Retrovirus activation in embryonal carcinoma cells by cellular promoters

Ivan Peckham,† Suzanne Sobel,‡ Jeff Comer,† Rudolf Jaenisch,§ and Eric Barklis†

†Vollum Institute for Advanced Biomedical Research and Department of Microbiology and Immunology, Oregon Health Sciences University, Portland, Oregon 97201 USA; ‡Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142 USA

Retrovirus expression in embryonal carcinoma (EC) cells is blocked at a postintegration stage of the viral life cycle, because of the inadequate function of the viral long terminal repeat (LTR) promoter in this cell type. However, rare sites in the EC cell genome permit provirus expression by undefined mechanisms. Our analysis of three expressed proviruses indicates that they have inserted into actively transcribed regions. Two of the three, examined in detail, integrated into the first introns of cellular transcription units in close proximity to active cellular promoters. One of these cellular genes is the probable murine homolog of the yeast ribosomal protein L3, responsible for trichodermin resistance. In all cases, virus activation appears to involve production of viral transcripts that are initiated in the 5′-flanking region, transcribed through the viral LTR, and subsequently spliced from a cellular donor to a viral acceptor. Our results suggest a general procedure for the isolation of active genes and promoters in different tissues.

[Key Words: Retrovirus, embryonal carcinoma, moloney murine leukemia virus; ribosomal protein L3; trichodermin resistance]

Received August 1, 1989; revised version accepted October 10, 1989.

Embryonal carcinoma (EC) cells are rapidly dividing stem cells which, in many respects, are similar to inner cell mass (ICM) cells of mammalian preimplantation embryos. Representing the earliest culturable stem cells of mammalian organisms, EC cells have been used extensively in the analysis of early events in differentiation [Sherman 1975; Strickland and Mahdavi 1978]. After growth in culture, ectopic placement of EC cells usually leads to unrestricted growth and formation of a teratocarcinoma [Bernstine et al. 1973]. However, a variety of murine embryonal stem (ES) cell lines can be placed in the ICM of a developing mouse embryo and will contribute to the formation of all tissue types in the generation of a mosaic animal [Evans and Kaufman 1981; Martin 1981; Bradley et al. 1984]. This ability of some ES cells has proved advantageous in the construction of transgenic animals [Gossler et al. 1986; Robertson et al. 1986]. One characteristic common to EC cells and preimplantation embryos is their failure to express viral genomes, including the Moloney murine leukemia virus (M-MuLV) genome [Jaenisch and Berns 1977; Teich et al. 1977]. In EC cells, retroviral integration is unimpaired, but replication is blocked at a postintegration step because [Stewart et al. 1982; Gautsch and Wilson 1983], at least partly, the promoter in the retroviral long terminal repeat (LTR) functions inadequately [Linney et al. 1984; Gorman et al. 1985].

Our previous work and the work of others have demonstrated that selection for retrovirus expression in EC cells may yield retrovirus host range mutants [Barklis et al. 1986; Hilberg et al. 1987; Weiher et al. 1987] or integration sites in the EC genome that are permissive for retrovirus expression [Sorge et al. 1984; Taketo et al. 1985; Barklis et al. 1986; Bhat et al. 1988]. Nevertheless, the number of EC cell chromosomal sites permitting retrovirus expression is very small [Barklis et al. 1986]. Previously, we isolated three proviral preintegration sites found to permit virus expression [Barklis et al. 1986]. Several independently isolated proviruses that are expressed in EC cells were also mapped to two of these regions. Although our results indicated that viral activation required the presence of EC cell DNA flanking the 5′ LTR, the mechanism by which activation occurred was unclear.

We examined the mechanism of retrovirus activation at three specific loci in the EC genome and have found revealing similarities. All three loci encode cellular transcripts that are expressed in uninfected, as well as infected, EC cells. Viral transcripts appear to be initiated in the 5′-flanking region, transcribed through the viral LTR, and subsequently spliced from a cellular donor to a viral acceptor. In two of these cases, we have isolated cDNA clones to the cellular transcripts. Analysis indicates that activated proviruses integrated into the first introns of the cellular transcription units, downstream from cellular promoters which function well in both differentiated and EC cells. In one case, viral integration...
occurred into the murine homolog of the yeast ribosomal protein gene L3, responsible for trichodermin resistance (Schultz and Friesen 1983). From these analyses, it appears that we have selected rare chromosomal loci downstream from very strong cellular promoters and splice donors. Our results suggest a general procedure for the isolation of active genes and promoters in different tissues.

Results

Activated proviruses have integrated into transcriptionally active regions

The A2, F2, and J1 proviral integrations (see Fig. 1) represent three rare sites in the F9 (Bernstine et al. 1973) EC cell genome, which are permissive for retrovirus expression. These three sites are recombinant MP10 proviral integrations into genes that are present twice in the F9 diploid genome. As MP10 proviruses, they carry the selectable neomycin gene in place of the M-MuLV gag, pol, and env genes but retain the viral LTR and regulatory regions. Within the proviruses, the neomycin-coding region is located downstream of the 5' LTR, a splice donor, and a splice acceptor so that expression may be from either the spliced or unspliced transcripts in permissive cells. Previous experiments demonstrated that in undifferentiated F9 cells, A2, F2, and J1 expression was dependent on cellular 5'-flanking DNA and that transcription of viral RNA did not initiate at the usual start site (Barklis et al. 1986). The scarcity of chromosomal regions that permit virus expression was underscored by isolation of the independently derived active MP10 proviruses E1 and D2 in the region of the F2 integration site and the active G2 provirus near the J1 integration site.

What is the mechanism of A2, F2, and J1 provirus activation of EC cells? As an initial step toward answering this question, we probed blots of RNA from uninfected F9 cells and from F9 cells harboring the A2, F2, or J1 proviruses (Fig. 2). Hybridization probes were from chromosomal regions flanking the 5' LTR of each provirus (depicted by heavy bars in Fig. 1). Interestingly, each flanking sequence probe detects cellular transcripts in uninfected cells. The A2 probe detects an RNA at 800 bases and a potential secondary band at 360 bases [data not shown]. The F2 probe hybridizes to prominent bands at 1400, 1020, and 700 bases, as well as barely detectable species at 2050 and 800 bases [Fig. 2A]. The predominant J1 cellular transcript is ~1450 bases, with a potential band at 350 base and an additional RNA at 750 base [Fig. 2B]. The J1 750-base RNA is enriched in cells containing the J1 provirus, possibly indicating an interruption of the cellular transcript by the provirus. Except for the J1 750-base band, the hybridization patterns of RNA from uninfected cells versus cells harboring activated proviruses are qualitatively equivalent—bands detected in uninfected cells are the ones appearing in cells with proviruses. However, with each probe, there is a discernible reduction of hybridization signal in RNA from the infected versus the parental F9 cells. Although this signal reduction could be within the margin of error for our hybridization procedure, an alternate explanation is that provirus insertion in one of each allelic pair reduces or blocks transcription from that allele.

Viral transcripts are of the spliced message size

There is a variety of mechanisms by which insertion into an actively transcribed region may activate retrovirus expression in EC cells. Two obvious possibilities are activation by a cellular enhancer (Taketo et al. 1985)

---

Figure 1. Activated retrovirus integrations. Illustrated are the integration sites of the A2, F2, and J1 proviruses, which are MP10 recombinant retroviruses that require cellular 5'-flanking DNA for expression in undifferentiated F9 cells (Barklis et al. 1986). Proviruses are bracketed by black boxes, indicative of the retroviral LTRs, proviral transcription units 5' to 3', are oriented left to right. The neomycin gene, which replaces the M-MuLV gag, pol, and env genes, is located downstream of a 440-bp intron, indicated by the splice donor (S.D.) and the splice acceptor (S.A.), and may be expressed from either the unspliced or spliced transcript. Cellular flanking DNA is located upstream (left) or downstream (right) of the proviral 5' and 3' LTRs, respectively. The locations and orientations of other activated MP10 proviruses (D2, E1, and G2), which integrated independently into other F9 cells, are indicated (not drawn to scale), by small arrows. Bold black lines beneath each map depict flanking sequence probes, which detect sequences present twice in the mouse diploid genome and were employed in our initial RNA blot hybridizations (see Fig. 2). All proviral NheI sites and all EcoRI sites are indicated. The cellular flanking sequence A2 DraI, F2 XbaI and BglIII, and J1 HindIII sites are also indicated. The scale is as shown.
or by a cellular promoter. In the former case, it is expected that enhancer activity from the cellular element would boost transcription from the viral promoter. In contrast, activation by a cellular promoter is expected to result in the formation of a readthrough transcript, containing both cellular and viral sequences. Previous analysis using RNase protection assays indicated that the A2, F2, and J1 viral transcripts do not start at the normal viral initiation site (Barklis et al. 1986), suggesting that cellular flanking elements do not activate the viral promoter. However, hybridization of RNA blots with flanking sequence probes did not detect any obvious readthrough transcripts [see Experimental methods]. Comparison to marker RNA mobilities and their log molecular weights indicates a major J1 RNA at ~1450 bases, a faint J1 RNA at 750 bases [more pronounced in B, J1 versus F9 lane], and a potential minor band at 350 bases. There are a series of F2 RNAs, including major bands at 1400, 1020, and 700 base, barely visible bands at ~2050 and 800 bases, and a cellular F2 transcript, detectable only with other F2 probes, at 500 bases.

A possible explanation for the apparent absence of readthrough-type transcripts is that they may be detected inefficiently with our probes if cellular sequences comprise a small proportion of the resultant RNA [see below]. Additionally, readthrough transcripts might be present at low levels because of termination and polyadenylation signals in the proviral 5' LTRs [Herman and Coffin 1987]. To detect viral transcripts, it was necessary to employ a full-length neomycin RNA probe radioslated to a high specific activity. In MP10-infected, permissive 3T3 cells, this probe detects an MP10 unspliced transcript of 4150 bases and a 3710-base spliced transcript [Fig. 3E,F]. As expected, undifferentiated, uninfected F9 cells do not demonstrate a detectable neo message [Fig. 3D]. However, in F9 cells, which carry the A2, J1, and F2 proviruses, individual RNAs corresponding to the size of the spliced viral RNA are evident [Fig. 3A–C]. The sizes of these RNAs, smaller than the 4150-base full-length viral message, are not consistent with the notion that they are simple readthrough products from an active promoter in the cellular flanking sequences. However, one possible explanation for the observed neo message sizes is that the RNAs were generated by readthrough from active cellular promoters, followed by splicing from cellular splice donor sites to the proviral splice acceptor sites. This model is consistent with our observations concerning the neo message sizes [Fig. 3] and the absence of transcription initiation at the standard viral start site (Barklis et al. 1986) and is supported by our analysis of cDNA clones to the cellular F2 and J1 messages [below].

**F2 and J1 proviruses have inserted into the 5' introns of cellular transcription units**

Clarification of the interaction of cellular and viral transcription units necessitated further characterization of the cellular mRNAs. Thus, we undertook the isolation of genomic...
of cDNA clones corresponding to RNAs encoded by two of the three integration regions, F2 and J1. Primary cDNA clones (cJ1-3400 and cF2-1900) were isolated by using flanking sequence probes to screen a Agt10 cDNA library from undifferentiated, uninfected F9 cells (see Experimental methods). Secondary clones (cF2-800, cJ1-2400, cJ1-1300) were isolated from the F9 cDNA library and a 14-day mouse embryo cDNA library using fragments of primary cDNA clones as probes [see Experimental methods].

To map the cellular F2 and J1 transcription units, all clones were restriction-mapped; critical portions for intron/exon mapping of cJ1-3400 and cJ1-2400 clones were sequenced, and all portions of cJ1-1300, cF2-1900, cF2-800, and the F2 and J1 genomic clones were sequenced. Transcription directions predicted by poly(A) tail sequences were confirmed by standard protocols (Melton et al. 1984; Barklis et al. 1985). Figure 4A diagrams our results for the F2 region. As shown, the 360 bp upstream from the F2 proviral integration site on the genomic clone is identical to the 5' 360 bp of our two cF2-1900 cDNA clones. The two cF2-800 cDNA clones are identical to the cF2-1900 cDNAs, except that two intron regions are removed in the smaller clones. The splice donor site for the 892-base 5' intron occurs 335 bp upstream from the F2 proviral integration site on our genomic clone, indicating that the F2 provirus has inserted into an intron. Because the 5' ends of all four F2 cDNA clones map to this region and because the genomic DNA immediately upstream contains a promoter that is active in undifferentiated F9 cells [see Fig. 5A; Table 1], we infer that integration has occurred in the first intron of the cellular transcription unit.

Analysis indicates that the J1 provirus has also integrated into the first intron of a transcription unit [Fig. 4B]. The J1 genomic clone is colinear with the completely processed cJ1-3400 clone for 1000 bp upstream of the J1 provirus, within the most 5' of the five introns present in the cJ1-3400 clone. The 5' ends of all J1 cDNA clones map just upstream of the 5' cJ1-3400 intron and just downstream of a strong cellular promoter (see

---

**Figure 4.** F2 and J1 cellular transcription units. Transcription maps of the F2 and J1 cellular transcription units were generated by a combination of procedures, including restriction digests, DNA and RNA blot hybridizations, DNA sequencing of portions of cJ1-3400 and cJ1-2400, and DNA sequencing of all portions of cJ1-1300, cF2-1900, cF2-800, and the F2 and J1 genomic regions. The orientations of transcription units are indicated by arrows and were determined by the presence of poly(A) tails on cDNA clones and confirmed by RNA blot hybridizations using single-stranded probes. Splice junctions were determined by DNA sequence comparison of genomic and different cDNA clones. Exon regions present in the smallest cDNA clones (cF2-800 and cJ1-1300) are shaded. On genomic maps, the upstream portion of the proviral 5' LTRs are illustrated as black boxes, whereas the cellular promoter regions [as determined in Fig. 5 and Table 1] are shown as black ovals. Restriction endonuclease sites are B (BanII), Bg (BglII), E (EcoRI), Ev (EcoRV), H (HindIII), N (Nhel), Pst (PstI), Ps (PvuII), R (Rsal), Sa (SalI), Sc (ScaI), SspI, Xba (XbaI), Xho (Xhol). (A) The F2 genomic region is an expanded view of the F2 Xbal–Nhel genomic region from Fig. 1. Two independent cF2-1900 clones were isolated from an F9 cDNA library, and two cF2-800 clones were isolated, one from the F9 cDNA library and one from a mouse embryo library. The 5' end of the sequenced cF2-1900 clone is identical to the F2 genomic region flanking the proviral 5' LTR. Not all Rsal sites on cF2-1900 and no Rsal sites on cF2-800 are shown. (B) The J1 genomic region is an expanded view of the J1 HindIII–Nhel genomic region from Fig. 1. The cJ1-3400 clone was isolated from the F9 cDNA library, possesses no poly(A) tail, and retains a genomically encoded EcoRI site at its 5' end. Independent cJ1-2400 and cJ1-1300 clones were isolated from both the F9 and mouse embryo cDNA libraries. The 5' end of the cJ1-3400 clone is identical to the J1 genomic region flanking the proviral 5' LTR and is colinear with the cJ1-2400 clones in the region containing the second, third, and fourth exons. There are two adjacent Rsal sites at the designated position in the J1 genomic clone. Not all Rsal sites on cJ1-3400 and no Rsal sites on cJ1-1300 are shown.
Identification of the cellular F2 and J1 promoters

On our genomic clones, 30–40 bp upstream from where the 5' ends of our longest cDNA clones map, we observed DNA sequences similar or identical to the consensus TATA box (Corden et al. 1980) sequence (see Fig. 5). Upstream of each TATA box, 70–150 bp, CCAAT box sequences (Dierks et al. 1983; Dorn et al. 1987) are present. Additional putative regulatory elements present in these regions are Sp1 sites (Kadonaga et al. 1987) in the J1 promoter (Fig. 5B, −152 to −147 and −68 to −63) and potential octamer (Sen and Baltimore 1986; Fig. 5A, −152 to −147) and E2F (Kovesdi et al. 1986; Fig. 5A, −190 to −183) sites in the F2 promoter.

As a functional test, we examined the promoter activity of the 313-bp F2 genomic Xbal–RsaI fragment (Figs. 4A and 5A, nucleotides −281 to +31) and the 214-bp J1 genomic HindIII–RsaI fragment (Figs. 4B and 5A, nucleotides −200 to +14). These fragments were inserted in either orientation upstream of the chloramphenicol acetyltransferase (CAT) gene and a synthetic polyadenylation signal (see Experimental methods): The resulting constructs were named F2ProCAT, F2RevCAT, J1ProCAT, and J1RevCAT. Additional constructs, F2LPCAT and J1LPCAT, included 1500 bp of additional 5' F2 genomic DNA and an additional 3400 bp of 5' J1 genomic DNA, respectively. As positive and negative controls, the constructs PyCAT (originally referred to as ΔMo + PyF101CAT by Linney et al. 1984), which is active in undifferentiated F9 cells, and MoCAT, which is only active in differentiated cells (Linney et al. 1984), were employed.

Our indirect assay for promoter activity followed the standard transient transfection protocol of Gorman et al. (1985) [see Table 1]. Our positive control plasmid PyCAT, employing the recombinant ΔMo + PyF101 promoter, expresses the CAT gene when transfected into undifferentiated cells (29% conversion). In contrast, the M-MuLV LTR promoter of MoCAT functions inadequately in these cells (2.1% of PyCAT), in agreement with previous results (Linney et al. 1984; Gorman et al. 1985; Barklis et al. 1986). The experimental constructs F2ProCAT and J1ProCAT yield CAT enzyme levels in excess of our positive control and 64- to 96-fold higher than the MoCAT control: These levels of CAT activity are the highest we observed in F9 cells [data not shown] and support the notion that activated F2 and J1 proviruses integrated downstream of strong cellular promoters. The orientation specificity of these promoters is demonstrated by the inactivity of the constructs F2ProCAT and J1RevCAT [Table 1], where the CAT gene is linked to the F2 promoter at the upstream Xbal site, and the J1 promoter is linked at the HindIII site [see Figs. 1, 4, and 5]. We cannot be certain as to whether the F2ProCAT and J1ProCAT constructs contain complete or truncated promoters. However, including an additional 1500 bp of upstream DNA from the F2 genomic clone does not appear to increase F2 promoter activity in F9 cells [cf. F2LPCAT versus F2ProCAT]. Interestingly, inclusion of 3400 bp of genomic J1 sequence flanking the J1 promoter HindIII site reduces CAT expression in F9 cells [cf. J1LPCAT vs. J1ProCAT]. This reduction of promoter activity could be due to the presence of a negative regulatory element upstream of the J1 genomic HindIII site.

Integration of the J1 provirus at the murine homolog of the yeast ribosomal protein gene L3

To investigate whether proviral preintegration sites correspond to known genes, we searched data bases for homologies to our sequences. However, our examination of the F2 genomic and cDNA clones was uninformative. Although a probe from the cF2-1900 cDNA clone identified discrete bands when hybridized to blots of restricted genomic DNA from all mammalian species tested, DNA sequence analysis of the cF2-800, cF2-1900, and F2 genomic clones yielded no significant homologies when compared with nucleotide sequences in the EMBL, NBRF, or GenBank libraries [data not shown]. Although our cDNA clones may not represent the natural, fully processed F2 cellular RNA, the absence of any clear long

Figure 5. Sequence of the F2 and J1 cellular promoters. DNA sequence of the F2 and J1 promoter regions is shown. Circled nucleotides corresponding to the 5' most nucleotide from the promoter region. The DNA sequence from the J1 genomic sequence (Corden et al. 1980), are boxed, as is the probable initiation codon of the putative J1 protein (see Fig. 6). Potential CCAAT box elements (Dierks et al. 1983; Dorn et al. 1987) are underlined. (A) F2 promoter region. The DNA sequence of the F2 Xbal–RsaI genomic fragment (see Fig. 4A) is shown. (B) J1 promoter region. The DNA sequence from the J1 genomic HindIII to the second of the two adjacent RsaI sites (see Fig. 4B) is shown.
open reading frames in any of our cDNA clones may reflect the possibility that the F2 proviral integration was at a transcribed pseudogene.

In contrast, comparison of the cJ1-1300 sequence to data banks revealed a remarkable homology to the yeast ribosomal protein gene L3 to which resistance to the drug trichodermin maps [Schultz and Friesen 1983]. The deduced translation product of cJ1-1300 is a 403-residue polypeptide [Fig. 6], as compared to a length of 385 residues for the yeast protein (Schultz and Friesen 1983). Additionally, probes from the J1 genomic clone and all genomic regions are strong cellular promoters that function as well conserved in the putative murine ribosomal L3 protein. The amino-terminal 21 residues of the yeast protein, essential for nuclear localization in yeast cells (Moreland et al. 1985), is well conserved in the putative J1 protein (90% identity). Interestingly, the first exon of the J1-1300 RNA is spliced precisely after the initiator methionine codon [Fig. 5B] at a donor site [ATG/GTACGT] that is slightly unusual for RNA polymerase II products [Mount 1982]. Although we are not certain that the murine ribosomal L3 protein is encoded only at the J1 chromosomal site, our data suggest that it might be. The J1 genomic probe detects a DNA fragment that is present once in the mouse haploid genome [Barklis et al. 1986], indicating that this active promoter region is unique. Additionally, probes from the J1 genomic clone and all J1 cDNA clones detect a unique 650-bp XbaI genomic fragment corresponding to the cJ1-3400 650-bp XbaI fragment [data not shown]. These data are consistent with the notion that the putative murine ribosomal L3 protein is encoded at a single locus and that the J1 provirus is activated by insertion into that locus.

### Discussion

M-MuLV expression is restricted in EC cells [Jaenisch and Berns 1977; Teich et al. 1977]. Virus integration occurs at roughly normal efficiency [Stewart et al. 1982; Gautsch and Wilson 1983], but steady-state levels of RNA are diminished 50- to 100-fold in EC cells relative to differentiatied cell types [Stewart et al. 1982]. Evidence suggests that EC cells lack at least some positive trans-activating factors necessary for effective promotion from the viral LTR promoter [Linney et al. 1984]. Additionally, the M-MuLV provirus contains at least two cis-active negative regulatory elements [Feuer et al. 1989], one in the LTR U3 region [Gorman et al. 1985], and one in the vicinity of the primer-binding site [PBS] [Barklis et al. 1986; Loh et al. 1987].

Some retroviral integrants manage to circumvent the EC expression block by mutation in one of the cis-active viral regulatory elements [Barklis et al. 1986; Hilberg et al. 1987]. Alternatively, virus expression may be permitted at certain loci in the EC cell genome [Sorge et al. 1984; Taketo et al. 1985]. For the recombinant MP10 retrovirus, the number of permissive chromosomal regions for provirus expression has been estimated to be ~100 [Barklis et al. 1986]. That these locations are rare has been underscored by the observation of multiple independent proviral integrations into a limited number of sites [Barklis et al. 1986].

What are the characteristics of genomic sites permissive for retrovirus expression in EC cells? Our evidence suggests that activation of MP10 proviruses in EC cells occurs by a very specific mechanism. All three activated proviruses examined, A2, F2, and J1, have integrated into transcriptionally active regions of the EC cell genome. None of the proviral transcripts is initiated at the normal viral transcription start site [Barklis et al. 1986], and all the viral transcripts are present at low levels and appear to be of the spliced size [Fig. 3]. Analysis of the F2 and J1 proviruses indicates that they integrated into the 5′-most introns and in the same orientation as the cellular transcription units that they perturbed [Fig. 4]. It is noteworthy that the other activated proviruses that integrated into these transcription units (G2, D2, and E1; see Fig. 1) also are localized in introns in the same orientations. At the 5′ ends of these genomic regions are strong cellular promoters that function as well as any that we have tested in undifferentiated cells [Fig. 5; Table 1]. Although we have not isolated clones to fusion transcripts, our results suggest that the activation of the A2, F2, and J1 proviruses occurs by transcription initiation at a strong cellular promoter, read through from cellular to proviral sequences, termination in the proviral 3′ LTR, and even-

### Table 1. Promoter activity in F9 cells

| Plasmid   | Conversion (%) | PyCAT (%) |
|-----------|---------------|-----------|
| PyCAT     | 29            | 100       |
| MoCAT     | 0.6           | 2.1       |
| F2ProCAT  | 59            | 203       |
| F2RevCAT  | 0.03          | 0.1       |
| F2LPCAT   | 50            | 172       |
| J1ProCAT  | 36            | 124       |
| J1RevCAT  | 0.04          | 0.1       |
| J1LPCAT   | 9             | 31        |

Promoter fragments inserted into a CAT expression plasmid were transfected into F9 cells and assayed for CAT activity (see Experimental methods). The positive (PyCAT) and negative (MoCAT) control plasmids have been described previously [Linney et al. 1984; Barklis et al. 1986]. All other constructs represent replacements of the M-MuLV promoter fragment in MoCAT by F2 or J1 genomic DNA fragments. F2ProCAT contains the 313-bp F2 genomic XbaI–RsaI [see Fig. 5A], 5′ to 3′ upstream of the CAT gene, whereas F2RevCAT contains the same fragment in the reverse orientation. F2LPCAT differs from F2ProCAT in that it contains an additional 1500 bp of F2 genomic DNA upstream of the XbaI site [up to the EcoRI site in Fig. 1]. J1ProCAT contains the 214-bp J1 genomic HindIII–RsaI fragment (~200 to +14 from Fig. 5B) inserted 5′ to 3′ upstream of the CAT gene, whereas the same fragment was inserted in J2RevCAT in reverse orientation. J1LPCAT differs from J1ProCAT in that it contains an additional 3400 bp of J1 genomic DNA upstream of the HindIII site. CAT activity is indicated as percent chloramphenicol conversion and as percent positive control. The F2RevCAT and J2RevCAT constructs were tested once; numbers for the others represent averages of two separate experiments.
tual splicing of cellular splice donor and proviral acceptor sites. The mechanism proposed is essentially the converse of proto-oncogene activation by readthrough from viral to cellular sequences [see Herman and Coffin 1987].

Although the F2 and J1 promoters each function well in both EC cells [Table 1] and in differentiated cells [data not shown], there is little similarity in the sequences [Fig. 5A,B] to indicate crucial cis-active promoter elements. In direct sequence comparisons, the greatest homology between promoters occurs between TATA boxes and transcription start sites where the sequences are as follows: TGNCTGCGC[AG][AG]TCCTCTCTCA (Fig. 5A, −35 to −6; Fig. 5B, −40 to −12). Notwithstanding, this region is devoid of consensus sequences to well-studied nuclear-binding proteins [Jones et al. 1988]. In fact, aside from TATA and CCAAT boxes, the J1 promoter has only two recognizable factor-binding sites: Sp1 sites [Kadonaga et al. 1987] at −152 to −147 and −68 to −63 [Fig. 5B]. The stronger F2 promoter [Table 1] has no apparent Sp1 sites but has nearly consensus octamer [Sen and Baltimore 1986; Fig. 5A, −152 to −147] and E2F [Kovesdi et al. 1986; Fig. 5A, −190 to −183] sites. The latter of these sites is intriguing, as E2F activity is enriched in adenovirus-infected cells, as well as undifferentiated F9 cells [Imperiali et al. 1984]. Meanwhile, the presence of a putative octamer site in an F9 active promoter is perplexing, because the EC octamer-binding factor, NF-A3 [Lenardo et al. 1989], has a proposed repressor role. It will be of interest to dissect F2 and J1 promoter function in these cells.

Our assessment that activated proviruses mark transcriptionally active genes in EC cells, is supported by the J1 proviral insertion into a recognizable gene. The cJ1-1300 cDNA clone is highly homologous to the yeast ribosomal protein L3 gene [Schultz and Friesen 1983], at both the nucleotide [data not shown] and amino acid [Fig. 6] levels. Excluding gaps, the putative J1 protein and the L3 protein are identical at 260 of 385 residues (68%), and the majority of substitutions are conservative. Be-

**Figure 6.** Homology of the putative J1 protein to the yeast ribosomal protein L3. The DNA sequence of the 1350-bp cJ1-1300 cDNA clone from its 5' end [+1 on Fig. 5B] through the poly(A) tail is shown at the top of each three lines. Directly below each line of nucleotide sequence is the deduced amino acid sequence of the putative J1 protein, and below the putative J1 protein sequence is the amino acid sequence of the yeast ribosomal protein L3 [Schultz and Friesen 1983]. Dashes in the yeast sequence indicate identities with the deduced J1 amino acid sequence, while brackets indicate amino acid deletions of the yeast sequence as compared to the J1 sequence. Excluding gaps, the two protein sequences are 68% identical.
Retrovirus activation by cellular promoters

cause the J1 genomic probe detects a single-copy gene [Barklis et al. 1986], we infer that the J1 locus is the murine ribosomal L3 protein gene. The localization of an expressed provirus in a highly transcribed housekeeping gene is consistent with our model of proviral activation. Interestingly, in one study where strong cellular promoters were selected by enhancer/promoter trap mechanisms [Fried et al. 1983], another a ribosomal protein gene, the ribosomal protein L7a gene encoded at the mouse *surfett* locus [Galllongo et al. 1989], was selected. Given a predictable activation mechanism such as the one we observe for MP10 proviruses in EC cells, it should be possible to develop retroviral vectors similar to recombinant bacterial transposons [Castilho et al. 1984; Way et al. 1984; Manoil and Beckwith 1985] and more convenient than transfection vectors [Gossler et al. 1989] for the isolation of active genes and promoters in different tissues.

Experimental methods

Cell culture and CAT assays

The F9 cell line [Bernstine et al. 1973] and NIH-3T3 fibroblasts were grown as described previously [Barklis et al. 1986]. F9 cell clones harboring the expressed A2, F2, and J1 proviruses were selected as described [Barklis et al. 1986]. NIH-3T3 cell populations expressing the MP10 provirus were generated by infection and G418 selection by the method of Cepko et al. [1984]. Transfection of cells with CAT constructs was by the method of Graham and van der Eb [1973], as modified by Parker and Stark [1979]. CAT assays were performed according to Gorman et al. [1985].

Recombinant libraries, plasmids, and viruses

The MP10 virus, and the A2, F2, and J1 proviral fragments have been described previously [Barklis et al. 1986]. SP6 *neo*, a plasmid for the generation of single-stranded RNA probes complementary to neomycin transcripts, was the gift of Dr. R. Cone (Cone et al. 1987). MoCAT and PyCAT (originally called ΔMO plus PyF101-CAT) have been described [Linney et al. 1984]. F2ProCAT, F2PrevCAT, F2LPCAT, J1ProCAT, J1RevCAT, and J1LPCAT, described in detail in the legend to Table 1, derive from MoCAT by replacement of the MoCAT M-MuLV promoter fragment with F2 and J1 genomic DNA fragments. The F9 and mouse embryo cDNA libraries each contained EcoRI-linked cDNAs cloned into *λgt10* and were generously provided by Dr Hee-Sup Shin and Dr. A. Stacey. Initially, flanking sequence probes were used to screen a λgt10 cDNA library from undifferentiated, uninfected F9 cells. From our initial screen, one F1 cDNA clone (cF1-3400) and two independent 1900-bp F2 cDNA clones (designated cF2-1900) were isolated. The 3400-base J1 cDNA clone hybridizes to only the 1350-base RNA and is terminated at one end by a chromosomally derived EcoRI site. The cF2-1900 clones retain obvious poly[A] tails and differ only slightly [14 bp] in their lengths at their 5’ ends. For secondary screens, two cDNA libraries were screened with probes from the cF1-3400 and cF2-1900 cDNA clones. Libraries were the aforementioned F9 cDNA library and a λgt10 cDNA library of 14-day mouse embryo RNA, probes were the 5’ 1600 bp from the cF2-1900 cDNA and the 5’ 1500 bp from the cF1-3400 cDNA. From these screens, virtually identical 800-bp F2 clones (designated cF2-800), 2400-bp J1 clones (designated cJ1-2400), and 1350-bp J1 clones (designated cJ1-1300) were isolated from each library. Generally, inserts in λ clones were subcloned into the plasmid vectors pGEM4 [Promega], Bluescribe, or Bluescript [Stratagene] for detailed analysis.

Hybridizations

RNA for Northern blot hybridizations was prepared from cell lines by the method of Chirgwin et al. [1979]. Cellular DNA for DNA blot hybridizations was isolated by the method of Jaenisch et al. [1983]. Ordinarily, probes for hybridizations were radiolabeled by either nick-translation [Landtlear et al. 1982] or by the random primer method of Feinberg and Vogelstein [1983]. For preparation of the high specific activity single-stranded RNA [ssRNA] neomycin probe, the method of Melton et al. [1984] was employed, using the SP6 *neo* plasmid of Cone et al. [1987], linearized at the EcoRI site. RNA blotting and hybridizations were performed as described by Barklis and Lodish [1983]. DNA blot hybridizations were as described by Southern [1975]. The λgt10 library screens were performed using the method of Benton and Davis [1977]. RNA sizes were calculated by comparison with mobilities of the following RNA standards: cytoplasmic β-actin [2000 base; Nudel et al. 1983], CHO-A [1170 base; Harpold et al. 1979], and cyclophilin [1000 base; Danielson et al. 1988].

Cloning and sequencing procedures

All nucleic acid restriction enzymes, modification enzymes, and polymerases were from Bethesda Research Laboratories, Boehringer–Mannheim Biochemicals, New England Biolabs, Promega, Stratagene, or U.S. Biochemical. All cloning procedures were performed essentially as described by Maniatis et al. [1982]. Deletions for DNA sequencing were generated by the method of Henikoff [1984], and sequencing was from double-stranded plasmids using the chain terminator procedure of Sanger et al. [1977].

Acknowledgments

This work was supported by grant MV416 from the American Cancer Society (National chapter) and by a grant from the Oregon Division of the American Cancer Society to E.B. Support to R.J. was from National Cancer Institute grant R35644339-03. We are grateful to Dr. Hee-Sup Shin for the F9 cDNA library and Dr. A. Stacey for the mouse embryo cDNA library. We thank R. Petersen for excellent technical assistance and J. Cossey for preparation of the manuscript.

Note added in proof

Sequence data described in this paper have been submitted to the EMBL/GenBank Data Libraries under accession number Y00225.

References

Barklis, E., R.C. Mulligan, R. Jaenisch. 1986. Chromosomal position or virus mutation permits retrovirus expression in embryonal carcinoma cells. *Cell* 47: 391–399.

Barklis, E., B. Pontius, K. Barfield, and H. Lodish. 1985. Structure of the promoter of the *Dictyostelium discoideum* pre-spore EB4 gene. *Mol. Cell Biol.* 5: 1465–1472.

Barklis, E. and H. Lodish. 1983. Regulation of *Dictyostelium discoideum* mRNAs specific for prespore or prestalk cells. *Cell* 32: 1139–1148.
Benton, N. and R.W. Davis. 1977. Screening lambdagt recombinant clones by hybridization to single plaques in situ. *Science* 196: 180.

Bernstine, E., M. Hooper, S. Grandchamp, and B. Ephrussi. 1973. Alkaline phosphatase activity in mouse teratoma. *Proc. Natl. Acad. Sci.* 70: 3899–3903.

Bhat, K., M.W. McBurney, and H. Hamada. 1988. Functional cloning of mouse chromosomal loci specifically active in embryonal carcinoma stem cells. *Mol. Cell Biol.* 8: 3251–3259.

Bradley, A., M. Evans, M.H. Kaufman, and E. Robertson. 1984. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* 309: 255–256.

Castilho, B.E., P. Olfson, and M.J. Casadaban. 1984. Plasmid insertion mutagenesis and lac gene fusion with mini-mu bacteriophage transposons. *J. Bacteriol.* 158: 488–495.

Cepko, C., B. Roberts, and R. Mulligan. 1984. Construction and applications of a highly transmissible murine retrovirus shuttle vector. *Cell* 37: 1053–1062.

Chirgwin, J., A. Przybyla, R. MacDonald, and W. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 18: 5294–5299.

Conc, R., A. Bernarous-Weber, D. Baorto, and R.C. Mulligan. Regulated Expression of a complete human beta-globin gene encoded by a transmissible retrovirus vector. *Mol. Cell Biol.* 7: 887–897.

Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and C. Chambon. 1980. Promoter sequences of eukaryotic protein-coding genes. *Science* 209: 1406–1414.

Danielson, P., S. Forss-Petter, M. Browing, L. Calavetta, J. Douglass, R. Milner, and J. Sutcliffe. 1988. pL1B5: A CDNA clone of the rat mRNA encoding cyclophilin. *DNA* 7: 261–267.

Dierks, P., A. van Ooyen, M.D. Cochran, C. Dobkin, J. Reiser, and C. Weissmann. 1983. Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit beta-globin gene in mouse 3T6 cells. *Cell* 32: 695–706.

Dorn, A., J. Bolleken, A. Staub, C. Benoist, and D. Mathis. 1987. A multiplicity of CCAAT box-binding proteins. *Cell* 50: 863–872.

Evans, M.J. and M.H. Kaufman. 1981. Establishment in culture of pluripotent cells from mouse embryos. *Nature* 289: 154–156.

Feur, G. P. Vogelstein. 1983. A technique for radiolabeling restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132: 6–13.

Feur, C., 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.

Fried, M., M. Griffiths, B. Davies, G. Bjursell, G. La Mantia, and L. Lania, L. 1983. Isolation of cellular DNA sequences that allow expression of adjacent genes. *Proc. Natl. Acad. Sci.* 80: 2117–2121.

Gautsch, J. and M. Wilson. 1983. Restriction of Moloney murine leukemia virus growth in teratocarcinoma: Involvement of factors other than DNA methylation. *Cold Spring Harbor Conf. Cell Proliferation* 10: 363–378.

Giamlongo, A., J. Yan, and M. Fried. 1989. Ribosomal protein L7a is encoded by a gene (surf-3) within the tightly clustered mouse surfet locus. *Mol. Cell Biol.* 9: 224–231.

Gorman, C., P. Rigby, and D. Lane. 1985. Negative regulation of viral enhancers in undifferentiated embryonic stem cells. *Cell* 42: 519–526.

Gossler, A., T. Doetschman, R. Korn, E. Serfling, and R. Kemler. 1986. Transgenesis by means of blastocyst-derived embryonic stem cell lines. *Proc. Natl. Acad. Sci.* 83: 9065–9069.

Gossler, A., A. Joyner, J. Rossant, and W. Skarnes. 1989. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* 244: 397–504.

Graham, R. and A. van der Eb. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virol.* 52: 456–467.

Harbold, M., R. Evans, M. Sallditt-Georgieff, and J. Darnell. 1979. Production of MRNA in Chinese hamster cells: Relationship of the rate of synthesis to the cytoplasmic concentration of nine specific mRNA sequences. *Cell* 17: 1023–1035.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* 28: 351–359.

Herman, S.A. and J.M. Coffin. 1987. Efficient packaging of read-through RNA in ALV:Implications for oncogene transcription. *Science* 236: 845–848.

Hilberg, F., C. Stocking, W. Ostertag, and M. Grez. 1987. Functional analysis of a retroviral host-range mutant: Altered long terminal repeat sequences allow expression in embryonal carcinoma cells. *Proc. Natl. Acad. Sci.* 84: 5232–5236.

Imperiale, M.J., H. Kao, T. Feldman, J.R. Nevins, and S. Strickland. 1984. Common control of the heat shock gene and early adenosine genes: evidence for a cellular E1A-like activity. *Mol. Cell Biol.* 4: 867–874.

Jaenisch, R. and A. Berns. 1977. Tumor virus expression during mammalian embryogenesis. In *Concepts in mammalian embryogenesis* (ed. M. Sherman), pp. 267–314. MIT Press Cambridge.

Jaenisch, R., K. Harbers, A. Schnieke, J. Lahler, I. Chumakov, D. Jahner, D. Grotkopp, and E. Hoffmann. 1983. Germine integration of Moloney murine leukemia virus at the Mov13 locus leads to recessive lethal mutation and early embryonic death. *Cell* 32: 209–216.

Jones, N., P. Rigby, and E. Ziff. 1988. Trans-acting protein factors and the regulation of eukaryotic transcription: Lessons from studies on DNA tumor viruses. *Genes Dev.* 2: 267–281.

Kadowaga, J.T., K.R. Carner, S.R. Masiarz, and R. Tjian. 1987. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* 51: 1079–1090.

Kovesdi, I., R. Reichel, and J.R. Nevins. 1986. Identification of a cellular transcription factor involved in E1A trans-activation. *Cell* 45: 219–228.

Landfair, S., P. Lefebvre, S. Chung, and H. Lodish. 1982. Transcriptional control of gene expression during development of Dictyostelium discoideum. *Mol. Cell. Biol.* 2: 1417–1426.

Lenardo, M.J., L. Staudt, P. Robbins, A. Kung, R.C. Mulligan, and D. Baltimore. 1989. Suppression of the IgH enhancer in teratocarcinoma cells associated with a novel octamer factor. *Science* 243: 544–546.

Linney, E., B. Davis, J. Overhauser, E. Chao, and H. Fan. 1984. Nonfunction of a Moloney murine leukemia virus regulatory sequence in F9 embryonal carcinoma cells. *Nature* 308: 470–472.

Loh, T.P., L.L. Silvert, and R.W. Scott. 1987. Proviral sequences that restrict retroviral expression in mouse embryonal carcinoma cells. *Mol. Cell. Biol.* 7: 3775–3784.

Maniatis, T., E. Fritsch, and J. Sambrook. 1982. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Manoil, C. and J. Beckwith. 1985. TnphoA:A transposon probe
for protein export signals. Proc. Natl. Acad. Sci. 92: 8129–8133.

Martin, G. 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. 78: 7632–7638.

Melton, D., P. Krieg, M. Rebagliati, T. Maniatis, K. Zinn, and M. Green. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. 12: 7035–7056.

Moreland, R., H.G. Nam, L. Hereford, and H. Fried. 1985. Identification of a nuclear localization signal of a yeast ribosomal protein. Proc. Natl. Acad. Sci. 82: 6561–6565.

Mount, S. 1982. A catalogue of splice junction sequences. Nucleic Acids Res. 10: 459–471.

Nudel, U., R. Zakut, M. Shani, S. Neuman, Z. Levy, and D. Yaffe. 1983. The nucleotide sequence of the rat cytoplasmic β-actin gene. Nucleic Acids Res. 11: 1759–1771.

Parker, B. and G. Stark. 1979. Regulation of Simian virus 40 transcription: sensitive analysis of the RNA species present early in infection by virus or viral DNA. J. Virol. 31: 360–369.

Robertson, E., A. Bradley, M. Kuehn, and M. Evans. 1986. Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vectors. Nature 323: 445–447.

Sanger, F., S. Nichen, and A. Colsen. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. 74: 5463–5467.

Schultz, L. and J. Friesen. 1983. Nucleotide sequence of the tcm1 gene [ribosomal protein L3] of saccharomyces cerevisiae. J. Bacteriol. 155: 8–14.

Sen, R. and D. Baltimore. 1986. Multiple nuclear factors interact with the immunoglobulin enhancer sequences. Cell 46: 705–716.

Sherman, M.I. 1975. Differentiation of teratoma cell line PCC4.aza 1 in vitro. In Teratomas and differentiation [ed. M.I. Sherman and D. Solter], pp. 189–205. Academic Press, New York.

Southern, E. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98: 503–517.

Sorge, J., A. Cutting, V. Erdman, and J. Gautsch. 1984. Integration-specific expression in embryonal carcinoma cells. Proc. Natl. Acad. Sci. 81: 6627–6631.

Stewart, C., H. Stuhlman, D. Jahner, and R. Jaenisch. 1982. De novo methylation, expression and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proc. Natl. Acad. Sci. 79: 4098–4192.

Stickland, S. and V. Mahdavi. 1978. The induction of differentiation in teratocarcinoma stem cells by retinoic acid. Cell 15: 393–403.

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. 82: 2422–2426.

Tetch, N., R. Weiss, G. Martin, and D. Lowy. 1977. Virus infection of murine teratocarcinoma stem cell lines. Cell 12: 973–982.

Way, J., M. Davis, D. Morisato, D. Roberts, and N. Kleckner. 1984. New Tn10 derivatives for transposon mutagenesis and for construction of LacZ operon fusions by transposition. Gene 32: 369–379.

Weil, H., E. Barkliss, W. Ostertag, and R. Jaenisch. 1987. Two distinct sequence elements mediate retroviral gene expression in embryonal carcinoma cells. J. Virol. 61: 2742–2746.
Retrovirus activation in embryonal carcinoma cells by cellular promoters.

I Peckham, S Sobel, J Comer, et al.

Genes Dev. 1989, 3:
Access the most recent version at doi:10.1101/gad.3.12b.2062