sites, the vector provides a means to identify and isolate promoters active in different cell types. Furthermore, the virus may be useful as an insertional mutagen, since selection for cell populations containing proviruses only in expressed sites is expected to reduce the number of integrants needed to screen for loss of gene function.

Retroviruses replicate by a complex life cycle involving (i) conversion of the viral RNA genome into double-stranded DNA by reverse transcriptase, (ii) integration of the viral DNA (or provirus) into the genome of the infected cell, (iii) transcription of the provirus as a cellular gene, and (iv) packaging of provirus transcripts into virions. Before integration, sequences near the 5′ (U5) and 3′ (U3) ends of the viral RNA duplicate such that the integrated provirus is flanked by long terminal repeats (LTRs), each containing the U3 and U5 regions (42, 44).

This life cycle enables retroviruses to move genes into the genomes of mammalian cells. Like other movable elements, retroviruses have been used both as probes for transcriptionally active chromosomal regions and as insertional mutagens (1, 4, 23, 24, 39, 41, 43, 44). However, several factors have undermined the practical use of retroviruses as genetic tools to study mammalian organisms. First, large genomes (3×10^9 nucleotides [nt]) necessitate screening large numbers of integrants in order to detect mutations in any specific gene. Second, mutations resulting from provirus integration are generally recessive, since most mammalian genomes are diploid. Third, enhancers in the LTRs may influence the expression of adjacent genes and thus interfere with detecting cellular sequences that regulate transcription in a tissue-specific manner. Finally, 3′ RNA processing signals and AUG codons within the lefthand LTR interfere with activation of provirus genes by nearby cellular promoters (9, 15, 16, 21, 41).

To address these problems, we have developed a retrovirus vector that selects for instances in which the provirus integrates in close proximity to a cellular promoter. By tagging transcriptionally active genes, the virus provides a mean to identify and isolate promoters active in different cell types. In this manner, the virus functions as a promoter trap. In addition, the vector may be an effective mutagen, since selection for cell populations containing proviruses integrated only in expressed sites is expected to reduce, by several orders of magnitude, the number of integrants needed to screen for the loss of gene function.

The strategy involved inserting a selectable marker (histidinol dehydrogenase [*hisD*]) into the 3′ LTR of an enhancerless Moloney murine leukemia virus (MoMuLV). The elongated LTR efficiently duplicated when the virus was passaged, placing *hisD* coding sequences in the 5′ LTR just 30 nt from the flanking cellular DNA. In this manner, integration removes intervening viral sequences that normally interfere with transcriptional activation of provirus genes by cellular promoters; since it lacks an enhancer, the provirus is not expected to affect expression of adjacent cellular sequences. The virus also contained a second selectable marker, neomycin phosphotransferase (*neo*), expressed from an internal promoter, to provide an independent measure of virus infectivity. The vector transduced histidinol resistance 2,500-fold less efficiently than neomycin resistance. Moreover, in each histidinol-resistant (His^r) line examined, *his* sequences in the 5′ LTR were expressed by transcripts initiating at a nearby cellular promoter.

MATERIALS AND METHODS

Plasmids. GgU3Hisen(−) and GgTKNeoU3Hisen(−) viruses were constructed from GgTKNeoen(−), a recombinant MoMuLV (P. Robbins and R. Mulligan, unpublished data). GgTKNeoen(−) was derived from pHSG-neo (18); it contains sequences extending from the 5′ LTR to the *Xho*I site at nt 1558, a bacterial *neo* gene expressed from the herpes simplex virus thymidine kinase (*tk*) promoter (32), and provirus sequences extending from the *Cla*I site (nt 7672) through the 3′ LTR and lacks sequences between the *Pvu*II and *Xba*I sites (nt 7933 to 8111) that contain the viral

* Corresponding author.

transcriptional enhancer. Coding sequences for the *Salmonella typhimurium hisD* gene, isolated from pSP1 (28) as a 1,350-nt *Bam*HI fragment, were ligated to the *Nhe*I site (nt 7846) in U3 of GgTKNeoen(−) to obtain GgTKNeo U3Hisen(−). GgU3Hisen(−) was derived from GgTKNeo U3Hisen(−) by deleting the *Bam*HI fragment containing the *tk* promoter and the *neo* gene.

Cells and viruses. Cells were grown in Dulbecco modified Eagle medium supplemented with 10% calf serum (NIH 3T3 and Ψ2 [31]) or 10% fetal calf serum (PA317 [33]). Cell lines producing recombinant retroviruses were derived after transfecting (29) Ψ2 or PA317 cells with plasmid DNAs and selecting in medium containing G418 (1 mg/ml; GIBCO Laboratories). After 10 days, G418-resistant (Neo^r) colonies were isolated and expanded in mass culture. To prepare virus stocks, 2×10^6 cells from each clone were seeded into 100-mm-diameter dishes. The next day, 2 ml of fresh medium was added, and after 2 h the medium was filtered through a 0.22-μm-pore-size membrane (Millipore Corp.) and stored at −70°C until use. Samples (1 ml) of different dilutions of virus stocks were added to 10^5 NIH 3T3 cells plated 1 day before infection. After incubation for 1 h at 37°C in the presence of Polybrene (8 μg/ml; Aldrich Chemical Co.), 9 ml of fresh medium was added. After incubating overnight, cells were grown for 10 days in selective medium containing 1 mg of G418 per ml or 4 mM L-histidinol (Sigma Chemical Co.). Colonies were fixed (10% [vol/vol] formaldehyde in phosphate-buffered saline) and stained with crystal violet before counting.

Southern hybridization analysis. Genomic DNA, extracted from His^r or Neo^r NIH 3T3 lines, was digested to completion with restriction endonucleases *Cla*I, *Cla*I and *Sal*I, *Hin*dIII and *Nde*I, fractionated in 1% (wt/vol) agarose gels, and transferred to nylon membranes (Zetabind; Cuno) as described previously (30). Blots were hybridized to ^{32}P-labeled probes prepared from the 1.35-kilobase-pair (kbp) *his* coding sequence by the random priming method (12).

Northern (RNA) hybridization analysis. Total cellular RNA was extracted from His^r or Neo^r NIH 3T3 cell lines by the guanidinium thiocyanate procedure. RNA was fractionated on formaldehyde-agarose gels and transferred to nylon membranes (GeneScreen Plus; Dupont, NEN Research Products) by electroblotting in 25 mM phosphate buffer. DNA-RNA hybridizations were carried out as previously described (45), using ^{32}P-labeled restriction fragments corresponding to *his* [1.35-kbp *Nhe*I; GgU3Hisen(−)], *neo* [1.2-kbp *Bgl*II-*Bam*HI; GgTKneoen(−)], and *gag* [1.1-kbp *Pvu*I-*Xho*I; GgU3Hisen(−)] sequences.

RNase protection assays. Cellular RNA (20 μg) was hybridized at 55°C for 10 h to ^{32}P-labeled RNA probes (4×10^5 cpm) in 30 μl of 80% (vol/vol) formamide, 0.4 M NaCl, 0.04 M piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (PIPES; pH 6; Sigma), and 0.001 M EDTA. Probes complementary to the provirus coding strand were prepared by using T3 RNA polymerase (Promega Biotec) to transcribe an 860-nt *Bam*HI-*Cla*I fragment of GgTKNeoU3Hisen(−) cloned into Bluescript KS(+) (Stratagene). After hybridization, samples were digested with RNases A (Boehringer Mannheim Biochemicals) and T1 (Bethesda Research Laboratories, Inc.) and processed for gel electrophoresis according to the instructions supplied by Promega Biotec. Protected fragments were separated on denaturing 6% polyacrylamide–8.3 M urea gels and visualized by autoradiography.

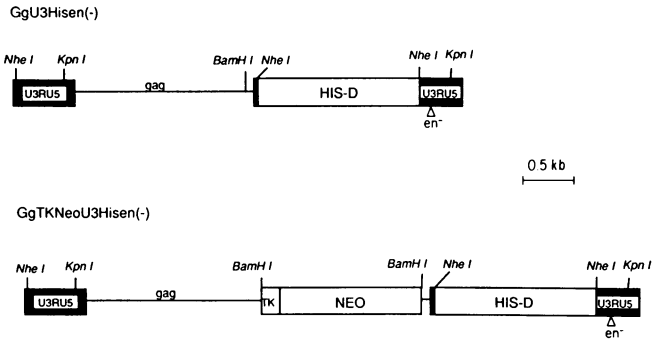

FIG. 1. Structures of recombinant retrovirus vectors. Gene sequences present in each virus are labeled as follows: HIS-D, *S. typhimurium* histidinol dehydrogenase; NEO, neomycin phosphotransferase; gag, truncated *gag* region of MoMuLV; TK, herpes simplex virus *tk* promoter. en$^-$ designates a 178-nt deletion in the 3′ LTR encompassing the viral enhancer.

RESULTS

LTR-mediated gene duplication. The retroviruses used in this study are shown in Fig. 1. GgU3Hisen(−) and GgTK NeoU3Hisen(−) were constructed from a Moloney murine leukemia provirus (18) by inserting the *hisD* coding sequence from *S. typhimurium* (*hisD*; 20) into the U3 region of the righthand LTR and by deleting enhancer sequences of the LTR (18). GgTKNeoU3Hisen(−) also contains a neomycin resistance gene under the transcriptional control of the herpes simplex virus *tk* promoter to provide an independent measure of virus titers.

Virus-producing cell lines were generated by transfecting GgTKNeoU3Hisen(−) into NIH 3T3 cells expressing packaging-defective ecotropic (Ψ2) and amphotropic (PA317) helper viruses. Viruses recovered from cloned producer lines were titered on NIH 3T3 cells, selecting in either G418 or L-histidinol (Table 1). Titers of the GgTKNeoU3Hisen(−) virus were high and similar to what we and others have obtained with other MoMuLV vectors (2, 5, 6, 18, 31, 33; data not shown), suggesting that the insertion of *his* sequences into the LTR did not markedly interfere with virus replication or integration. As is generally observed, ecotropic Ψ2 cells produced higher yields of virus than did amphotropic PA317 cells (5, 31, 33); however, the ratio of *his* to *neo* colony-forming units was similar for each virus stock, about 3.8×10^{-4}.

Structures of the integrated GgTKNeoU3Hisen(−) proviruses in six independent Neo^r and His^r clones were analyzed

TABLE 1. Titers of GgTKNeoU3Hisen(−) viruses produced by cloned PA317 and Ψ2 cell lines[a]

Clone	Titer (CFU/ml per 10^7 producer cells)		Ratio (10^{-4})
	Histidinol	Neomycin (10^4)	
PAC3	40	13	3.0
PAC7	50	18	2.9
PAC9	125	50	2.5
ΨC7	625	140	4.5
ΨC9	1,030	190	5.4
ΨC10	750	160	4.7

[a] PA317 and Ψ2 cells were transfected with pGgTKNeoU3Hisen(−), and Neo^r clones were selected in G418. Virus stocks were prepared by incubating 2 ml of medium with 2×10^6 cells of each Neo^r clone. NIH 3T3 cells (10^5) were exposed to various dilutions of virus and subsequently placed into G418 or L-histidinol selective medium. After 10 days, resistant colonies were stained and counted.

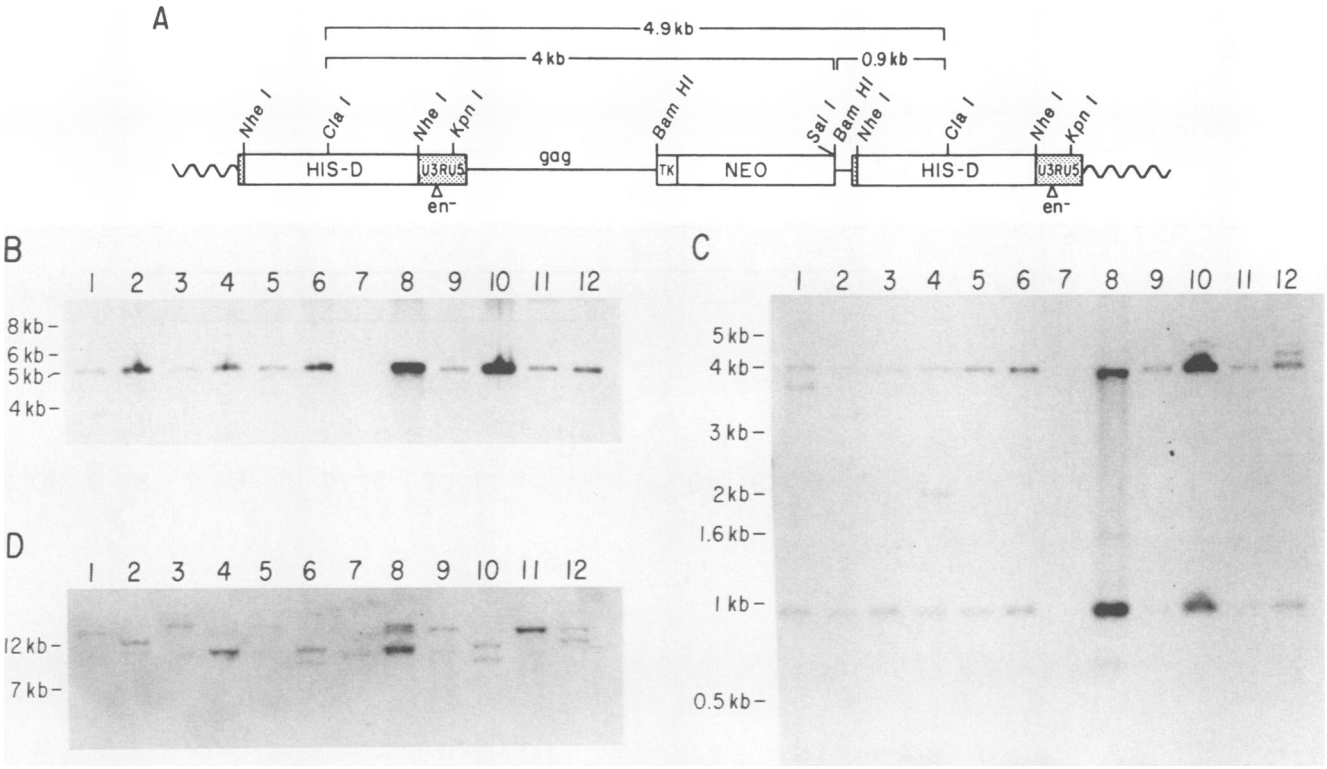

FIG. 2. Structure of integrated proviruses. (A) Restriction map of GgTKNeoU3Hisen(−) proviruses. (B to D) Southern blot analysis of GgTKNeoU3Hisen(−) proviruses in NIH 3T3 cell lines. Cell DNAs (approximately 10 μg per lane) digested with *Cla*I (B), *Cla*I and *Sal*I (C), and *Hin*dIII (D) were fractionated on agarose gels, blotted to nylon filters, and hybridized to *his* probes in lanes as follows: 1, PAC3(NEO); 2, PAC3(HIS); 3, PAC7(NEO); 4, PAC7(HIS); 5, PAC9(NEO); 6, PAC9(HIS); 7, ΨC7(NEO); 8,ΨC7(HIS); 9, ΨC9(NEO); 10, ΨC9(HIS); 11, ΨC10(NEO); 12, ΨC10(HIS). Clones selected in G418 (NEO) and L-histidinol (HIS) are indicated by parentheses.

by Southern blot hybridization (Fig. 2). Regardless of the initial selection, all but one clone contained proviruses in which *his* sequences had duplicated as part of the LTR (Fig. 2A to C). Thus, *Cla*I and *Sal*I endonucleases generated fragments expected from proviruses flanked by LTRs containing *his*. Cleavage fragments of 4.9 or 4 kbp and 0.9 kbp were generated after digestion with *Cla*I alone and together with *Sal*I, respectively. Only the neomycin-selected ΨC7 line lacked the expected provirus fragments; however, this line lacked most if not all *his* sequences and expressed aberrant *neo* transcripts, suggesting that the provirus DNA had rearranged. Additional bands of varying sizes represent fragments that extend from *Cla*I sites in the provirus to sites into the flanking cellular DNA.

To estimate the number of proviruses per cell, cellular DNAs were digested with *Hin*dIII (Fig. 2D) or *Nde*I (data not shown), enzymes that do not cut within the provirus. All clones contained one to three proviruses, and in all cases hybridization patterns were unique, confirming that each line was an independent clone.

Relatively few proviruses acquire the ability to express *his*. The ratio of neomycin and histidinol titers indicated that provirus integration was 2,500-fold less likely to convert cells to a His^r phenotype than to a Neo^r phenotype. In principle, the potential to express histidinol resistance could be an intrinsic but inefficient property of each provirus. For example, translation of 3′ *his* sequences in transcripts initiated at the *tk* promoter (Fig. 3A) could allow some *his* expression. Alternatively, *his* expression may require secondary events, such as mutations or transcriptional activation by adjacent cellular sequences.

The results of several experiments suggested that the capacity to transduce histidinol resistance is not an intrinsic property of the infecting virus. Cells initially selected in G418 did not survive when transferred to medium containing L-histidinol (Table 2), indicating that most proviruses did not confer histidinol resistance. In addition, the number of doubly resistant colonies produced after plating GgTKNeoU3Hisen(−)-infected cells in medium containing both G418 and L-histidinol was similar (within a factor of 2) to the number of colonies obtained after selection in L-histidinol alone (data not shown). This result implies that only a subset of the proviruses conferring neomycin resistance was capable of expressing histidinol resistance. Finally, the ability to passage histidinol resistance did not require the *tk* promoter, since histidinol titers for the GgTKNeoU3Hisen(−) and GgU3Hisen(−) viruses were nearly identical (Table 3).

Although secondary events apparently influence the ability of U3His vectors to transduce *his*, results of several experiments suggested that *his* expression was not activated by mutations within the provirus. First, the proviruses in His^r clones lacked gross sequence rearrangements, as judged by Southern blot analysis (Fig. 2). Second, proviruses rescued from His^r clones after superinfection with wild-type MoMuLV did not transduce *his* any more efficiently (as compared with *neo*) than did the original U3His vector (Table 4).

***his* transcripts initiate in the flanking cellular DNA.** To further examine why only certain proviruses expressed *his*, transcription of provirus sequences in His^r and Neo^r clones was analyzed by Northern blot hybridization (Fig. 3B and

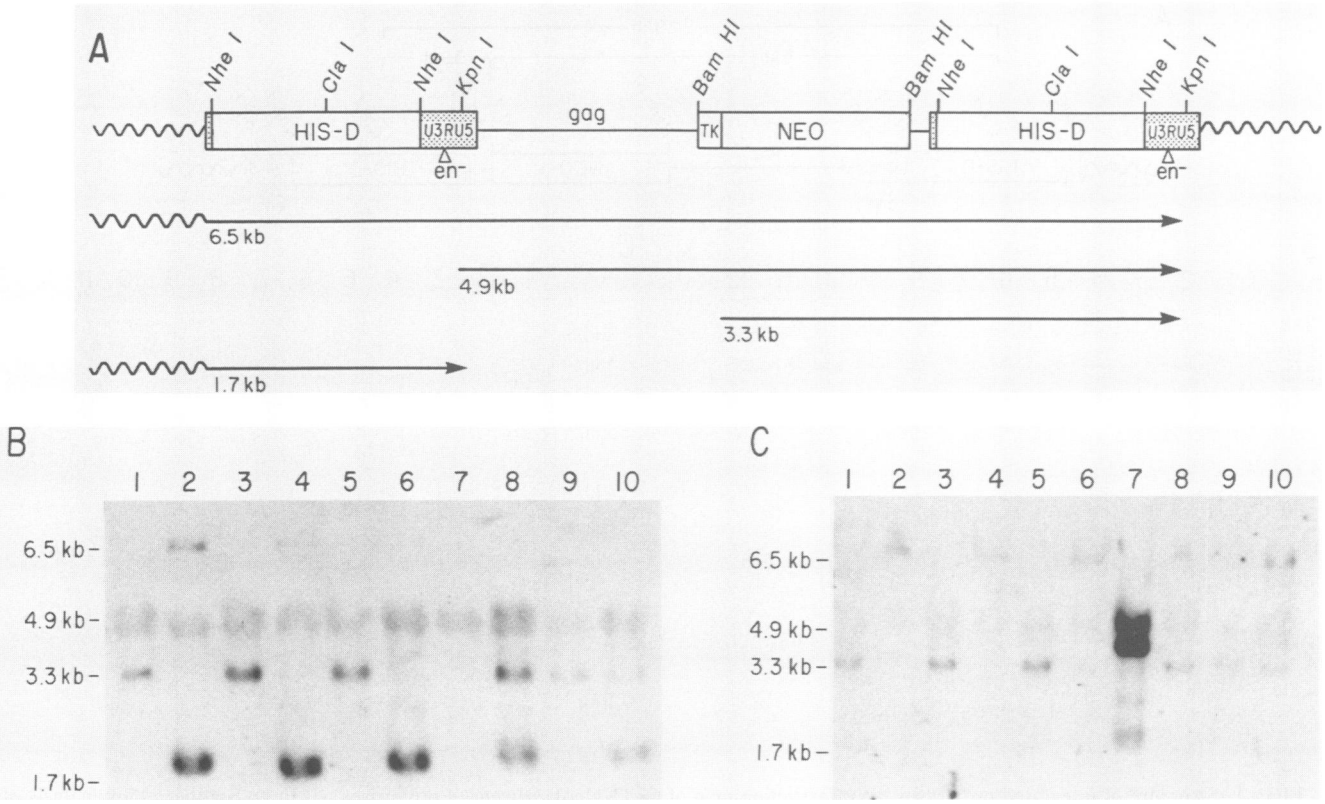

FIG. 3. Analysis of provirus transcripts. (A) GgTKNeoU3Hisen(−) provirus structure, predicted transcripts, and sites of transcription initiation in Neo^r and His^r clones. Total cellular RNA (10 μg per lane) was extracted from Neo^r and His^r clones, fractionated on formaldehyde-agarose gels, transferred to nylon membranes, and hybridized to *his* (B)- or *neo* (C)-specific probes in lanes as follows: 1, PAC3(NEO); 2, PAC3(HIS); 3, PAC7(NEO); 4, PAC7(HIS); 5, PAC9(NEO); 6, PAC9(HIS); 7, ΨC7(NEO); 8, ΨC7(HIS); 9, ΨC10(NEO); 10, ΨC10(HIS). (Clone designations are as for Fig. 2.) The sizes of major RNA species as indicated were estimated from RNA standards (Bethesda Research Laboratories).

C). All lines, whether selected in G418 or in L-histidinol, expressed 4.9- and 3.3-kbp provirus transcripts, whereas lines selected in L-histidinol expressed two additional transcripts of 6.5 and 1.7 kbp. These RNAs were mapped according to their ability to hybridize to *his*-, *neo*-, and *gag*-specific probes. Briefly, all four transcripts hybridized to *his* (Fig. 3B); the 6.5-, 4.9-, and 3.3-kbp RNAs hybridized to *neo* (Fig. 3C), and only the 6.5- and 4.9-kbp species hybridized to *gag* (data not shown). These data suggested that the 4.9- and 3.3-kbp RNAs started at the 5′ LTR and at the *tk* promoter, respectively, and terminated in the 3′ LTR, whereas, the 6.5- and 1.7-kbp RNAs in His^r clones appeared to initiate outside the provirus and terminate at polyadenylation sites in the 3′ and 5′ LTRs, respectively (Fig. 3A).

The sizes of the smaller (1.7 kbp) transcripts in His^r clones were never quite the same but varied by as much as 100 bp (Fig. 3B). This is the result one might expect if the provi-

TABLE 2. Cross-resistance of 3T3 cells expressing GgTKNeoU3Hisen(−) proviruses[a]

Clone	Colonies/10^5 cells	
	Neo^r	His^r
His^r		
PAC3	2	
PAC7	0	
PAC9	6.0×10^2	
ΨC7	1.2×10^2	
ΨC9	1.0×10^5	
ΨC10	2.0×10^3	
Neo^r		
PAC3		0
PAC7		0
PAC9		0
ΨC7		0
ΨC9		0
ΨC10		0

[a] Cells from Neo^r and His^r clones induced by GgTKNeoU3Hisen(−) were plated in selective media containing L-histidinol and G418, respectively. His^r and Neo^r colonies were counted after 10 days.

TABLE 3. Transduction of histidinol resistance by GgTKNeoU3Hisen(−) and GgU3Hisen(−) viruses[a]

Line	His^r (CFU/ml per 10^7 cells)[b]	
	GgTKNeoU3Hisen(−)	GgU3Hisen(−)
PA317	65 ± 71 (8)	40 ± 27 (4)
Ψ2	82 ± 122 (8)	185 ± 172 (7)

[a] PA317 and Ψ2 cells were transfected with GgTKNeoU3Hisen(−) or GgU3Hisen(−) DNA, and virus-producing lines were isolated in histidinol and G418 selective media. Supernatants from individual Neo^r [GgTKNeoU3Hisen(−)] or His^r [GgU3Hisen(−)] producers were titered on NIH 3T3 cells. Colonies were counted after 10 days in medium containing L-histidinol.

[b] Averages and standard deviations were obtained from several producer lines (numbers in parentheses).

TABLE 4. Transduction of histidinol and neomycin resistance by viruses rescued from His^r 3T3 cells expressing GgTKNeoU3Hisen(−) proviruses[a]

Line	Titer (CFU/ml per 10^7 cells)	
	Neo^r colonies	His^r colonies
NIH 3T3	0	0
ΨC7	3,500	0
ΨC9	1,660	0
ΨC10	2,500	0

[a] His^r clones and NIH 3T3 cells (10^5 cells) were infected with MoMuLV (10^6 PFU) and grown to confluence. Stocks of rescued virus were prepared by incubating 2 ml of medium with 10^7 infected cells for 2 h. NIH 3T3 cells (10^5) were exposed to various dilutions of virus and subsequently placed into selective medium containing G418 or L-histidinol. Resistant colonies were stained and counted after 10 days.

ruses were located at different distances from cellular promoters, and the size of each transcript depended on the amount of appended cellular RNA. In most cases, cell-derived sequences are expected to be short, since efficient translation of native histidinol dehydrogenase requires that the first AUG in the hybrid transcript be the initiating codon for histidinol dehydrogenase (7, 17, 25, 27). For the average mammalian gene, these sequences would average 50 to 100 nt (26). While transcription probably starts in sequences immediately adjacent to the provirus, these results do not exclude the possibility that integration has occurred 3′ to a splice acceptor site.

To confirm that transcripts in histidinol-selected clones initiated within the flanking cellular DNA, total RNA was extracted and analyzed by an RNase protection assay (Fig. 4A). ^{32}P-labeled RNA probes complementary to the provirus coding strand, extending from the *Cla*I and *Bam*HI sites in GgTKNeoU3Hisen(−), were prepared by using T3 RNA polymerase. RNA extracted from virus-producing lines or from histidinol- and neomycin-selected lines protected an 860-bp fragment. This corresponds to transcripts colinear with the provirus template, including both MoMuLV and *his* sequences. However, RNA from His^r lines generated an additional fragment of 720 bp, exactly the size expected for transcripts colinear with provirus sequences extending from the *Cla*I site to the 5′ end of the LTR.

Steady-state levels of the 3.3-kbp *tk neo his* mRNA were either undetectable or significantly lower in His^r clones than in Neo^r lines (Fig. 3B and C). Accordingly, His^r clones exhibited variable resistance to G418 (Table 2). Although the mechanisms are unknown, the phenomenon may reflect suppression of the *tk* promoters by an upstream cellular promoter not unlike promoter interference described elsewhere (10, 11).

DISCUSSION

This study exploited features of the retrovirus life cycle to construct MoMuLV vectors that identify cellular promoters. Retrovirus vectors containing *hisD* coding sequences inserted into the U3 region exhibited a normal ability to be passaged to recipient cells, generating proviruses flanked by *his*. Selection for *his* expression produced cell clones in which *his* sequences in the 5′ LTR were expressed on transcripts initiating in flanking cellular sequences. By cloning sequences linked to the proviruses in His^r cells, the vectors can be used to isolate cellular promoters.

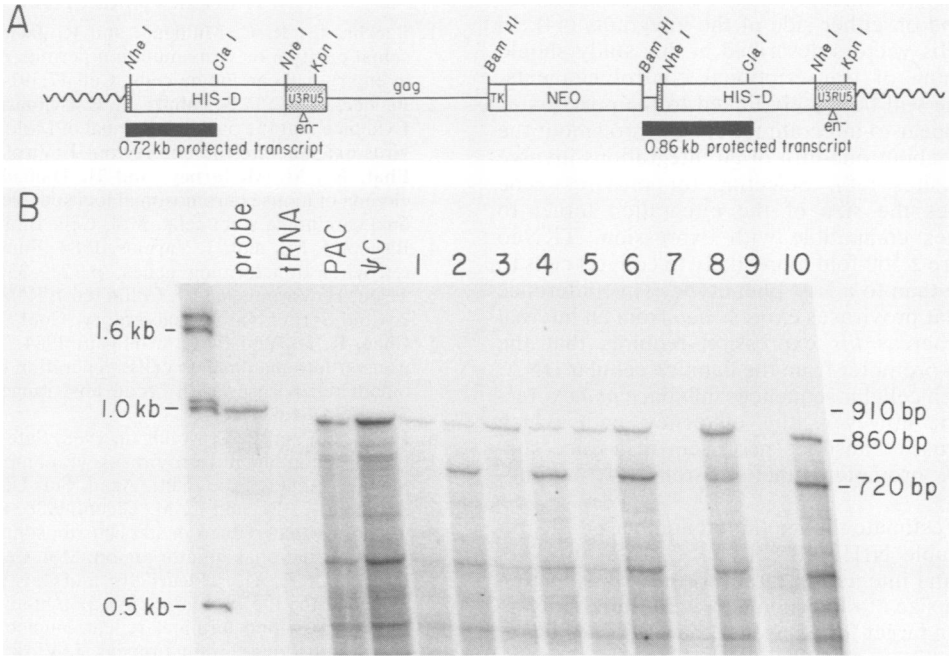

FIG. 4. RNase protection analysis of provirus transcripts. A 910-nt ^{32}P-labeled RNA probe, complementary to provirus sequences extending from the *Cla*I site in the 3′ LTR to the *Bam*HI site at the 3′ end of *neo*, was prepared and used in RNase protection assays as described in Materials and Methods. Transcripts extending through *his* sequences located in 3′ and 5′ LTRs should protect fragments of 860 and 720 nt, respectively (solid bars in panel A). (B) RNA samples from PAC, an amphotropic producer line, and ΨC, an ecotropic producer line, were analyzed. Lanes: 1, PAC3(NEO); 2, PAC3(HIS); 3, PAC7(NEO); 4, PAC7(HIS); 5, PAC9(NEO); 6, PAC9(HIS); 7, ΨC7(NEO); 8, ΨC7(HIS); 9, ΨC10(NEO); 10, ΨC10(HIS); probe, native probe; tRNA, control protection analysis using tRNA. (Clone designations are as for Fig. 2.) Sizes of end-labeled DNA fragments are indicated at the left; mobilities of predicted RNA fragments are indicated at the right.

It is interesting that the MoMuLV LTR can accommodate sequences as long as 1.3 kbp without seriously affecting the ability of the virus to be passaged. *his* sequences were placed 30 nt 3′ from the end of U3, downstream of sequences required for replication and integration into cellular DNA (14, 34, 35, 44). At 1.7-kbp, the elongated LTR is significantly longer than either the normal MoMuLV LTR (0.59 kbp), the 1.1-kbp LTR of mouse mammary tumor virus (8), or the 1.3-kbp LTR of a nondefective MoMuLV containing *dhfr* cDNA sequences (40). Whether retroviruses can tolerate longer LTR sequences remains to be determined.

Several investigators have isolated cellular promoters or enhancers by linking random DNA fragments to the coding sequence of a selectable marker (e.g., transforming or antibiotic resistance genes), introducing the DNA into recipient cells and selecting for cell clones that result if the gene is expressed (13, 22, 36, 37, 46). However, this approach suffers from several limitations that the strategy described here avoids. First, DNA-mediated gene transfer is less efficient than retrovirus transduction, particularly in certain cell types. Second, introduced genes are frequently amplified in cells surviving selection. This increases background and necessitates screening multiple clones or performing secondary transfections in order to identify clones containing only one gene copy. Third, potential promoter-enhancer elements identified after DNA transfer are not expressed in their normal chromosomal locations.

Transfected enhancerless genes have been used to identify transcriptionally active chromosome regions (3, 19, 22). In some cases, expression appeared to be regulated in a tissue-specific manner (3). Similarly, integration-specific activation of provirus genes has been observed in cells in which the LTR is transcriptionally inactive (1). However, cloning transcriptional activators by this approach is difficult because elements such as enhancers may be located at considerable distance and on either side of the integrants (44). In contrast, the U3His vectors described in this study should simplify the cloning of transcriptional control elements, since at least some will be closely linked to the provirus.

Retroviruses appear to integrate randomly throughout the genome, although about one-fifth of all integrations involve highly preferred sites (38). Selecting for provirus gene expression reduces the size of the integration target to genomic sequences compatible with expression. TKNeo U3His viruses were 2,500-fold more likely to convert cells to a Neo^r phenotype than to a His^r phenotype. This difference reflects the fact that proviruses express *neo* from an internal (*tk*) promoter, whereas *his* expression requires that the provirus acquire a promoter from the flanking cellular DNA. The extent to which cellular sequences influence *neo* expression is unknown; however, the efficiency with which TKNeoU3His viruses transduce neomycin resistance suggests that the *tk* promoter functions relatively autonomously.

It is possible to estimate the maximum number of integration sites that enable NIH 3T3 cells to express histidinol resistance. The total integration target for *neo* transduction cannot exceed the size of the genome (3×10^9 bp), implying that the integration target for *his* is less than 10^6 nt. Integrations that activate *his* are also likely to occur near transcriptional start sites to avoid appending AUG codons upstream of *his* coding sequences. In genes, the distance from the transcriptional start site to the first AUG averages 50 to 100 nt (26), while in random sequences, AUG triplets are distributed every 64 nt. Dividing the maximum integration target by the size of the average integration site yields a maximum of 1×10^4 to 2×10^4 integration sites capable of expressing *his* at levels sufficient to confer resistance.

Although this calculation probably overestimates the number of integration sites, the value is still small enough to suggest that U3His vectors may be used as insertional mutagens. In principle, the provirus would serve as a molecular tag to clone any gene whose function is linked to an observable phenotype, assuming that the gene is expressed at levels sufficient to promote histidinol resistance. Cells derived from a collection of 10^4 to 10^5 independent His^r clones should contain proviruses in all expressed sites, including transcriptionally active genes. Consequently, selection for His^r cell populations should reduce the number of integrants needed to screen for the loss of gene function. Recovery of clones expressing null phenotypes may be enhanced by the use of hypodiploid cells (e.g., CHO) and in cases in which gene inactivation leads to a selectable phenotype (e.g., tumor-suppressing genes or antioncogenes). Alternatively, homozygous loss of gene function may be accomplished by breeding mice derived from embryonal stem cells infected with U3His vectors and that contain germ line integrations of the provirus.

ACKNOWLEDGMENTS

We thank L. Gouw for technical assistance, K. Hunter and N. Speck for providing MoMuLV helper virus, and P. Robbins and R. C. Mulligan for the GgTKNeoen(−) plasmid.

This work was supported by Public Health Service grant P01 CA42063 from the National Cancer Institute to H.E.R. and partially supported by Cancer Center core grant 14051 to P. A. Sharp and a Benevolentia Foundation grant (C & A Brenninkmeyer, The Netherlands) to H.v.M. H.v.M. was supported by the Deutsche Forschungsgemeinschaft, Bonn-Bad Godesberg, Federal Republic of Germany. H.E.R. is a Rita Allen Foundation scholar.

LITERATURE CITED

1. **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.
2. **Bender, M. A., T. D. Palmer, R. E. Gelinas, and D. Miller.** 1987. Evidence that the packaging signal of Moloney murine leukemia virus extends into the *gag* region. J. Virol. **61**:1639–1646.
3. **Bhat, K., M. McBurney, and H. Hamada.** 1988. Functional cloning of mouse chromosomal loci specifically active in embryonal carcinoma stem cells. Mol. Cell. Biol. **8**:3251–3259.
4. **Bishop, J. M., and H. Varmus.** 1984. Functions and origins of retroviral transforming genes, p. 249–356. *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.
5. **Cone, R. D., and R. C. Mulligan.** 1984. High efficiency gene transfer into mammalian cells: generation of helper free recombinant retroviruses with broad host range. Proc. Natl. Acad. Sci. USA **81**:6349–6353.
6. **Danos, O., and R. C. Mulligan.** 1988. Safe and efficient generation of recombinant retroviruses with amphotropic and ecotropic host ranges. Proc. Natl. Acad. Sci. USA **85**:6460–6464.
7. **Donahue, T. F., and A. M. Cigan.** 1988. Genetic selection for mutations that reduce or abolish ribosomal recognition of the *HIS4* translational initiator region. Mol. Cell. Biol. **8**:2955–2962.
8. **Donehower, L. A., B. Fleurdelys, and G. L. Hager.** 1983. Further evidence for the protein coding potential of mouse mammary tumor virus long terminal repeat: nucleotide sequence of an endogenous long terminal repeat. J. Virol. **45**:941–949.
9. **Dougherty, J. P., and H. Temin.** 1987. Promotorless retroviral vector indicates that there are sequences in U3 required for 3′ RNA processing. Proc. Natl. Acad. Sci. USA **84**:1197–1201.
10. **Emerman, M., and H. M. Temin.** 1984. Genes with promotors in retrovirus vectors can be independently suppressed by an epigenetic mechanism. Cell **39**:459–467.
11. **Emerman, M., and H. M. Temin.** 1986. Quantitative analysis of

gene suppression in integrated retrovirus vectors. Mol. Cell. Biol. **6**:792–800.
12. **Feinberg, A. P., and B. Vogelstein.** 1984. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137**:266–267.
13. **Fried, M., M. Griffith, B. Davies, G. Bjursell, G. L. Mantia, and L. Lania.** 1983. Isolation of cellular DNA sequences that allow expression of adjacent genes. Proc. Natl. Acad. Sci. USA **80**:2177–2181.
14. **Fujiwara, T., and K. Miyuuchi.** 1988. Retroviral DNA integration: structure of an integration intermediate. Cell **54**:497–504.
15. **Fung, Y. K. T., A. M. Fadly, L. B. Crittenden, and H. J. Kung.** 1981. On the mechanism of retrovirus-induced avian lymphoid leukosis: deletion and integration of the proviruses. Proc. Natl. Acad. Sci. USA **78**:3418–3422.
16. **Fung, Y. K. T., W. G. Lewis, K. J. Kung, and L. B. Crittenden.** 1983. Activation of the cellular c-erbB by LTR insertion: molecular basis for induction of erythroblastosis by avian leukosis virus. Cell **33**:357–368.
17. **Geballe, A. P., and E. S. Mocarski.** 1988. Translational control of cytomegalovirus gene expression is mediated by upstream AUG codons. J. Virol. **62**:3334–3340.
18. **Guild, B. C., M. H. Finer, D. E. Housman, and R. C. Mulligan.** 1988. Development of retrovirus vectors useful for expression of genes in cultured murine embryonal cells and hematopoietic cells in vivo. J. Virol. **62**:3795–3801.
19. **Hamada, H.** 1986. Activation of an enhancerless gene by chromosomal integration. Mol. Cell. Biol. **6**:4179–4184.
20. **Hartmann, S. C., and R. C. Mulligan.** 1988. Two dominant-acting selectable markers for gene transfer studies. Proc. Natl. Acad. Sci. USA **85**:8047–8051.
21. **Hayward, W. S., B. G. Neel, and S. M. Astrin.** 1981. Activation of a cellular onc-gene by promotor insertion in ALV-induced lymphoid leukosis. Nature (London) **290**:475–480.
22. **Hiller, S., M. Hengstler, M. Kunze, and R. Knippers.** 1988. Insertional activation of a promoterless thymidine kinase gene. Mol. Cell. Biol. **8**:3298–3302.
23. **Jaenisch, R., D. Jaehner, P. Nobis, I. Simon, J. Loehler, K. Harbers, and D. Grotopp.** 1981. Chromosomal position and activation of retroviral genomes inserted into the germ line of mice. Cell **24**:519–529.
24. **King, W., M. Patel, L. Lober, S. Goff, and M. Nguyen-Huu.** 1985. Insertion mutagenesis of embryonal carcinoma cells by retroviruses. Science **228**:554–558.
25. **Kozak, M.** 1980. Evaluation of the "scanning model" for initiation of protein synthesis in eucaryotes. Cell **22**:7–8.
26. **Kozak, M.** 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. **15**:8125–8148.
27. **Kozak, M.** 1987. Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. Mol. Cell. Biol. **7**:3438–3445.
28. **Lugo, T., B. Handelin, B. Killary, D. Housman, and R. E. Fournier.** 1987. Isolation of microcell hybrid clones containing retroviral vector insertions into specific human chromosomes. Mol. Cell. Biol. **7**:2814–2820.
29. **Luthman, H., and G. Magnusson.** 1983. High efficiency polyoma DNA transfection of cloroquine treated cells. Nucleic Acids Res. **11**:1295–1307.
30. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **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**:153–159.
32. **McNight, S.** 1980. The nucleotide sequence and transcript map of the herpes thymidine kinase gene. Nucleic Acids Res. **8**:5949–5964.
33. **Miller, A. D., and C. Buttimore.** 1986. Redesign of retrovirus packaging cell lines to avoid recombination leading to helper virus production. Mol. Cell. Biol. **6**:2895–2902.
34. **Panganiban, A. T.** 1985. Retroviral DNA integration. Cell **42**:5–6.
35. **Panganiban, A. T., and H. M. Temin.** 1984. Circles with two tandem LTRs are precursors to integrated retrovirus DNA. Cell **36**:673–679.
36. **Santangelo, G. M., J. Tornow, C. S. McLaughlin, and K. Moldave.** 1988. Properties of promoters cloned randomly from the *Saccharomyces cerevisiae* genome. Mol. Cell. Biol. **8**:4217–4224.
37. **Santangelo, G. M., J. Tornow, and K. Moldave.** 1986. Cloning of open reading frames and promotors from the Saccharomyces cerevisiae genome: construction of genomic libraries of random small fragments. Gene **46**:181–186.
38. **Shih, C. C., J. P. Stoye, and J. M. Coffin.** 1988. Highly preferred targets for retrovirus integration. Cell **53**:531–537.
39. **Sorge, J., A. E. Cutting, V. D. Erdman, and J. W. Gautsch.** 1984. Integration-specific retrovirus expression in embryonal carcinoma cells. Proc. Natl. Acad. Sci. USA **81**:6627–6631.
40. **Stuhlmann, H., R. Jaenisch, and R. C. Mulligan.** 1989. Construction and properties of replication-competent murine retroviral vectors encoding methotrexate resistance. Mol. Cell. Biol. **9**:100–108.
41. **Taketo, M., E. Gilboa, and M. 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.
42. **Varmus, H.** 1988. Retroviruses. Science **240**:1427–1435.
43. **Varmus, H. E., N. Quintrell, and S. Ortiy.** 1981. Retroviruses as mutagens: insertion and excision of a non-transforming provirus alters expression of a resident transforming provirus. Cell **25**:23–36.
44. **Varmus, H., and S. R. Swanstrom.** 1984. Replication of retroviruses. p. 75–134. *In* R. Weiss, N. Teich, H. Varmus, and R. Coffin (ed.), RNA tumor viruses, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
45. **von Melchner, H., and D. E. Housman.** 1988. The expression of the neomycin-phosphotransferase in human promyelocytic leukemia cells (HL60) delays their differentiation. Oncogene **2**:137–140.
46. **Weber, F., J. Villiers, and W. Schaffner.** 1984. An SV40 "enhancer trap" incorporates exogenous enhancers or generates enhancers from its own sequences. Cell **36**:983–992.