Retrovirus-mediated insertion of expressed and non-expressed genes at identical chromosomal locations

Brent Berwin and Eric Barklis*

Vollum Institute for Advanced Biomedical Research and Department of Microbiology and Immunology, Oregon Health Sciences University, Portland, OR 97201, USA

Received February 16, 1993; Revised and Accepted April 19, 1993

ABSTRACT

During retrovirus replication, a cellularly derived tRNA is annealed to the viral RNA at the primer binding site (PBS) to prime reverse transcription, and both the tRNA and the PBS become copied and matched together on complementary proviral DNA strands prior to integration. Using a viral PBS single base pair mutant which affects provirus expression in undifferentiated cells, we show that reversion to wild type (wt) occurs at a frequency of approximately 50%. Daughter cell lines containing wt or mutant proviruses at identical chromosomal sites have been isolated, supporting a model where an integrated PBS-mismatched provirus was copied before mismatch correction could occur. Virus expression in daughter cells containing the mutant provirus was 100-fold higher than in cells bearing the wt counterpart. Additionally, proviral 5' DNA and cellular 5' flanking DNA became methylated in daughter cells containing wt but not mutant integrants. These results strongly support the current model of retrovirus reverse transcription, and indicate that the wt PBS region contains an element which suppresses virus expression and directs the methylation of viral and neighboring cellular DNA.

INTRODUCTION

The Moloney murine leukemia virus (M-MuLV) uses cellular tRNA as its primer for reverse transcription (Figure 1; 1,2). The tRNA^{pro} binds via its 3' nucleotides to an eighteen base pair primer binding site (PBS), viral nucleotides 146–163, just 3' of the 5' long terminal repeat (LTR; 1,2,3). During reverse transcription, the tRNA^{pro} primes synthesis of the minus strand to the 5' end of the viral genome RNA where synthesis stops and the first template jump occurs (see Figure 1). The newly synthesized DNA minus strand R region anneals to a homologous 3' R viral RNA region and minus strand synthesis continues to the 5' end of the viral PBS. Plus strand synthesis then is initiated at the polypurine tract (PPT) at the 5' end of the 3' LTR and extends downstream, apparently stopping after having copied 18 base pairs of the primer tRNA^{pro}. A second jump then results

in the annealing of the LTR regions as well as the copied portions of the tRNA^{pro} to the copied viral PBS region. After elongation, the double stranded DNA product is ready for integration into the host cell genome (1,2). Both biochemical (1) and genetic (4) results support the notion that the tRNA^{pro} is copied during reverse transcription, while it also has been shown that PBS mutations revert at high frequency (3,5). Theoretically, should a mutation be present in the viral PBS, the mismatched viral and tRNA^{pro} PBS copies will anneal resulting in an integrated provirus containing a mismatch (Figure 1). After integration, mismatch correction may occur before DNA replication resulting in a homologous mutant or PBS wildtype (wt) region. Upon post-integration DNA replication, if mismatch correction has not occurred, one proviral copy will be wt while the other will be fully mutant (Figure 1, bottom line). In either case, one might predict a theoretical 50% reversion to wt of an initially mutant PBS virus (5).

Ordinarily, mutations in the tRNA PBS of M-MuLV yield no scorable difference from wt virus. An exception is the B2 mutation, a single base pair change from G to A at M-MuLV nucleotide 160 (5,6). Wild type M-MuLV expression is repressed in undifferentiated embryonal carcinoma (EC) cells, such as the F9 cell line (5,7,8,9,10), while the B2 mutation in the PBS alleviates the repression of expression of M-MuLV in undifferentiated EC cells (5,10). However, differentiated cells, such as NIH 3T3 fibroblasts, do not exhibit this repression of wt M-MuLV gene expression (5,10). The wt M-MuLV repression in EC cells is due at least partly to differences in RNA levels: detectable viral RNA levels in EC cells are less than 1% of that in infected fibroblast cells (5,7,8,10,11,14). Viral proteins also are not necessary for RBS mediated repression of expression of wt M-MuLV in EC cells (5,8,12). Since the proviral PBS region mediates repression at the DNA level in undifferentiated EC cells, we refer to this second function as the repressor binding site (RBS; 12,13). Repression via the RBS appears to be independent of methylation, as it occurs prior to provirus methylation and in azacytidine treated cells (8,12,14,15), although methylation does not appear to be independent of repression (12).

In order to control for effects of chromosomal DNA on provirus expression and vice versa, we decided to see if we could

* To whom correspondence should be addressed

obtain wt and B2 versions of a retrovirus at identical insertion sites. During these studies, we have found that the B2 mutation reverts at about the predicted 50% frequency. Through the reversion of the B2 mutation, we have isolated independent cell lines containing B2 and wt versions of M-MuLV at identical integration sites in the host cells' genomes. These cell lines allowed us to control for flanking sequence effects. We have found that the wt provirus expression was repressed 100-fold relative to its mutant counterpart. Additionally, while the mutant proviral DNA and chromosomal region were unmethylated, 5' portions of the wt provirus as well as 5' flanking DNA were methylated at the tested sites. These results support the tRNA/PBS copying and mismatch pairing mechanism, and demonstrate the influence of the M-MuLV RBS on virus gene expression and proviral and chromosomal methylation.

METHODS AND MATERIALS

Cells and viruses

F9 EC cells, NIH 3T3 fibroblasts, and Psi2 (16) and PA317 (17) packaging cell lines were grown as described previously (5,18). DNA transfections were done by the method of Parker and Stark (19); infections of cells with virus stocks were performed as described by Cepko et al. (20). B2BAG producer cell lines were created either by direct transfection of Psi2 packaging cells, or by infection of Psi2 cells with media supernatants from transiently transfected PA317 cells, as described by Jones et al. (18). For selections, infected F9 and 3T3 cells were split 48 hours post-infection onto G418 at 5×10^5 cells/plate. For 3T3 cells, 0.8 mg/ml of the 50% pure G418 powder (Gibco) was used; for F9 cells, 0.5 mg/ml was used. For the purpose of cloning, infected or transfected Psi2 cells were selected, cloned with cloning rings (Belco) and expanded. For isolation of first generation F9 clones, F9 cells at 10^5 cells per 150 mm plate were infected, then selected 36 hours post-infection on the same plates without splitting: colonies were cloned and expanded as above to yield first generation clones. Second generation clones were created by splitting of first generation cells at a concentration of 50,000 cells/150mm plate and growing colonies from single cells. Colonies then were cloned with cloning rings and expanded.

The B2BAG provirus (Figure 2) is comprised of B2recirc (5) from the 5' LTR to the SpeI site at nt 730, and BAG (18,24) from the SpeI site at nt 730 through the 3' LTR. B2BAG is a 7630 bp provirus with the *E. coli* β -galactosidase gene being driven by the M-MuLV 3' LTR promoter, the pBR322 origin of replication, and the SV40 early promoter (also retaining replication origin capabilities) driving expression of the selectable neomycin gene. The B2BAG construct also contains the B2 mutation in the PBS (see text). B2BAG restriction sites include: BamHI: 1020, 4120; BclI: 2370, 4800; BglII: 4800; ClaI: 1850, 6880; EcoRI: 4030, 5800; EcoRV: 140, 220, 2140, 6820, 6910, 7170, 7240; HindIII: 4480; KpnI: 450, 7510; NheI: 30, 7060; PvuI: 860, 1140, 1880, 2350, 2810; SacI: 410, 2950, 7440; SalI: 4120; SmaI: 450, 4120, 7510; SpeI: 730, 6690; SspI: 2250; XbaI: 300, 7330; XhoI: 6400; no internal sites for DraI or SacII (Figure 2).

Enzyme assays and FACS analysis

β -Galactosidase (β -Gal) assays were as previously described by Norton and Coffin (18,21). One β -gal activity unit is 1 nMol

0-nitrophenyl- β -D-galactopyranoside (ONPG) hydrolyzed per min (21). Total protein quantification was done as described by Bradford (22). β -Gal histochemical staining, modified from established protocols (23,24), was performed as follows: adherent cells were washed with phosphate buffered saline (PS), fixed in 0.5% glutaraldehyde in PS at room temperature for 15 min, washed again with PS, drained, bathed 5 min in PS, and then treated with a solution of 5mM potassium ferricyanide, 5 mM potassium ferrocyanide, 1mM $MgSO_4$, and 1mg/ml 5-Bromo-4-chloro-3-indolyl- β -D-galactosidase (X-gal) overnight. After staining, plates were viewed at $200\times$ magnification and the percentage of blue-staining (β -gal producing) cells was determined. Microscope photography was performed by attaching a 35mm Contax 137 camera containing Kodak Tmax 400 film to a Zeiss IM35 inverted microscope, total magnification was $100\times$.

Fluorescence activated cell sorting (FACS) analysis was performed using fluorescein di- β -D-galactopyranoside (FDG; Molecular Probes) as a substrate, following the protocol of Nolan et al. (31), with the expert assistance of A.Scotch Weinberg. Briefly, one million trypsinized cells in 0.1 ml of stain solution (10 mM Hepes, pH 7.2 plus 4% fetal calf sera in PS) were incubated at $37^\circ C$ for 10 min, after which 0.1 ml of $37^\circ C$ 2 mM FDG was added. Following FDG addition, incubations continued at $37^\circ C$ for 1 min, and were stopped by addition of 1.8 ml of ice cold stain solution containing 1 microgram/ml propidium iodide. Cells were kept at $4^\circ C$ and analyzed with respect to both side scatter and fluorescein fluorescence on a Becton Dickinson cell sorter.

Labeling, hybridizations and polymerase chain reactions (PCR)

Probe labeling for Southern blot hybridization was done by the random-priming labeling method (25). The DNA probe for the 5' portion of B2BAG and corresponding flanking region was the B2BAG BamHI fragment, nt 1020–4120; the probe for the 3' portion was the B2BAG EcoRI-ClaI fragment, nt 5800–6880. The SpeI–BamHI (nt 730–1020) fragment was the probe for the Southern blot of the PCR products. DNA preparation, blotting and hybridizations were essentially as described previously (26,27,28,29). Digestions were performed as suggested by the manufacturers' guidelines, and nylon filters (Zetabind; American Bioanalytical) were used for Southern blotting.

Polymerase chain reactions (PCR) were used to amplify PBS-containing portions of the F9 clones from genomic DNA (30), and plasmid B2BAG was used as a control. The PCR primers used were an 18-mer corresponding to M-MuLV nucleotides 38–55, 5'-CGTGTATCCAATAAACCC-3', and a 17-mer corresponding to the 5' end of β -Gal, 5'-GTTGGGTAACGCC-AGGG-3'. PCR reactions of 20 μ l final volume contained 1 μ g genomic DNA, 2 units Vent DNA polymerase (New England Biolabs), 100 ng of each of the primers, and final concentrations of 10mM KCl, 10mM $(NH_4)_2SO_4$, 20mM Tris–HCl (pH 8.8), 2mM $MgSO_4$, 0.1% Triton X-100, and 0.32mM of each of the four dNTP's. Reactions were run for 50 cycles (1 minute at $94^\circ C$, 2 minutes at $55^\circ C$ and 2 minutes at $72^\circ C$) on a Techne machine. One half of each PCR reaction was precipitated, resuspended, and cut with restriction enzyme AlwI according to the manufacturer's specifications. Both cut and uncut samples were run on a 1.2% agarose gel and blotted and hybridized as described above.

Figure 1. Model for retroviral reverse transcription. The predicted outcome for reverse transcription of a Moloney murine leukemia (M-MuLV) primer binding site (PBS) mutant is illustrated. The cellularly derived tRNA primer and the PBS are depicted as shown, with the wild type (wt) PBS region indicated as a black ellipse and the mutant as a hatched ellipse. For convenience, only one of the viral RNA strands is shown as a thin black line with its 5' end at the left and a poly A tail (AAAAA). DNA strands appear as heavy black lines, and DNA or RNA R, U5, and U3 regions are noted, as is the polypurine tract (PPT) which primes plus strand synthesis. DNA strands representing the host cell chromosomal integration site are shown as white boxes with vertical black lines, and on the bottom line, proviruses are illustrated on a smaller scale, not indicating the separate DNA strands. As shown, when a PBS mutation is present on the viral RNA, it may be mismatched with a wt tRNA (line 1, top). During reverse transcription, both the PBS (line 3) and the tRNA (line 4) are copied and mismatched (line 5), resulting in the integration of a PBS-region mismatched provirus (line 6). If replication and cell division occurs before mismatch correction, wt and mutant proviruses will be found at identical sites in different daughter cells. Additional details are provided in the text. This figure was adapted from previous studies (1,2,4).

RESULTS

Reversion of the M-MuLV primer binding site B2 mutation

The B2 mutation is a G to A mutation at nt 160 in the M-MuLV primer binding site (PBS; 5). This mutation alleviates the repression of proviral gene expression in EC cells acting at the DNA level and affecting RNA levels in a virus protein-

Figure 2. Proviral map of the B2BAG construct. B2BAG is a 7630 bp proviral variant of the BAG β -galactosidase expression vector, where the Moloney murine leukemia virus (M-MuLV) long terminal repeat (LTR) promoter drives expression of the *E. coli* β -galactosidase gene and an internal SV40 early promoter also retaining replication origin capabilities (SVori) drives expression of the selectable neomycin gene. Other features of this construct are the pBR322 bacterial origin of replication (pBRori), and a mutant tRNA^{pro} primer binding sequence (PBS), where the second G of the sequence GGGATC has mutated to an A residue, creating the sequence GAGATC. This mutation occurs at position 160 of M-MuLV viral genome, and originally was referred to as the B2 mutation. Restriction site locations for all BamHI, BglII, ClaI, EcoRI, HindIII, NheI, SalI, SmaI, SspI, XbaI and XhoI sites are shown. Approximate map positions for all occurrences of selected enzymes are as follows: BamHI: 1020, 4120; Bcl: 2370, 4800; BglII: 4800; ClaI: 1850, 6880; EcoRI: 4030, 5800; EcoRV: 140, 220, 2140, 6820, 6910, 7170, 7240; HindIII: 4480; KpnI: 450, 7510; NheI: 30, 7060; PvuI: 860, 1140, 1880, 2350, 2810; SacI: 410, 2950, 7440; SalI: 4120; Sma: 450, 4120, 7510; SspI: 730, 6690; SspI: 2250; XbaI: 300, 7330; XhoI: 6400; no internal sites for DraI or SacII.

independent fashion (5,7,8,10). Kempler et al. (12) also have found that the silencer in the PBS region, referred to as the repressor binding site (RBS), has a diminished effect on distant versus proximal promoters. Because our previous studies on the function of the M-MuLV RBS involved infectivity assays which theoretically could be affected by differences in integration or integration site selection, we sought to obtain matched cell lines containing wt and mutant M-MuLV proviruses at equivalent sites. Based on the current model of retrovirus reverse transcription which entails copying of both the retroviral RNA PBS and the cellular tRNA reverse transcription primer, it seemed possible to generate wt and B2 segregants from a single mismatched provirus by the mechanism detailed in the Introduction (Figure 1). However, as an initial test, we wished to verify that the B2 mutation would revert at high frequency in our virus.

To examine the frequency of B2 reversion, the B2BAG virus (Figure 2) was chosen for analysis. B2BAG derives from the BAG virus (18,24) and is essentially a wt M-MuLV featuring the B2 mutation at nt 160, the β -galactosidase gene being driven by the M-MuLV LTR, and an internal SV40 early promoter which drives expression of the selectable neomycin gene: previously we showed that the wt RBS repressed β -gal but not neo expression in F9 EC cells in a construct similar to B2BAG (12). For reversion analysis, we directly transfected B2BAG into Psi2 virus packaging cells. Producer lines thus generated should release viruses which theoretically deliver 50% wt and 50% B2 mutant genomes. We also infected Psi2 cells with media supernatants from B2BAG transfected PA317 virus packaging cells. We deliberately allowed infected cells to divide prior to splitting onto selective media (see Methods) in order to create infected Psi2 producer clones which should make either wt BAG or B2BAG viruses. Three Psi2 transfection stocks and eleven Psi2 infection stocks were made. These stocks were used to infect F9 (EC) and 3T3 cells which were selected for G418 resistance. Then, pools of at least 100 G418 resistant clones from each infection were assayed for β -gal activity (Table 1). As expected, all Psi2 stocks produced virus that infected 3T3 cells and expressed β -gal well, since the RBS does not affect fibroblast

Figure 3. Histochemical staining for β -galactosidase expression in the R6, R6A4 and R6C2 cell lines. Primary clone R6 (A) and subclones R6A4 (B) and R6C2 (C) were histochemically stained for β -galactosidase expression analysis, and photographed under identical conditions on a Zeiss inverted microscope. The black bar indicates 200 microns.

expression. Also as expected, transfection stocks which theoretically could deliver B2 and wt virus gave high β -gal activities (300–530 U/mg) in pools of selected F9 cells. However, six of the eleven Psi2 lines produced by infection

Table 1. β -galactosidase activities of B2BAG infected cells.

Virus Source (Psi2 line made by)	β -galactosidase activity (U/mg protein) on selected cells	
	3T3 cells	F9 cells
infection clone A	1500	8
infection clone B	1900	500
infection clone C	2000	29
infection clone D	800	11
infection clone E	2200	38
infection clone F	1500	57
infection clone G	1500	500
infection clone H	2400	600
infection clone I	1200	320
infection clone J	1500	12
infection clone K	800	650
transfection pool A	400	300
transfection pool B	400	300
transfection pool C	300	530

Producer lines were generated either by transfection or infection of Psi2 cells. For transfected cell lines, Psi2 cells were transfected with the B2BAG construct, selected for G418 resistance and pooled as populations of resistant cell lines from at least 100 independent colonies per pooled population. Infected producer lines were generated by infection of Psi2 cells with media from PA317 cells which were transiently transfected with the B2BAG construct. Infected G418 resistant Psi2 lines were cloned from expanded isolated colonies. For assays, infected 3T3 or F9 cells and G418 resistant colonies (at least 100) were pooled as populations for β -galactosidase assays. β -galactosidase assays and protein determinations were performed as described in the Materials and Methods.

(clones A,C,D,E,F,J) delivered virus which transduced the neo gene but had low (0–60 U/mg) β -gal activities in F9 cells, while the remaining five (clones B,G,H,I,K) made virus that generated decent β -gal signals (300–650 U/mg) in infected F9 cells. These results were entirely consistent with a predicted 50% reversion rate for the B2 mutation.

Isolation of expressed and non-expressed proviruses at a single chromosomal site

As an attempt to isolate wt and mutant genomes at identical sites, EC cells, plated at low dilution on 150 mm maxi plates, were infected with B2BAG virus from transfected Psi2 cells (see Methods and Materials). The EC cells, without having been split, were then selected for G418 resistance. The G418 resistant colonies were cloned, expanded, and portions were histochemically stained for β -gal activity. We initially looked for heterogeneous staining of cells from a single clone as an indication of genotypic heterogeneity, although incomplete staining could yield false positives. Of the primary clones (Table 2), 4 of 16 (N8, N13, R5, R7) exhibited >99% of the cells staining blue, consistent with homogeneous cell lines containing B2 provirus. Six of the 16 primary clones (N1, N9, R13, R15, R17, R18) stained <1% blue, indicative of reversion to wt (correction to wt prior to replication). The remaining 6 primary clones (N6, N10, N15, R6, R8, R12) exhibited intermediate staining at 10% or 95%. Although none of the primary clones showed the 50% staining that one would predict if replication were to have occurred prior to correction and all progeny grew equally well, we decided to isolate pure secondary clones from R7 (a control), N6, N15, and R6 for examination by histochemical staining (Table 1). N15 and the control R7 yielded secondary clones which stained with great variability; we are unsure whether this was due to incomplete staining or if epigenetic shut-off of β -gal occurred in some, but not all, daughters of a pure clone. N6

Table 2. Histochemical staining of infected F9 clones and subclones.

primary clones		percentage of β -galactosidase positive cells in							
		N6 subclones		N15 subclones		R6 subclones		R7 subclones	
N1	<1%	A1	20%	A1	30%	A4	<1%	A1	40%
N6	10%	A2	10%	A2	50%	A6	<1%	A2	10%
N8	>99%	A3	10%	A6	30%	B1	<1%	A3	60%
N9	<1%	A4	20%	B3	30%	C2	80%	A4	70%
N10	95%	A5	10%	B6	10%	C5	<1%	B4	90%
N13	>99%	A6	10%	C2	30%	D6	<1%	B5	10%
N15	95%	B3	10%	C3	30%			C1	30%
R5	>99%	B5	10%	C4	30%			C2	20%
R6	10%	C1	20%	C5	40%			C3	60%
R7	>99%	C3	10%	D3	20%			C5	40%
R8	95%	C4	10%					D2	10%
R12	95%	C5	20%					D4	30%
R13	<1%	C6	10%					D5	20%
R15	<1%	D1	10%					D6	40%
R17	<1%	D2	10%						
R18	<1%	D3	20%						
		D5	20%						

Plates of F9 clones and subclones were histochemically stained for β -galactosidase activity as described in the Materials and Methods. After staining, plates were viewed at a 200 $\times$ magnification and the percentage of blue-staining cells out of a total of 100 cells was determined. For primary clones, results were rounded to the nearest 5%, for subclone determinations, results were rounded to the nearest 10%.

daughters showed little variability, all staining 10–20%. However, the R6 primary clone yielded 6 secondary subclones, 5 of which showed <1% staining (A4, A6, B1, C5, D6) and one (C2) in which most of the cells (80%) stained blue. We found the R6 results to be of particular interest because the percentage of blue secondary clones (1 of 6) was consistent with the percentage of blue-staining primary cells (10%), and because there was a dramatic difference between staining and non-staining cells in the primary and secondary clones. Examples of the staining differences are shown in Figure 3. Figure 3a shows that the primary R6 clone stained about 10% very blue, the rest white. However, subclone R6A4 (Figure 3b) stained completely white while subclone R6C2 (Figure 3c) stained 80% blue.

Although the histochemical staining data for the R6, R6A4 and R6C2 cell lines were consistent, fluorescence activated cell sorting (FACS) analysis provided more quantitative information on the percentages of β -gal expressing cells. To perform the FACS analysis, suspensions of cells from each of the cell lines were incubated with the substrate fluoroscein di- β -D-galactopyranoside (FDG; 31) after which the live cells were analyzed for fluorescence of the FDG product. The R6 primary cell line exhibited two populations of cells (Figure 4a): 93–94% had fluorescence equal to uninfected F9 cells, while 6–7% had a fluorescence two logs higher: this was consistent with the histochemical staining results. As expected, the R6A4 subclone showed 0% of its cells to be in the high fluorescence class (Figure 4b), whereas the R6C2 subclone showed 94% of its cells to be in the high fluorescence class (Figure 4c). Although we are unsure why the remaining 6% of the R6C2 cells did not fluoresce highly, it could be that these represented a low percentage of unfiltered dead cells, or a subpopulation of cells with epigenetically inactivated expression of β -gal (23). These FACS results were corroborated by a β -gal assay on cell lysate samples (Table 3). R6C2 exhibited high β -gal levels (over 1000 U/mg protein), similar to the B4 daughter cell line of the R7 primary clone. The R6 primary cell line showed roughly 10% of the β -gal activity of R6C2, in agreement with the FACS

Table 3. β -galactosidase activities of infected F9 clones.

F9 cell line	β -gal activity (U/mg protein)
F9 control	13.8
clone R6	163
clone R6 subclone A4	14.8
clone R6 subclone A6	14.4
clone R6 subclone B1	12.7
clone R6 subclone C2	1375
clone R6 subclone C5	0.8
clone R6 subclone D6	10.0
clone R7 subclone B4	943

F9 cells were infected with a supernatant from Psi2 cells which had been transfected with the B2BAG construct. After infection, G418 resistant cell clones were isolated and expanded, and subclones of those clones were isolated and expanded. Thus, R6 represents a primary clone from the initial infection, while R6A4, R6A6, R6B1, R6C2, R6C5, and R6D6 represent subclones of the R6 clone. R7B4 is a subclone of a separate primary clone (R7). Enzyme and protein assays were as described in Table 1.

analysis and histochemical staining. The rest of the R6 subclones (R6A4, R6A6, R6B1, R6C5, R6D6) had baseline β -gal activities.

One potential explanation as to the differences between expressing and non-expressing cells in R6, as well as the difference in expression between R6A4 and R6C2, is that the original R6 population was a mixture of two independent clones derived from independent integration events. To prove that the clones were derived from a single integration we examined the genomic integration sites by Southern blot hybridization (Figure 5). Initially we used the BamHI fragment from B2BAG as a hybridization probe. This probe does not detect a band in the control F9 cells (lane A), but detects bands of approximately 10 kbp in EcoRI-digested DNA from the R6, R6A4 and R6C2 cell lines (lanes B–D). This indicates that in the latter three cell lines, there are EcoRI sites in the chromosomal flanking DNA approximately 6 kbp upstream of the proviral 5' LTRs. The 3'

Figure 4. Fluorescence activated cell sorting (FACS) analysis of β -galactosidase expression in the R6, R6A4 and R6C2 cell lines. Suspensions of cells from clone R6 (A), subclone R6A4 (B), and subclone R6C2 (C) were incubated with the β -galactosidase substrate fluorescein di- β -D-galactopyranoside (FDG), after which live cells were analyzed for fluorescence of the fluorescein product on a Becton Dickinson sorter. Assay conditions were as described in the Materials and Methods, and sorted cells are plotted on the basis of side scatter (y axis) and fluorescence of the fluorescein product (x axis). In each panel, at least 2,000 cells were sorted. For R6, 6.46% of the cells were in the high fluorescence class; for R6A2, 0% were high fluorescence; and 93.79% of the R6C2 cells were highly fluorescent.

proviral probe (the B2BAG 3' EcoRI–ClaI fragment) showed the distances from the 3' proviral EcoRI site to the nearest 3' flanking site to be equal (about 6.2 kbp past the 3' LTR) for the proviruses in all three cell lines (lanes E–G). Similarly, the BglII fragments from the proviruses to the 5' (lanes H–J) and 3' (lanes

Figure 5. Insertion sites of R6, R6A4 and R6C2 proviruses. Genomic DNA was isolated from F9 cells (lane A) or cells containing the R6 (lanes D, G, J, M, P, S), R6A4 (lanes C, F, I, L, O, R) or R6C2 (lanes B, E, H, K, N, Q) proviruses. The genomic DNA was then digested with restriction endonucleases EcoRI (lanes A–G), BglII (lanes H–M), DraI (lanes N–P) or XbaI (lanes Q–S) for subsequent Southern blot hybridizations using proviral 5' (lanes A, B, C, D, H, I, J, N, O, P, Q, R, S) or 3' (lanes E, F, G, K, L, M) probes. The probe for 5' proviral and flanking region fragments was the B2BAG BamHI fragment, while the probe for 3' region fragments was the B2BAG EcoRI–ClaI fragment (nt 5800–6880). Migration of 23.1, 9.4, 6.6, 4.4, 2.3 and 2.0 kbp marker DNA fragments is as indicated on the left of each panel. Please note that the apparent reduced mobilities of bands in lanes F, I, and L correlate with differences in the electrophoresis of these genomic DNA samples, as evidenced by reduced mobilities of repetitive DNA bands in the ethidium bromide stained gels (not shown).

Figure 6. PCR amplification of proviral B2BAG DNA. Polymerase chain reactions (PCR), performed as described in the Methods and Materials, were used to amplify a 657 bp PBS-containing region from 1 μ g of genomic DNA isolated from the R6 clone (lane E), and its daughters R6D6 (lane B), R6C2 (lane C), and R6A4 (lane D), as well as the R7B4 cell line (lane A). B2BAG plasmid DNAs at 0.1 ng (lane F) and 1 ng (lane G) were used as controls. The reactions utilized primers corresponding to 5' LTR and 5' β -gal regions. Half of each reaction was fractionated on a 1.2% agarose gel, blotted and hybridized with a labeled probe from the B2BAG SpeI–BamHI fragment (nt 730–1020). The standard was positioned by comparison to the migration of PhiX DNA marker fragments, as indicated.

K–M) flanking sites were equal. DraI, which does not cut in the proviruses, yielded identical fragments (lanes N–P), as did 5' probed HindIII and 3' probed BamHI cuts (data not shown). In all three cell lines XbaI, which cuts in each LTR, yielded predicted identical 7 kbp fragments (lanes Q–S) indicating no

Figure 7. AlwI digestion of amplified proviral DNA. Half of each of the PCR reactions from Figure 6 was retained in order to test the diagnostic cutting ability of the restriction enzyme AlwI upon them: a wt PBS (GGATC) contains an AlwI site, while the mutant B2 PBS (AGATC) does not. PCR products were digested with AlwI, fractionated on a 1.2% agarose gel, blotted and hybridized with a probe from the B2BAG SpeI–BamHI fragment (nt 730–1020). Lane designations are as follows: A, R7B4; B, R6D6; C, R6C2; D, R6A4; E, R6; F, B2BAG plasmid; G, B2BAG plasmid. Note that the B2BAG plasmid, which is cut at nt 1020 but not in the PBS region, yields a 605 bp fragment and is designated uncut. Marker DNA migrations are as shown.

apparent differences in the proviruses. These Southern blot hybridizations indicate that the proviruses are integrated at identical chromosomal sites, consistent with the notion that they derived from a single integration event. However, we do not know why only 10% of the R6 population was positive for β -gal, rather than the 50% we predicted. Possible, but untested, explanations are that the discrepancy was an artefact of cloning; that there was selection against β -gal expressing cells during R6 expansion; and/or there was an epigenetic turn-off of β -gal expression.

While expression differences from proviruses at identical chromosomal locations were consistent with the model in Figure 1, other causes for such differences could be invoked. If provirus expression differences were due to the PBS region differences, then expressed proviruses should contain the B2 mutation while non-expressed proviruses should be wt: these could be distinguished by the ability of restriction endonuclease AlwI to cut wt (GGATC) but not the B2 mutant (AGATC) in the PBS region. To test the diagnostic cutting ability of AlwI, the proviral PBS regions were amplified by polymerase chain reactions (PCR). As shown in Figure 6, primers corresponding to 5' LTR and 5' β -gal regions amplified 657 bp fragments of the R7B4, R6D6, R6C2, R6A4 and R6 proviruses (lanes A–E). The B2BAG SpeI–BamHI fragment was used to probe these fragments, which are identical in size to the similarly amplified B2BAG plasmid (lanes F and G). In Figure 7, the PCR fragments from Figure 6 were digested with AlwI. Note that the amplified B2BAG plasmid produced a hybridizing fragment of approximately 605 bp, which corresponds to the PCR product which was cut at the BamHI site (also an AlwI site, nt 1020) but not at the PBS. We assume that the predicted other 52 bp fragment was too small for detection by blot hybridization. The R6 primary clone (lane E) gave a major band at about 475 bp (corresponding to a PCR product which was cut at both the BamHI site and the PBS), and a minor band the same size as that from the B2BAG plasmid. The small 160 bp fragment, nt

Figure 8. Methylation of proviruses and flanking regions. Genomic DNA from the R6 primary cell line (lanes C,F,I,L,O,R), and the R6A4 (lanes B,E,H,K,N,Q) and R6C2 (lanes A,D,G,J,M,P) secondary cell lines was digested overnight with SmaI (lanes A–C), EcoRI plus ClaI (lanes D–F), EcoRI plus XhoI (lanes G–I), SacII (lanes J–L) or SacII plus BglII (lanes M–O). Digestion products were electrophoresed on a 0.9% agarose gel, Southern blotted, and hybridized to either the B2BAG BamHI fragment (lanes A–O) or the B2BAG EcoRI–ClaI fragment (lanes P–R). Size standards (in kbp) were determined by comparison to the migration of lambda phage fragments.

38–168, was not represented by the probe and was not detected. The 475 bp band in lane E indicated that reversion of the B2 mutation had occurred, while the intensity of the 605 versus the 475 bp bands was consistent with the percentage of expressing to non-expressing cells in the R6 cell population. As expected, the two R6 subclones which did not express β -gal, R6D6 and R6A4 (lanes B and D), were cut by AlwI, while the expressing clones R6C2 (lane C) and R7 secondary clone R7B4 (lane A) were uncut. In summary, the appearance of the AlwI cut proves that there is PBS reversion. The cutting of the AlwI directly correlated with non-expression in the proviruses, and the simplest explanation is that differences in β -gal expression resulted from proviral PBS sequence differences.

Influence of RBS differences on provirus and flanking sequence methylation patterns

Since the R6A4 and R6C2 cell lines contained their proviruses at identical insertion points and presumably differ only in the PBS region, these cell lines provided an excellent opportunity to determine RBS effects on methylation patterns. To do so, we performed Southern blot hybridizations of genomic DNA that was cut with a variety of methylation sensitive enzymes (Figure 8). Our 5' β -gal probe indicated that the DNA methylation sensitive restriction endonuclease SmaI cut R6C2 genomic DNA to completion at both the 5' LTR and internal sites (lane A), while SmaI fragments of >20 kbp were detected from the R6 and R6A4 DNA (lanes B,C): this supports the interpretation that the 5' LTR SmaI site of the latter two cell lines was methylated, at least partially. EcoRI plus the methylation sensitive enzyme ClaI cut to completion in the R6C2 DNA (lane D, 5' probe), producing a 2.3 kbp fragment (totally proviral) and a barely visible 7.8 kbp fragment, presumably from the proviral 5' ClaI site to an EcoRI site in the 5' flanking region (based on size). The EcoRI–ClaI cut R6 DNA exhibited a faint band of 2.3 kbp, a barely visible band of 7.8 kbp (not readily

GENOMIC MAP OF THE R6, R6A4, AND R6C2 PROVIRUSES

Figure 9. Map of the R6, R6A4 and R6C2 chromosomal integration region. The chromosomal region where the R6 provirus and its daughters lie is shown. The provirus region is depicted as two boxes indicative of 5' and 3' long terminal repeat (LTR) regions, bracketing a heavy black line; chromosomal DNA is shown as a thin black line. All known and mapped sites for BglII (B), ClaI (C), EcoRI (E), SacII (S), SmaI (Sm), XbaI (Xb), and XhoI (X) are indicated. Sites tested for DNA methylation as shown in Figure 8 are marked in triangles. White triangles indicate tested sites that were unmethylated in R6A4 cells, while black triangles indicate methylated R6A4 sites.

detected), and a significant amount of ClaI-uncut DNA at 10 kbp (lane F); while the R6A4 DNA appeared to be exclusively methylated at the ClaI site (lane E). A 5' probing of EcoRI–XhoI cut R6C2 DNA indicated that cutting occurred at the internal EcoRI as well as a flanking region XhoI site to produce a 5.4 kbp fragment (lane G); the R6A4 and R6 DNA was partially methylated at the 5' flanking region XhoI site (lanes H,I). SacII does not cut in the provirus, but the 5' probe indicated a different size fragment for R6C2 DNA (lane J) than for R6A4 and R6 (lanes K,L). This indicates that at least one SacII site in the flanking region was methylated in R6 and R6A4, but not in R6C2. R6 and R6A4 DNA cut with SacII plus BglII (5' probe) gave BglII sized 5' bands of approximately 6.6 kbp (lanes N,O), but R6C2 DNA was cut at a SacII site in 5' the flanking region (lane M).

The data from Figure 8, lanes A–O, indicated methylation pattern differences between expressed and non-expressed 5' proviral and 5' chromosomal flanking regions. In contrast, all three proviruses were unmethylated at 3' sites. As shown in lanes P–R, EcoRI–XhoI cut DNA probed with a 3' fragment gave the same size bands in all cases: a 600 bp 3' EcoRI–XhoI fragment from the provirus and a 1.9 kbp fragment from the proviral XhoI to a 3' flanking XhoI site. These bands indicate that the 3' flanking and proviral XhoI sites are unmethylated in all cases. In summary (Figure 9), of six tested methylation sites, the two 3'-most sites were unmethylated in all cell lines. However, two sites in the 5' region of the provirus, and two sites in the 5' flanking region were methylated in wt background (R6A4) but not B2 background (R6C2). These results show that, integration sites being equal, the M-MuLV RBS influences the state of provirus and 5' flanking region DNA methylation.

DISCUSSION

As previously observed (3,5), PBS mutations revert at high frequency. Using this observation of PBS reversion along with the model of reverse transcription (1,2), we designed a protocol for the isolation of different daughter segregants of a single proviral integration (Figure 1). From a single integration event, we have isolated a parental clone, R6, and two different daughters, R6A4 and R6C2. R6C2 expressed β -gal at high levels, while expression was inhibited in R6A4 (Table 3; Figures 3,4). Proviral insertion occurred at an identical site in R6A4 and R6C2 (Figure 5), and the only observed difference between the cell lines was detected at the AlwI site, within the proviral PBS (Figure 7). While it is possible that other proviral changes could have occurred between R6A4 and R6C2, no differences were observed amongst the XbaI genomic cuts (Figure 5). Given our protocol, the existence of R6A4 and R6C2 is direct proof that

both the PBS and tRNA primer are copied during M-MuLV reverse transcription of a single virus. These segregants also demonstrate that mismatch correction in mammalian cells is not completely efficient.

The differences in expression observed in R6A4 and R6C2 provide information concerning the M-MuLV stem cell specific repressor binding site (RBS), which overlaps the M-MuLV PBS. The RBS effect is not due to differences in wt and mutant integration, nor is it integration site dependent. In agreement with the findings of Kempler et al. (12), we saw no apparent RBS effect on neomycin expression based on the isolation of G418 resistant EC cell colonies, suggesting that RBS silencing diminishes over a distance. It was previously shown that RBS repression did not appear to be dependent on methylation (12). However, we have shown that proviral and flanking sequence methylation patterns are influenced by the state of the RBS. In particular, the R6A4 wt RBS-containing provirus became methylated at 5' but not 3' sites, while the mutant R6C2 provirus remained relatively unmethylated (Figures 8,9). RBS differences affected methylation in 5' flanking regions, but not in 3' flanking regions (Figures 8,9). We believe that the 5' versus 3' methylation patterns in the R6A4 region could be caused by a diminution of RBS effects over distance, but also could be a consequence of selective pressure for neomycin gene expression in the original R6 clone. Theoretically, the mechanism for the RBS effect on provirus methylation could be a direct one where the wt RBS acts as a signal for de novo methylation. Alternatively the mechanism could be an indirect one, where lack of expression is a signal for methylation. Our bias is towards the indirect mechanism, although examples where provirus methylation in EC cells occurs rapidly (11) or gradually (14,15) are both available. What causes the methylation pattern differences of 5' DNA flanking the R6A4 and R6C2 proviruses? One possibility is that expression of R6C2 provirus caused demethylation of previously methylated flanking DNA. A second possibility is that R6A4 provirus' silencing caused methylation of previously unmethylated flanking DNA. We favor the second possibility, as EC genomic DNA is relatively unmethylated (32) and because the R6A2 provirus presumably underwent a transition from an unmethylated reverse transcription product to a methylated integrant.

A potential line of inquiry regarding the R6A4 cell line involves screening by FACS to obtain second site β -gal positive revertants. Also, we currently are attempting to clone out R6 flanking DNA, which should be useful in future studies on how the PBS affects proviral and flanking sequence chromatin structure in the R6A4 and R6C2 cell lines. Obviously, the method that we have described for the isolation of expressed and non-expressed

sequences at identical chromosomal locations is not a general one as has been described elsewhere (33,34). Indeed, it appears limited to retroviral PBS sequences. Nevertheless, the matched R6A4 and R6C2 cell lines should be useful in examination of M-MuLV provirus host cell DNA interactions.

ACKNOWLEDGEMENTS

We thank Oliver Nanassy for help during the early phases of this work. Our FACS analysis could not have been performed without the expert assistance of A.Scotch Weinberg. Thanks also to Breton Freitag, Mike Yamauchi and Robert Gray for reagents and discussion during the course of this work. This work was supported by grant 1-Fy91-0049 from the March of Dimes Birth Defects Foundation, grant MV416B from the American Cancer Society, and grant 1 R01 CA53332-01 from the National Cancer Institute.

REFERENCES

1. Gilboa, E., Mitra, S.W., Goff, S. and Baltimore, D. (1979) *Cell*, **18**, 93–100.
2. Weiss, R., Teich, N., Varmus, H. and Coffin, J. (eds.) (1984) *RNA Tumor Viruses*. Cold Spring Harbor Laboratory, Cold Spring Harbor.
3. Rhim, H., Park, J. and Morrow, C.D. (1991) *J. Virol.*, **65**, 4555–4564.
4. Colicelli, J. and Goff, S.P. (1986) *J. Virol.*, **57**, 674–677.
5. Barklis, E., Mulligan, R.C. and Jaenisch, R. (1986) *Cell*, **47**, 391–399.
6. Shinnick, T., Lerner, R. and Sutcliffe, J. (1981) *Nature*, **293**, 543–548.
7. Loh, T.P., Sievert, L.L. and Scott, R.W. (1987) *Mol. Cell Biol.*, **7**, 3775–3784.
8. Loh, T.P., Sievert, L.L. and Scott, R.W. (1990) *Mol. Cell Biol.*, **10**, 4045–4057.
9. Teich, N., Weiss, R., Martin, G. and Lowy, D. (1977) *Cell*, **12**, 973–982.
10. Weiher, H., Barklis, E., Ostertag, W. and Jaenisch, R. (1987) *J. Virol.*, **61**, 2742–2746.
11. Stewart, C., Stuhlman, H., Jahner, D. and Jaenisch, R. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4088–4092.
12. Kempler, G., Freitag, B., Berwin, B., Nanassy, O. and Barklis, E. (1993). *Virology*, **93**, in press.
13. Petersen, R., Kempler, G. and Barklis, E. (1991) *Mol. Cell Biol.*, **11**, 1214–1221.
14. Gautsch, J. and Wilson, M. (1983) *Cold Spring Harbor Conf. Cell Proliferation* **10**, 363–378.
15. Niwa, O., Yokota, Y., Ishida, H. and Sugahara, T. (1983) *Cell*, **32**, 1105–1113.
16. Mann, R., Mulligan, R. and Baltimore, D. (1983) *Cell*, **33**, 153–159.
17. Miller, A. and Buttimore, C. (1986) *Mol. Cell Biol.*, **6**, 2895–2902.
18. Jones, T. A., Blaug, G., Hansen, M. and Barklis, E. (1990) *J. Virol.*, **64**, 2265–2279.
19. Parker, B. and Stark, G. (1979) *J. Virol.*, **31**, 360–369.
20. Cepko, C., Roberts, B. and Mulligan, R. (1984) *Cell*, **37**, 1053–1062.
21. Norton, P. and Coffin, J. (1985) *Mol. Cell Biol.*, **5**, 281–290.
22. Bradford, M.M. (1976) *Anal. Biochem.*, **72**, 248–254.
23. Hoeber, R., Migchielson, A., van der Jagt, R., van Ormondt, H. and van der Eb, A. (1991) *J. Virol.*, **65**, 904–912.
24. Walsh, C. and Cepko, C. (1988) *Science*, **241**, 1342–1345.
25. Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.*, **132**, 6–13.
26. Jaenisch, R., Harbers, K., Schnieke, A., Lohler, J., Chumakov, I., Jahner, D., Grotkopp, D. and Hoffman, E. (1983). *Cell*, **32**, 209–216.
27. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor University Press, Cold Spring Harbor.
28. Peckham, I., Sobel, S., Comer, J., Jaenisch, R. and Barklis, E. (1989) *Genes and Development*, **3**, 2062–2071.
29. Southern, E. (1975) *J. Mol. Biol.*, **503**–517.
30. Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. (1985) *Science*, **230**, 1350–1354.
31. Nolan, G.P., Fiering, S., Nicolas, J-F. and Herzenberg, L.A. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2603–2607.
32. Antequera, F., Boyes, J. and Bird, A. (1990) *Cell*, **62**, 503–514.
33. Hasty, P., Ramirez-Solis, R., Krumlauf, R. and Bradley, A. (1991) *Nature*, **350**, 243–246.
34. Valancius, V. and Smithies, O. (1991) *Mol. Cell Biol.*, **11**, 1402–1408.