

Relationship between Retroviral DNA Integration and Gene Expression

JOANNE BARNES WEIDHAAS, ELIZABETH LLOYD ANGELICHIO,
SABINE FENNER, AND JOHN M. COFFIN*

*Department of Molecular Biology and Microbiology, Tufts University,
Boston, Massachusetts 02111*

Received 2 March 2000/Accepted 16 June 2000

Although retroviruses can integrate their DNA into a large number of sites in the host genome, factors controlling the specificity of integration remain controversial and poorly understood. To assess the effects of transcriptional activity on integration in vivo, we created quail cell clones containing a construct with a mini-gene cassette, whose expression is controlled by the papilloma virus E2 protein. From these clones we derived transcriptionally active subclones expressing the wild-type E2 protein and transcriptionally silent subclones expressing a mutant E2 protein that binds its target DNA but is unable to activate transcription. By infecting both clones and subclones with avian leukosis virus and using a PCR-based assay to determine viral DNA integration patterns, we were able to assess the effects of both protein binding and transcriptional activity on retroviral DNA integration. Contrary to the hypothesis that transcriptional activity enhances integration, we found an overall decrease in integration into our gene cassette in subclones expressing the wild-type E2 protein. We also found a decrease in integration into our gene cassette in subclones expressing the mutant E2 protein, but only into the protein binding region. Based on these findings, we propose that transcriptionally active DNA is not a preferred target for retroviral integration and that transcriptional activity may in fact be correlated with a decrease in integration.

Integration, or the insertion of a double-stranded DNA copy of the viral genome into the hosts' genomic DNA, is a central event in the retrovirus life cycle. While the DNA breaking and joining reactions mediating integration are biochemically well understood (5, 6, 7, 9, 10, 18), the determinants of retroviral integration site selection have been difficult to elucidate.

In vitro integration systems have provided a powerful tool with which to study the determinants of integration site preferences on the DNA level. These assays have shown that hot spots for integration can be created by changes in local DNA structure, such as by the methylation of a run of alternating CpG dinucleotides (17) or by the creation of nucleosome-associated regions of DNA in minichromosomal DNA (26, 27). Favored integration sites in nucleosome-associated regions were shown to be due to DNA bending (24), with the most distorted sites within the nucleosome core being the most preferred for integration (25). Consistent with this idea, several DNA binding proteins known to create sharp bends in their target DNA, such as the *Escherichia coli* integration host factor, also create hot spots for integration within their binding site regions (3). By contrast, the binding of some other DNA binding proteins, such as bacterial transcriptional repressors, have been shown to suppress integration in the vicinity of their binding sites (28). Despite the wealth of information from in vitro systems, the effect of DNA binding proteins on integration into chromosomal DNA has never been determined.

Attempts to study integration in vivo have been difficult due to the scarcity of integration events in the large mammalian genome. Early in vivo studies with murine leukemia virus and avian sarcoma-leukosis virus found that integration was not sequence specific and that a large number of sites in the host

genome could serve as integration targets (5, 39). Other in vivo studies have suggested a specificity in target site selection for certain regions of the chromosome, such as those that are transcriptionally active (31) or those associated with other features, such as DNase I hypersensitivity (11, 29, 30, 40). All of these early in vivo studies suffered from potential biases such as small sample sizes, the isolation of stably integrated proviruses, and the selection of cloned proviruses. A system was designed in our laboratory that enabled study of large numbers of integration events by using a virus with a selectable marker and creating libraries of clones with provirus together with host flanking sequences. Analysis of these libraries found a small number of highly preferred sites for integration (33). However, recent work by Carreau et al. studying integration site libraries from human immunodeficiency virus-infected cells found no evidence for highly preferred sites or for any increase in the efficiency of integration near transcriptionally active DNA (8).

Most recently, a PCR-based assay was developed in our laboratory that enabled study of integration into newly infected cells and avoided any possible biasing of observed results through cloning (42). This assay was sensitive enough to detect a single integration event within a population of 5 million cells, enabling the study of a large pool of unselected integration events simultaneously. Initially, the assay was used to study integration into 11 randomly chosen regions of the avian genome. It was found that while all of the regions tested were used for retroviral integration at a frequency not significantly different from that expected for random, certain nucleotide positions within these regions were used at up to 280-fold more than random frequency. We hypothesized from these findings that while all or most regions of the genome were accessible for integration, strong integration site preferences could be determined at the local DNA level. These initial studies were unable to determine what role, if any, transcriptional activity of target DNA or protein binding had on retroviral integration.

In this report, we describe a study in which the primary goal

* Corresponding author. Mailing address: Department of Molecular Biology and Microbiology, Tufts University, 136 Harrison Ave., Boston, MA 02111. Phone: (617) 636-6528. Fax: (617) 636-8086. E-mail: jcoffin_par@opal.tufts.edu.

was to determine and separate the roles of transcriptional activity and protein binding on retroviral integration *in vivo*. Our strategy was to establish cell lines carrying a minigene, the expression of which could be regulated by the presence or absence of an appropriate transcriptional regulator, the bovine papillomavirus (BPV) E2 protein, and to then monitor integration patterns into this minigene as a function of the level of E2-stimulated transcriptional activity. In addition, a mutant of E2 that could bind its target DNA without activating transcription allowed us to separate the effects of protein binding from transcriptional activity on retroviral integration.

We found that E2-mediated activation of the transcription of our minigene led to an overall decrease in integration events both within the E2 binding region and within the actively transcribed gene. In contrast, expression of the mutant E2 protein led to a decrease in integration only into the regulator's binding region, with no change in integration frequency within the untranscribed gene. In agreement with earlier *in vitro* work, our findings show that protein binding *in vivo* can suppress integration in the vicinity of protein binding sites. However, contrary to earlier predictions, our findings also suggest that transcriptional activity is not associated with increased retroviral integration and in fact might be associated instead with a decrease in integration frequency *in vivo*.

MATERIALS AND METHODS

Cells and virus. The QT6 cell line used in this study was originally derived from chemically induced tumors of Japanese quail (22). The virus used, RAV-1, is a replication-competent exogenous member of the avian sarcoma-leukosis virus genus which was rescued from a molecular clone (32).

Plasmids. pMJG1, a kind gift from M. Grosse, was derived from pTKGH (Allegro Scientific) by the insertion of three E2 binding sites between the *NdeI* and *HindIII* sites in the multiple cloning site of pTKGH (1). Plasmids pCGE2 and pCGE2 340-G (4) are derivatives of pBluescript (Stratagene) and pSP65 (Promega). They both contain a cytomegalovirus (CMV)-driven version of the entire 410-amino-acid E2 protein and are identical except for a single C-to-G amino acid change at position 340. Plasmid pCB60-95 (2), a kind gift from J. A. T. Young, contains the neomycin gene cassette and was used for cotransfection and selection with pMJG1. Plasmid pME18SHyg B contains the hygromycin gene cassette and was used for cotransfection and selection with the pCGE2 plasmids.

Derivation of QT6 clones and subclones. Clones and subclones were derived from QT6 cells by transfection with the plasmids described above using the Lipofectamine method (Gibco BRL). For each transfection, 2.7 μ g of the desired construct and 0.3 μ g of the selectable DNA were incubated in 300 μ l of Lipofectamine for 45 min. A 2.7-ml amount of serum-free Dulbecco's modified minimal essential medium was then added, and this mixture was placed on the cells for 22 h before serum-containing medium was added. Clones were selected with neomycin (300 to 500 μ g/ml) and subclones were selected with hygromycin B (200 to 400 μ g/ml). Appropriate clones and subclones were expanded and infected with RAV-1 as described below.

Southern analysis of clones and subclones. Genomic DNA was digested with restriction enzymes, transferred to nylon membranes, and hybridized to a random-primer labeled probe derived from pMJG1 according to standard protocols (21). Three different sets of enzymes were used to ensure that the entire plasmid was present and to determine the copy number. The first digestion was with *XmnI* and *SacI* to ensure that the upstream portion of the plasmid was present, the second was with *EcoRI* to ensure that the downstream end was present, and the third was with *SacI* alone to determine the copy number.

Detection of the E2 protein. The E2 protein was detected by Western blotting using an ECL (enhanced chemiluminescence) kit (Amersham). The membrane (NEN) containing the samples of interest was rocked for 1 h with the primary antibody, a monoclonal mouse antibody to the E2 protein (B202; a kind gift from D. Breiding), which was diluted 400-fold in 5 ml of Tris-buffered saline-0.2% Tween plus 2.5% milk. Next, the secondary antibody, a horseradish peroxidase-labeled mouse antibody that binds to the primary antibody, was added to the membrane at a 1,000-fold dilution in 5 ml of the same solution. The membrane was rocked for 1 h at room temperature and washed according to the manufacturer's instructions. The results were visualized by autoradiography.

Infection of cells. Large amounts of infectious virus were produced by first infecting a plate of QT6 cells with 1 ml of frozen RAV-1 stock plus 1 \times Polybrene (15 μ g/ml). The primary infected plate was expanded, and supernatants were monitored for the level of reverse transcriptase activity. When infected cells were efficiently producing virus (two to three passages) and were almost confluent, the medium was replaced with 8 to 9 ml of fresh medium. Supernatant from these

cultures was collected 16 to 18 h later, filtered through a 0.22- μ m pore size filter, and immediately used to infect the test cells. The cells to be infected were plated at a density of 2×10^6 per 100-mm-diameter culture dish in 11 ml of medium 16 to 18 h before infection. They were subjected to three rounds of infection as follows. The medium was removed immediately before infection, and 2 ml of RAV-1-containing supernatant with 1 \times Polybrene was added to the cells for 45 min. At the end of this incubation, 8 ml of regular uninfected growth medium was added for 45 min. The second round of infection was identical to the first, and the third differed in that no Polybrene was added to the viral supernatant before it was added to the cells. To minimize selection for or against cells with specific integration sites as well as minimize reduplication of integration events by cell division, the genomic DNA for study was collected 2 to 3 days after infection by standard procedures (21).

Immunoassay. Production of the human growth hormone (hGH) protein by cells was first assayed using an immunological assay (hGH-Transient Gene Expression System kit; Nichols Institute Diagnostics), which used two antibodies to the hGH protein, one of which was avidin labeled and one of which was labeled with 125 I. Serum from the cells was first combined with the two antibodies, and then a biotin-coated bead was added. This mixture was rocked for 4 h at room temperature. The beads were washed and then counted in a gamma counter, and the amount of hGH per sample was determined.

RNase protection assay. Total cellular RNA from clones and subclones was isolated using an RNeasy kit (Qiagen) according to the manufacturer's recommendations. RNA was then eluted in water, treated with DNase I, and repurified using the RNeasy kit prior to A_{260/280} determination. RNA was divided into appropriate amounts and frozen at -70°C until use.

A [32 P]UTP-labeled antisense riboprobe was generated from pBluescript containing a 110-bp fragment from the fourth exon of the hGH gene by using a Riboprobe Systems kit (Promega) according to the manufacturer's recommendations. Threefold dilutions of cellular RNA in the range of 1.5 to 22.5 μ g (brought to equal total RNA levels with tRNA) were hybridized to the riboprobe (50,000 cpm/sample) for 16 h at 45°C . The samples were treated with RNases A (4 μ g/ml) and T₁ (11 U/ml) for 45 min at 30°C and then with a sodium dodecyl sulfate (SDS)-proteinase K solution for 30 min at 37°C . Samples were then extracted with phenol-chloroform, ethanol precipitated, resuspended in 95% formamide loading buffer, preheated, and loaded onto a 5% polyacrylamide gel containing 8 M urea. Gels were analyzed with a Storm PhosphorImager (Molecular Dynamics). Bands were quantitated using Imagequant software (Molecular Dynamics).

PCR assay to detect *in vivo* integration. PCRs were performed using 20 μ g of the infected cell genomic DNA, an amount equivalent to approximately 10^7 cells. At our estimated multiplicity of infection of two to three proviruses per cell, we predicted approximately 2.5×10^7 integration events in every 20 μ g of DNA analyzed per PCR. Since the haploid genome is 10^9 bp, and we were examining approximately 200-bp regions in each experiment, we expected to see only two to three integration events per copy of the target DNA in each region analyzed by PCR.

The PCRs for analysis of integration site distribution within a given region were prepared as follows. Genomic DNA isolated from infected or uninfected QT6 clones and subclones was diluted to 1 μ g/ μ l, heated at 100°C for 5 min, and then placed in a 80°C heating block. Twenty microliters (20 μ g) of the DNA was added to 50 μ l of a reaction mixture (10 mM Tris-HCl [pH 8.3], 3 mM MgCl₂, 50 mM KCl, 0.01% gelatin, 411 μ M each deoxynucleoside triphosphate, 0.6 μ M each primer [DNA specific and virus specific]) and 3.75 U of *Taq* polymerase (AmpliTag; Cetus-Perkin Elmer), overlaid with 50 μ l of mineral oil, and prewarmed to 80°C for 5 min. The reaction mixtures were transferred directly into a PCR machine preheated to 80°C , heated to 94°C for 5 min, and then amplified for 29 cycles at 94°C for 1 min, 69°C for 1.5 min, and 72°C for 2 min. For the final step in the last cycle, the samples were heated to 72°C for 3 min. The entire PCR mixture was then purified using a QIAquick PCR purification kit (Qiagen) according to the manufacturer's directions and elution in a final volume of 50 μ l of 10 mM Tris-HCl (pH 8.3).

PCR products were visualized by extension of an end-labeled primer. Ten microliters of each purified PCR product was dried and annealed with approximately 0.2 pmol of an internally nested γ - 32 P-labeled primer (10^6 counts per reaction) in 1 \times reaction buffer (40 mM Tris-HCl [pH 7.5], 20 mM MgCl₂, 50 mM NaCl). Extension was carried out for 30 min at 42°C . Samples were analyzed on a prewarmed 6% polyacrylamide denaturing gel under standard conditions, with sequencing ladders derived from each region run in parallel to provide size standards. These gels were then dried for 30 min and exposed to a Phosphor-Imager screen overnight.

Analysis of integration events. Analysis was performed using the Imagequant software in conjunction with a Storm PhosphorImager, both from Molecular Dynamics. Band intensity was determined by densitometry, and the total number of integration events per region was calculated. For each region, the same number of PCRs was analyzed for clones and subclones to compare the number of integration events.

Oligonucleotides. Oligonucleotides used in this study (Table 1) were selected using the PRIMER version 0.5 program (20), which selects primer pairs compatible with specific reaction component concentrations and annealing temperatures. Primers were synthesized and purified by M. Berne (Tufts University).

TABLE 1. Oligonucleotides for PCR, primer extension, and sequencing^a

Name	Sequence of primer used for:	
	PCR	Primer extension or sequencing
U3-RAV	ATCGTCGTGCACAGTGCCTTT	
JB-1	CTTGTCTGTAAGCGGATGCCG	
JB-4		GGGAGCAGACAAGCCCGT
S-4 ^a		GCGTATCACGAGGCCCTTT
JB-10	CGTTGCTCGCGTTTGCTG	
JB-11		GTTTGCTGGCGGTGTCCC
S-11 ^a		CGAATTCGGTAGGGTAGCTC
JB-12	CCGCTTAACAGCGTCAACAGC	
JB-13		ACAGCGTGCCGCAGATCC
S-13 ^a		CAGATGCAGTCGGGGCGG
JB-14	TCCTGGAGCAGGGAGAGTGC	
JB-15		GTGCTGGCCTCTTGCTCTCC
S-15 ^a		ATGTGAGTATCGCCATGTAAGC
JB-16	CGAAAATGCAGGCAGATGAGC	
JB-17		CACGCTGAGTGAGGTTCCCA
GPDH-PCR ^b	GGGAGAGATGGTAAAAGTCGGA	
GPDH-P.E. ^b		GTCAACGGATTTGGCCGTATTG

^a Used for sequencing.^b Complementary to the first intron region of the quail glyceraldehyde-3-phosphate gene (41).

RESULTS

Experimental design. We chose to use the BPV E2 protein as our transcriptional regulator, since it requires only a simple promoter to activate downstream transcription, and there are known E2 mutants that bind their target DNA without activating transcription (12). The E2 protein binds as a dimer to its target sequence in DNA and when bound to multiple binding sites can enhance downstream transcription (13, 23, 34, 35). E2-mediated transcriptional enhancement is believed to require interaction with at least one additional cellular factor, such as Sp1 (19, 38), which is thought to assist in the recruitment of TFIID to the promoter site (14). The mutant E2 protein (E2 340-G) was shown to be identical to the wild-type

(wt) E2 protein in that it is DNA binding competent, dimeric, and localized to the nucleus but unable to support transcription (12).

Clones of QT6 cells were created by transfection with plasmid pMJG1 (12), consisting of a minigene cassette with the hGH cDNA sequence under the control of an E2-dependent transcription control element (Fig. 1). The hGH protein is a useful reporter because its mRNA is quite stable, it is secreted by the transfected cells into the medium, and it can be measured using a simple immunological assay.

Subclones were derived from clones following transfection with plasmids expressing either the wt or mutant E2 protein under control of the CMV early gene promoter (4). Transcriptionally active subclones (wt subclones) with at least a three-fold induction of hGH as determined by the immunoassay were successfully isolated from two different clones (clones 1 and 2 [Table 2]); a subclone containing the mutant E2 protein (mutant subclone) was also isolated from clone 2. The presence and copy number of pMJG1 were confirmed to be identical by Southern analysis between clones and subclones used in this study (data not shown), with three copies of the plasmid in clone 1 and nine in clone 2. Some of the copies of pMJG1 in clone 2 appeared to be in tandem. The presence of the E2 protein in subclones was confirmed by Western analysis (Fig. 2).

To confirm and better assess the level of transcriptional activity in our subclones, in addition to the immunological assay we used an RNase protection assay to directly quantitate levels of hGH mRNA. From analysis of three separate experiments, we found by densitometry that both wt subclones had

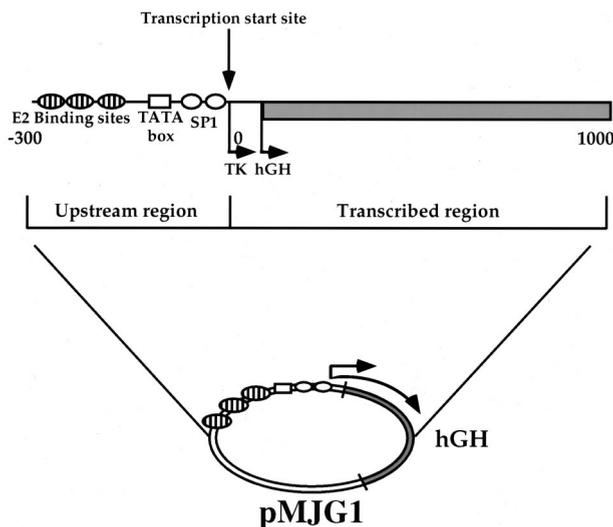


FIG. 1. The E2-driven expression system. Plasmid pMJG1 was used to create clones by transfection of QT6 cells. The plasmid contains three E2 binding sites, followed by a TATA box, two Sp1 binding sites, the herpes simplex virus TK gene promoter, and the hGH coding region, as indicated. The arrow above the expanded region shows location of the transcription start site. The TK gene promoter cap site and hGH translation initiation site are depicted with arrows below the line.

TABLE 2. Properties of QT6 clones and subclones

Clone no.	hGH (ng/ml) ^a			pMJG1 copy no. ^b
	Parent	wt subclone	Mutant subclone	
1	0.35	1.2	ND ^c	3
2	0.85	6.0	1.4	9

^a As measured by the hGH immunoassay (see Materials and Methods).^b As determined by Southern analysis (see Materials and Methods).^c ND, not determined.

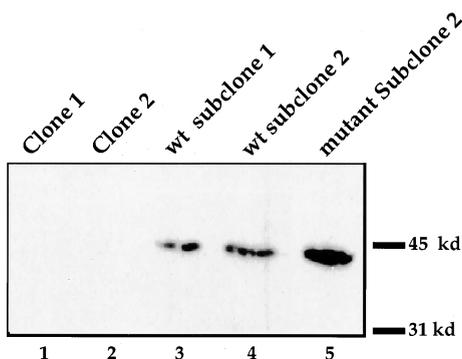


FIG. 2. Expression of E2 protein in subclones. Extracts from clones and subclones transfected with the wt or mutant E2 expression construct were analyzed by SDS-polyacrylamide gel electrophoresis followed by Western blotting and ECL detection of E2 protein using a monoclonal antibody (B202) for the BPV E2 protein that binds to both wt and mutant E2 protein (both 42 kDa).

an approximate five- to sevenfold induction of hGH expression, and our mutant subclone had approximately twice the level of the hGH RNA expression as did its parent clone (Fig. 3). These findings were similar to the results which we found using our immunologic assay and confirm that there is a direct increase in transcription of our hGH reporter in the presence of the wt E2 protein.

Detection of integration events. To determine patterns and frequencies of integration into the hGH gene, we isolated DNA from the clones and subclones 2 to 3 days after infection and subjected it to the PCR assay previously described by Withers-Ward et al. (42) (Fig. 4). This assay is exquisitely sensitive, enabling detection of a single molecule resulting from a specific integration event against a background of millions of events at other sites. PCR was performed with a primer

complementary to one of five sites within pMJG1 and a primer complementary to the viral long terminal repeat. An example of integration patterns obtained using this assay (Fig. 5) shows the results of integration into the JB-10 region of clone 2 (left) and into its wt subclone (right). The patterns shown were derived from replicate PCR amplifications of three independent infection experiments. Each band represents a single PCR-amplified radiolabeled integration event from one infected cell, with a darker intensity indicating multiple integration events at the same location (i.e., from different cells in the pool). As observed previously (42), the integration patterns are highly nonrandom, with some sites used quite frequently and others not at all. The use of many of the same sites for integration in multiple analyses from separate infection experiments, as indicated by arrows, supports the conclusion that these sites represent local hot spots for integration, not fortuitous reduplication by cell division. As is clearly visible in Fig. 5, while some hot spots remain unchanged between the parent clone and its wt subclone, there were significantly fewer integration events in the presence of the wt E2 protein than in its absence. There were also distinct changes in the distribution of integration sites when E2 was present, discussed in detail below.

We analyzed a large segment of the minigene, beginning 5' of the E2 binding sites and extending 1 kb into the hGH coding region. For each region we analyzed a total of three to eight PCRs from three separate infections and pooled the results. Comparing overlapping results from contiguous regions, we found that the same events were amplified with different primers (data not shown). To better visualize and analyze our results, we used PhosphorImager analysis to quantitate the position and intensity of the bands. Below we present separately the results obtained for integration events upstream and downstream of the transcription start site.

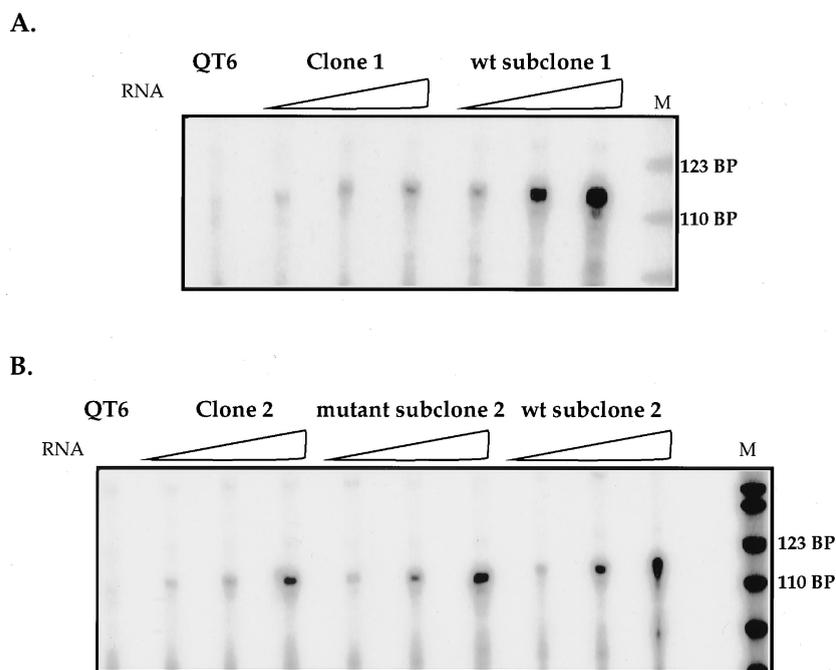


FIG. 3. Expression of hGH RNA in clones and subclones. Total RNA from clone 1 (A) and clone 2 (B) and their E2-expressing subclones was extracted and annealed with a ³²P-labeled riboprobe. The protected fragment after RNase digestion was analyzed by polyacrylamide gel electrophoresis. Levels of cell RNA used were 2.5, 7.5, and 22.5 μg for clone 1 and its subclone and 1.5, 4.5, and 13.5 μg for clone 2 and its subclones. Gels were analyzed with a PhosphorImager. M, size markers.

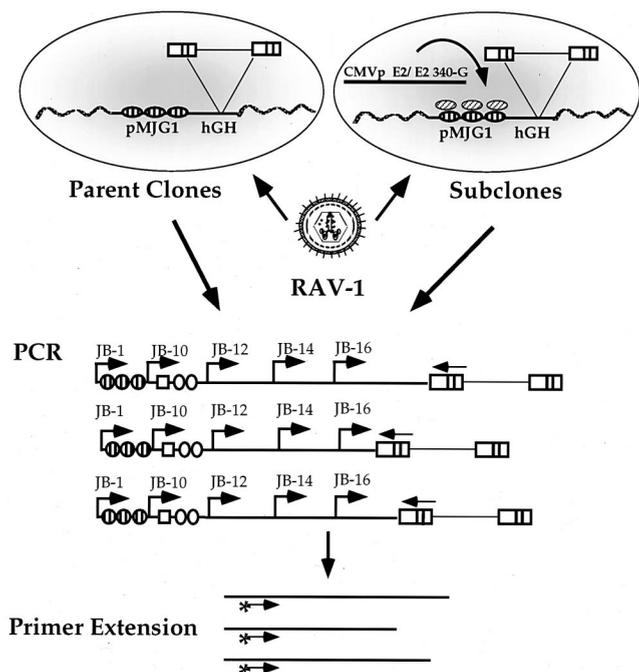


FIG. 4. Detection of integration events in clones and subclones. Clones and subclones were infected with RAV-1, the genomic DNA was collected, and PCR was performed. Two oligonucleotide primers, one derived from a sequence in pMJG1 (JB-*n*) and the other from a sequence in the viral U3 region (U3-RAV), were used to amplify integration events. The resulting PCR products were used as templates to extend an end-labeled primer, and the samples were separated by polyacrylamide gel electrophoresis.

Integration upstream of the transcription start site. The integration patterns into the area upstream of the transcription start sites in clones 1 and 2 and their respective subclones are shown in Fig. 6. In contrast to previous predictions, we found that the transcriptionally active wt subclones showed no enhancement of integration but in fact showed a noticeable decrease in integration events into this region. In general, the differences in integration that we observed between each parent clone and its wt subclone were common to the two sets and were as follows: first, a decrease in integration upstream and within the E2 binding sites; second, a decrease in integration into the areas upstream of the TATA box; third, a loss of integration directly into the Sp1 sites; and finally, a decrease in integration in the region of the transcription start site.

In the mutant subclone 2, integration into the upstream region had features resembling both the wt subclones and the parent clones. First, in the area containing the E2 binding sites as well as in the area upstream of the TATA box, there was a general decrease in integration similar to that seen in the wt subclones. However, in the region containing the two Sp1 sites as well as in the area 5' of the transcription start site, the mutant subclone, similar to the parent clones, showed a heavy use of this region for integration.

Integration into the hGH coding region. We next examined the frequency of integration downstream of the transcription start site, as shown for clone 1 and its wt subclone in Fig. 7. Again, no increase in integration targeting resulting from transcriptional activity could be observed. In fact, there was a decrease in the amount of integration into this area compared to the parent clone. These results were confirmed by analysis of clone 2 and its wt subclones. By contrast, levels of integration into the untranscribed hGH gene in the mutant subclone 2

were not significantly different from integration levels in parent clone 2 (data not shown).

Integration into nonregulated sequences. Although we knew from the Southern analysis that the number and distribution of our hGH minigene constructs were identical among the clones and their subclones, we also wanted to ensure that the differences in integration frequency observed were specific for transcriptional activity and not due to some unknown effect of the E2 protein on infectability or on integration in general. We therefore first compared the amounts of integrated viral DNA between clones and subclones by Southern analysis and found that the levels of integrated viral DNA were equivalent (data not shown), implying the absence of significant differences in all early steps of infection.

Next, to address the issue of possible nonspecific effects on integration more directly, we compared patterns of integration between clones and subclones into the glyceraldehyde-3-phosphate dehydrogenase (GPDH) gene, a multicopy housekeeping gene (41). As can be seen in Fig. 8, the number of integration events into the GPDH gene did not vary significantly with expression of the wt or mutant E2 protein, as it did in the hGH minigene in the same experiment.

To further protect against distortion of our analysis by any minor differences among clones and subclones, we used the number of integration events into the GPDH gene to normalize the relative frequency of integration into our minigene. With these normalized values, we were able to calculate the relative amount of integration between the subclones and their parent clones (Fig. 9). These calculations revealed that integration into the upstream regions of both clone 1 and clone 2 was reduced by approximately 60% in the presence of the wt

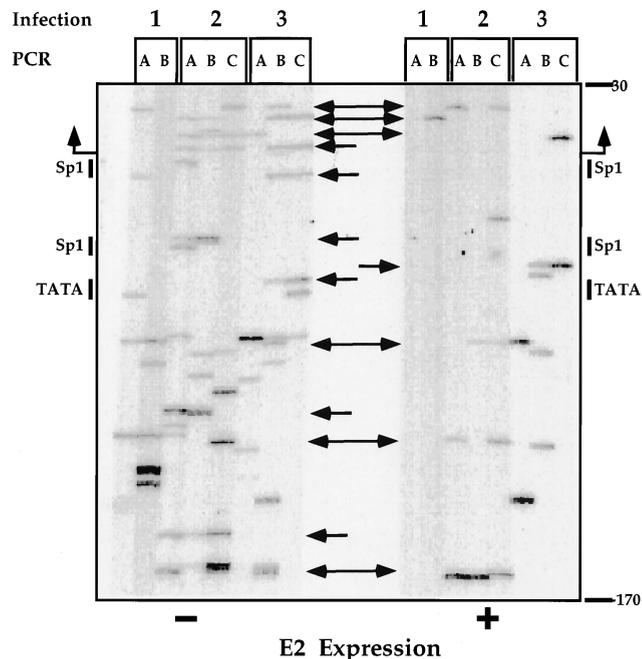


FIG. 5. Pattern of integration into the upstream region of pMJG1. Parent clone 2 and its wt subclone were infected with RAV-1, and the integration site distribution was determined by PCR with primer JB-10 (Fig. 4). Locations of the TATA box, Sp1 sites, and transcription start site (arrow) are shown. DNA from three different infection experiments (1, 2, and 3) was divided into two or three samples of 20 μ g each (A, B, and C) and analyzed in independent PCRs. The double-headed arrows indicate hot spots for integration conserved between the clone and subclone, and single-headed arrows indicate hot spots found in only the clone or the subclone.

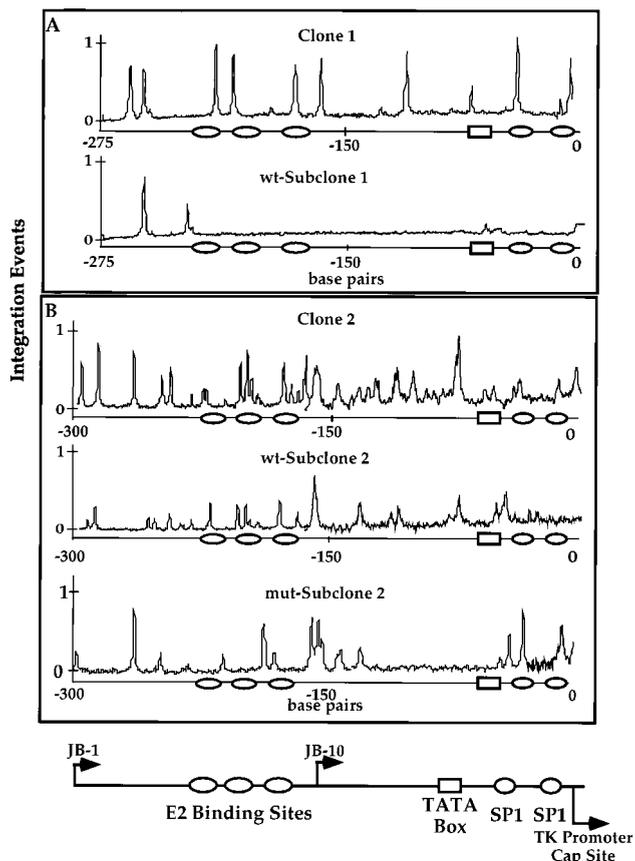


FIG. 6. Effects of the E2 protein on integration into the upstream region. The integration patterns and frequencies into the upstream region of pMJG1 of clone 1 (A) and clone 2 (B) and their subclones as determined by densitometry are shown as the sum of the products of all like reactions divided by the number of reactions analyzed. The location along the pMJG1 construct is shown with the E2 binding sites, TATA box, and Sp1 sites marked. The location in base pairs is shown on the abscissa, with 0 representing the transcription start site. The right half of the pattern in panel B is derived from the gel shown in Fig. 5. mut-Subclone 2, mutant subclone 2.

E2 protein and by approximately half that amount in the presence of the mutant E2 protein. In the transcribed region, in the presence of the wt E2 protein (and increased transcriptional activity) there was a smaller but still significant decrease in integration, whereas in the presence of the mutant E2 protein there was no effect on integration.

DISCUSSION

Transcriptionally active DNA has long been hypothesized to be preferred for integration (30, 31, 40), based on studies of small numbers of selected events and the logic that transcriptionally active DNA would provide a more suitable environment for expression of the integrated DNA provirus. The experiments described here were designed to test this hypothesis by directly assessing the effects of transcriptional activity and protein binding on retroviral integration into chromosomal DNA. In our model system, we found that enhancement of transcriptional activity as well as protein binding without transcriptional stimulation did not lead to enhancement of integration. Rather, both were associated with an overall decrease in integration events. While we did not study the difference in integration between DNA in heterochromatin and euchroma-

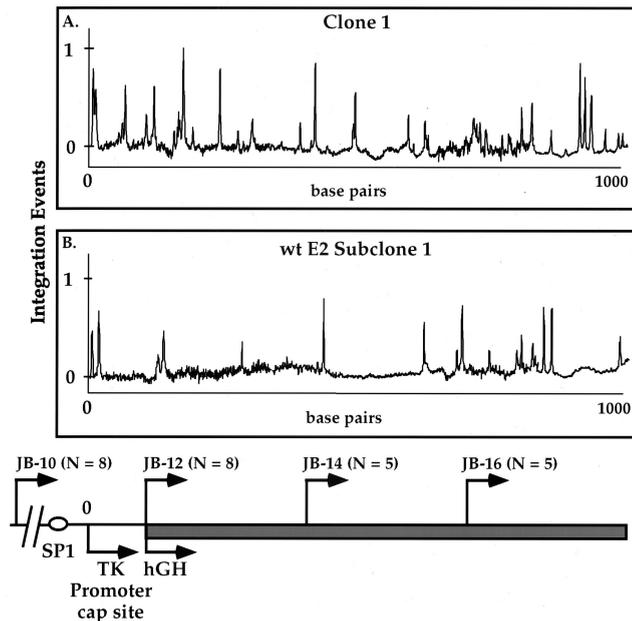


FIG. 7. Integration into the transcribed region of clone 1. Results from PCR analysis of integration into the 1,000 bp downstream of the transcription start site in DNA from infected clone 1 (A) and its wt subclone (B) are plotted as described for Fig. 6. The results from different primers are again normalized for comparison.

tin in this model, our results clearly do not support the hypothesis of a positive link between increased transcriptional activity and targeting of integration.

The minigene system. The model that we chose for this study was a simple minigene construct consisting of an hGH cDNA reporter sequence downstream of the herpes simplex virus

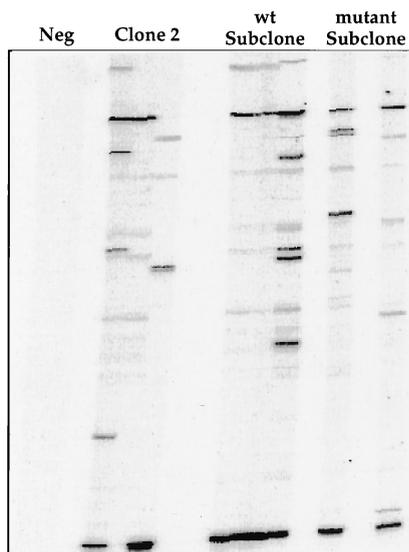


FIG. 8. Integration into a housekeeping gene in clones and subclones. The pattern and frequency of integration of RAV-1 DNA into the GPDH gene in clone 2 and its subclones are shown. The negative control on the left is DNA from infected QT6 cells not transfected with pMJG1. PCR was performed using the U3-RAV and GPDH-PCR primers (Table 1). The DNA used for analysis was one sample from each of the two or three infections. Similar results were obtained with clone 1 and its subclones (data not shown).

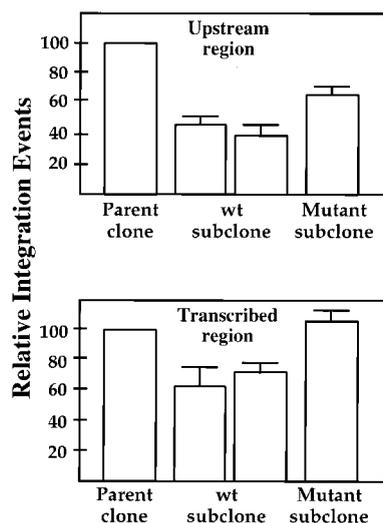


FIG. 9. Summary of integration into pMJG1. The integration events were normalized for each type of infected clone by dividing the number of events into the indicated regions of pMJG1 by the number of integration events into GPDH. The standard errors were calculated based on the differences in the relative frequency of integration among regions amplified with separate PCR primers.

thymidine kinase (TK) gene promoter (which provides the cap site, TATA sequence, and two Sp1 sites). Expression of this gene is controlled by binding of the BPV E2 protein to an array of three binding sites upstream of the TK gene promoter. This artificial gene system was chosen over naturally inducible sequences (such as metallothionein or heat shock genes) because of its compactness and simplicity and out of concern that inducing agents such as heavy metal or heat might affect integration directly. It also allowed us to use a mutant E2 protein to separate effects due to protein binding from those due to transcriptional activation. Finally, its presence in multiple copies in the cell lines tested allowed us to collect more integration events than into a single-copy gene. Also, the presence of multiple copies of the target sequence reduced concern over possible effects due to specific positions. The similarity of our results in two separate sets of clones and subclones gives strong support that our results are correct and not due to effects of location in the genome or copy number. Thus, although the construct is artificial, and we cannot rule out that different interactions with specific transcriptional control elements might occur in some genes, we are confident that our results will apply in a general way to most or all natural genes. Direct analyses of integration into other types of cellular sequences are under way in our laboratory.

Decrease in integration in the factor binding region. In the region upstream of the transcription start site, the presence of the E2 protein was associated with an overall decrease in integration events in both the wt subclones and the mutant subclone compared to their parental clones, although the loss was greater in the wt subclones. We hypothesize that the loss of integration into this area reflects binding of the transcriptional initiation complex to the DNA, which thereby blocks accessibility to the integration machinery. Indeed, binding of transcriptional regulatory proteins to DNA has been shown to interfere with integration of retroviral DNA *in vitro* (28). In the wt subclones, the decrease in integration included the entire area upstream of the TATA box extending through the Sp1 sites up to the transcription start site. In the mutant subclone, by contrast, we did not see a loss of integration into the Sp1

sites or the region 5' of the transcription start site. We hypothesize that the difference between the two may reflect the inability of the mutant E2 protein to successfully recruit the Sp1 protein, and thereby the transcriptional initiation complex, which would also explain its failure to activate transcription.

Integration within the coding region. The region downstream of the transcription start site also showed a decrease in integration in the presence of the E2 protein, but only in our wt subclones. This effect was not seen in the mutant subclone, even though there was a twofold increase in the basal level of hGH expression in the presence of the mutant protein. Although proteins must bind to the DNA to induce transcription, the absence of a decrease in integration in the mutant subclone suggests that the decrease in integration seen in the wt subclones was due to their transcriptional activity. The decrease in integration observed could be due to direct interference of integration by the transcriptional apparatus itself or might reflect indirect effects, such as displacement of nucleosomes (36, 37) and loss of associated hot spots. It could also reflect additional changes in the conformation of the DNA with transcriptional activity not yet appreciated.

Effects of DNA structure on integration. Changes in DNA structure, particularly bends due to the association with nucleosomes (27) or introduced by DNA binding proteins (3), have been shown to introduce hot spots for integration of retroviral DNA *in vitro*. Similar hot spots have also been observed in phased chromatin-associated DNA in cells (26). We did not observe the creation of obvious hot sites as a function of E2 binding or transcriptional activation, although E2 is known to introduce bends into its DNA target (23). We would propose that *in vivo* such effects of individual proteins are blurred by the binding of additional proteins and protein complexes. Our observation of decreased integration into transcriptionally active DNA supports *in vitro* work showing that nucleosomal DNA is preferred for integration, since transcriptionally active DNA has been shown to be dynamic and involves shifting of nucleosomes resulting in out-of-phase nucleosomes (36, 37).

Transcriptional activity and integration. Based on the results of this study, we consider it improbable that there is any specific interaction of the retroviral integration apparatus, either with a component of the transcription machinery or with a transcription-associated change in DNA structure. A well-established example of the former interaction is found in the case of the retrovirus-like Ty3 element of yeast, where a specific interaction of the preintegration complex and a polymerase III-specific transcription factor directs integration to the upstream region of tRNA genes (16). The finding that a protein related to a generalized yeast transcription factor can interact with human immunodeficiency virus type 1 integrase (15) has been taken to suggest a similar effect for retroviruses, but the relevance of this interaction in cells remains to be established.

These studies have provided insight into the long-standing question of the effects of transcriptional activity on retroviral integration. We found that transcriptionally active DNA is not preferred for integration over the same DNA when it is less active. Rather, increasing transcription led to a decrease in integration, most likely due to direct or indirect blocking of integration by components of the transcriptional machinery. This approach may eventually provide a useful tool for analysis of DNA chromosomal structure *in vivo* and lead to a better understanding of the changes in DNA structure that occur during transcription.

ACKNOWLEDGMENTS

We thank Elliot Androphy, Marty Grosse, and David Breiding for supplying components of the E2 system and the E2 antibody. We also thank John Strasswimmer for assistance with Western blot analysis of the E2 protein and John Aschoff for guidance and materials for the RNase protection assay.

This work was supported by grant R35 CA from the National Cancer Institute to J.M.C. J.M.C. is American Cancer Research Professor of Molecular Biology and Microbiology.

REFERENCES

- Barsoum, J., S. S. Prakash, P. Han, and E. J. Androphy. 1992. Mechanism of action of the papillomavirus E2 repressor: repression in the absence of DNA binding. *J. Virol.* **66**:3941–3945.
- Bates, P., J. A. T. Young, and H. E. Varmus. 1993. A receptor for subgroup A Rous sarcoma virus is related to the low density lipoprotein receptor. *Cell* **74**:1043–1051.
- Bor, Y.-C., F. D. Bushman, and L. E. Orgel. 1995. In vitro integration of human immunodeficiency virus type 1 cDNA into targets containing protein-induced bends. *Proc. Natl. Acad. Sci. USA* **92**:10334–10338.
- Breiding, D. E., F. Sverdrup, M. J. Grosse, N. Moscuro, W. Boonchai, and E. J. Androphy. 1997. Functional interaction of a novel cellular protein with the papillomavirus E2 transactivation domain. *Mol. Cell. Biol.* **17**:7208–7219.
- Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop. 1987. Correct integration of retroviral DNA in vitro. *Cell* **49**:347–356.
- Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop. 1989. Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proc. Natl. Acad. Sci. USA* **86**:2525–2529.
- Bushman, F. D., T. Fujiwara, and R. Craigie. 1990. Retroviral DNA integration directed by HIV integration protein in vitro. *Science* **249**:1555–1558.
- Carteau, S., C. Hoffmann, and F. Bushman. 1998. Chromosome structure and human immunodeficiency virus type 1 cDNA integration: centromeric alphoid repeats are a disfavored target. *J. Virol.* **72**:4005–4014.
- Engelman, A., K. Mizuuchi, and R. Craigie. 1991. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* **67**:1211–1222.
- Fujiwara, T., and K. Mizuuchi. 1988. Retroviral DNA integration: structure of an integration intermediate. *Cell* **54**:497–504.
- Goodenow, M. M., and W. S. Hayward. 1987. 5' long terminal repeats of *myc*-associated proviruses appear structurally intact but are functionally impaired in tumors induced by avian leukemia viruses. *J. Virol.* **61**:2489–2498.
- Grosse, M. J., J. Barsoum, S. S. Prakash, and E. J. Androphy. 1996. The BPV-1 E2 DNA-contact helix cysteine is required for transcriptional activation but not replication in mammalian cells. *Virology* **217**:301–310.
- Haugen, T. H., L. P. Turek, F. M. Mercurio, T. P. Cripe, B. J. Olson, R. D. Anderson, D. Seidl, M. Karin, and J. Schiller. 1988. Sequence-specific and general transcriptional activation by the bovine papillomavirus-1 E2 transactivator require an N-terminal amphipathic helix-containing E2 domain. *EMBO J.* **7**:4245–4253.
- Hoey, T., R. O. J. Weinzierl, G. Gill, J.-L. Chen, B. D. Dynlacht, and R. Tjian. 1993. Molecular cloning and functional analysis of *Drosophila* TAF110 reveal properties expected of coactivators. *Cell* **72**:247–260.
- Kalpana, G. V., S. Marmon, W. Wang, G. R. Crabtree, and S. P. Goff. 1994. Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science* **266**:2002–2006.
- Kirchner, J., C. M. Connolly, and S. B. Sandmeyer. 1995. Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. *Science* **267**:1488–1491.
- Kitamura, Y., Y. M. H. Lee, and J. M. Coffin. 1992. Nonrandom integration of retroviral DNA in vitro: effect of CpG methylation. *Proc. Natl. Acad. Sci. USA* **89**:5532–5536.
- Lee, Y. M. H., and J. M. Coffin. 1990. Efficient autointegration of avian retrovirus DNA in vitro. *J. Virol.* **64**:5958–5965.
- Li, R., J. D. Knight, S. P. Jackson, R. Tjian, and M. R. Botchan. 1991. Direct interaction between Sp1 and the BPV enhancer E2 protein mediates synergistic activation of transcription. *Cell* **65**:493–505.
- Lincoln, S. E., M. J. Daly, and E. S. Lander. 1991. Primer: a computer program for automatically selecting PCR primers. Massachusetts Institute of Technology, Cambridge, Mass.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Moscovici, C., M. G. Moscovici, H. Jimenez, M. M. C. Lai, M. J. Hayman, and P. K. Vogt. 1977. Continuous tissue culture cell lines derived from chemically induced tumors of Japanese quail. *Cell* **11**:95–103.
- Moskaluk, C., and D. Bastia. 1988. DNA bending is induced in an enhancer by the DNA-binding domain of the bovine papillomavirus E2 protein. *Proc. Natl. Acad. Sci. USA* **85**:1826–1830.
- Muller, H.-P., and H. E. Varmus. 1994. DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. *EMBO J.* **13**:4704–4714.
- Pruss, D., F. D. Bushman, and A. P. Wolffe. 1994. Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proc. Natl. Acad. Sci. USA* **91**:5913–5917.
- Pryciak, P. M., H.-P. Muller, and H. E. Varmus. 1992. Simian virus 40 minichromosomes as targets for retroviral integration in vivo. *Proc. Natl. Acad. Sci. USA* **89**:9237–9241.
- Pryciak, P. M., A. Sil, and H. E. Varmus. 1992. Retroviral integration into minichromosomes in vitro. *EMBO J.* **11**:291–303.
- Pryciak, P. M., and H. E. Varmus. 1992. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell* **69**:769–780.
- Robinson, H. L., and G. C. Gagnon. 1986. Patterns of proviral insertion and deletion in avian leukemia virus-induced lymphomas. *J. Virol.* **57**:28–36.
- Rohdewohld, H., H. Weiher, R. Wolf, 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.
- Scherdin, U., K. Rhodes, and M. Breindl. 1990. Transcriptionally active genome regions are preferred targets for retrovirus integration. *J. Virol.* **64**:907–912.
- Sealy, L., M. L. Privalsky, G. Moscovici, C. Moscovici, and J. M. Bishop. 1983. Site-specific mutagenesis of avian erythroblastosis virus: erb-B is required for oncogenicity. *Virology* **130**:155–178.
- Shih, C.-C., J. P. Stoye, and J. M. Coffin. 1988. Highly preferred targets for retrovirus integration. *Cell* **53**:531–537.
- Spalholz, B. A., P. F. Lambert, C. L. Yee, and P. M. Howley. 1987. Bovine papillomavirus transcriptional regulation: localization of the E2-responsive elements of the long control region. *J. Virol.* **61**:2128–2137.
- Spalholz, B. A., Y.-C. Yang, and P. M. Howley. 1985. Transactivation of a bovine papilloma virus transcriptional regulatory element by the E2 gene product. *Cell* **42**:183–191.
- Studitsky, V. M., D. J. Clark, and G. Felsenfeld. 1994. A histone octamer can step around a transcribing polymerase without leaving the template. *Cell* **76**:371–382.
- Studitsky, V. M., D. J. Clark, and G. Felsenfeld. 1995. Overcoming a nucleosomal barrier to transcription. *Cell* **83**:19–27.
- Ushikai, M., M. J. Lace, Y. Yamakawa, M. Kono, J. Anson, T. Ishiji, S. Parkkinen, N. Wicker, M.-E. Valentine, I. Davidson, L. P. Turek, and T. H. Haugen. 1994. *trans* activation by the full-length E2 proteins of human papillomavirus type 16 and bovine papillomavirus type 1 in vitro and in vivo: cooperation with activation domains of cellular transcription factors. *J. Virol.* **68**:6655–6666.
- Varmus, H. E., and R. Swanstrom. 1984. Replication of retroviruses, p. 369–512. In R. A. Weiss et al. (ed.), *RNA tumor viruses*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Vijaya, S., D. L. Steffen, and H. L. Robinson. 1986. Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. *J. Virol.* **60**:683–692.
- Wiskirchen, R., G. Siemeister, M. Hartl, and K. Bister. 1993. Sequence and expression of a glyceraldehyde-3-phosphate dehydrogenase-encoding gene from quail embryo fibroblasts. *Gene* **128**:269–272.
- Withers-Ward, E. S., Y. Kitamura, J. P. Barnes, and J. M. Coffin. 1994. Distribution of targets for avian retrovirus DNA integration in vivo. *Genes Dev.* **8**:1473–1487.