JOURNAL OF VIROLOGY, June 2002, p. 5540–5547 Vol. 76, No. 11
0022-538X/02/$04.00+0 DOI: 10.1128/JVI.76.11.5540–5547.2002

Isolation and Analysis of Retroviral Integration Targets by Solo Long Terminal Repeat Inverse PCR

Yi Feng Jin,[1] Toshio Ishibashi,[2] Akio Nomoto,[1] and Michiaki Masuda[1,3]*

Department of Microbiology, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033,[1] Department of Otolaryngology, Social Insurance Central General Hospital, Tokyo 169-0073,[2] and Department of Microbiology, School of Medicine, Dokkyo University, Tochigi 321-0293,[3] Japan

Received 5 July 2001/Accepted 4 February 2002

Upon retroviral infection, the genomic RNA is reverse transcribed to make proviral DNA, which is then integrated into the host chromosome. Although the viral elements required for successful integration have been extensively characterized, little is known about the host DNA structure constituting preferred targets for proviral integration. In order to elucidate the mechanism for the target selection, comparison of host DNA sequences at proviral integration sites may be useful. To achieve simultaneous analysis of the upstream and downstream host DNA sequences flanking each proviral integration site, a Moloney murine leukemia virus-based retroviral vector was designed so that its integrated provirus could be removed by Cre-*loxP* homologous recombination, leaving a solo long terminal repeat (LTR). Taking advantage of the solo LTR, inverse PCR was carried out to amplify both the upstream and downstream cellular flanking DNA. The method called solo LTR inverse PCR, or SLIP, proved useful for simultaneously cloning the upstream and downstream flanking sequences of individual proviral integration sites from the polyclonal population of cells harboring provirus at different chromosomal sites. By the SLIP method, nucleotide sequences corresponding to 38 independent proviral integration targets were determined and, interestingly, atypical virus-host DNA junction structures were found in more than 20% of the cases. Characterization of retroviral integration sites using the SLIP method may provide useful insights into the mechanism for proviral integration and its target selection.

The RNA genome of retroviruses is reverse transcribed into a double-stranded DNA copy, which is then integrated into the host chromosome as a provirus. Viral elements, such as integrase (IN) and the terminal structures of viral DNA, that are required for retroviral integration have been extensively characterized. It has also been shown that selection of the proviral integration targets is nonrandom (15, 29–31, 37, 47) and that the central domain of IN plays a role in determining the target specificity (3, 36). The efficiency of chromosomal sites to become a preferred integration target appears to be affected by several factors, such as transcriptional activity (35, 46), DNase I hypersensitivity (9, 32, 33, 45), methylation (15), GC content (5, 16, 34), nuclear scaffold attachment (20), nucleosome structure (27, 28, 30, 31), and DNA structure of a higher order (14, 22, 24, 27, 28). However, these results were obtained by in vitro studies using artificial target DNA or analysis of a small number of in vivo integration sites, and it is still unclear why proviral integration takes place at certain target sites more often than others. To elucidate the mechanism of target selection for proviral integration, it may be useful to compile and analyze a large number of nucleotide sequences corresponding to proviral integration sites in the context of actual cellular chromosome. Traditionally, cellular flanking DNA of a provirus was cloned by the time-consuming method of genomic library construction and screening. Although the invention of PCR technology has led to rapid and simple methods, such as inverse PCR (40) and vectorette PCR (2), those techniques generally amplify either an upstream or downstream flanking sequence at one time, but not both. Therefore, when polyclonal populations of the cells harboring a provirus at different chromosomal sites are analyzed, it is difficult to tell which upstream sequence and downstream sequence are derived from the same integration site. To avoid this problem, we constructed a Moloney murine leukemia virus (Mo-MuLV)-based retroviral vector which carries the *loxP* sequence in each of the 5′ and 3′ long terminal repeats (LTRs). A previous study by Masuda et al. (19) indicated that proviral DNA of the *loxP*-carrying vector could be excised by Cre-mediated homologous recombination, leaving a solo LTR. It was also shown that inverse PCR taking advantage of the solo LTR allowed simultaneous cloning of the upstream and downstream flanking sequences in a single plasmid (19). In this study, we examined whether this method, termed solo LTR inverse PCR (SLIP), could be used to analyze polyclonal cell populations for proviral integration sites. The SLIP method was performed on five independent polyclonal populations of the vector-transduced cells representing a total of 151 integration events, and host DNA sequences corresponding to 38 proviral integration sites were successfully determined. The unique ability of the SLIP method to simultaneously characterize the upstream and downstream cellular flanking sequences revealed that the virus-host DNA junctions of more than 20% of the examined integration sites had aberrant structures which differed from the canonical 4-bp duplication expected for Mo-MuLV integration. The results suggested that the SLIP method is useful for characterizing proviral integration targets, whose analysis may provide novel insights into the mechanism of retroviral integration and its target selection.

(Portions of this study were performed by Yi Feng Jin in

* Corresponding author. Mailing address: Department of Microbiology, Dokkyo University School of Medicine, Tochigi 321-0293, Japan. Phone: 81-282-87-2131. Fax: 81-282-86-5616. E-mail: m-masuda@dokkyomed.ac.jp.

partial fulfillment of the requirements for a Ph.D. degree at the Graduate School of Medicine, the University of Tokyo.)

MATERIALS AND METHODS

Vector plasmid. The Mo-MuLV-based vector, TSN-lox, was described previously (19). Briefly, it carries the herpes simplex virus type 1 thymidine kinase gene (*tk*), the simian virus 40 replication origin, and the neomycin resistance gene (*neo*). It also has the *loxP* sequence in the R region of both the 5′ and 3′ LTRs. The LTL-lox vector used in this study (see Fig. 1) was constructed by removing the simian virus 40 origin and *neo* from TSN-lox. For this purpose, the TSN-lox vector plasmid was digested by *Bam*HI, *Cla*I, and *Xho*I, and the *Bam*HI-*Xho*I, *Cla*I-*Bam*HI, and *Xho*I-*Cla*I fragments containing the 5′ and 3′ LTRs and the *tk* gene, respectively, were ligated.

Cell culture. PT67 packaging cells expressing the *gag* and *pol* genes of Mo-MuLV and the *env* gene of 10A1 MuLV (21) were purchased from Clontech and grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum. The tk^- Rat2 cells (41) were also grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum. For selection of tk^+ and tk^- cells, cells were grown in the presence of hypoxanthine-aminopterin-thymidine (HAT) supplement (Life Technologies, Inc.) and 100 μM bromovinyldeoxyuridine (BVdU), respectively, as described previously (19).

Virus. LTL-lox vector virus was prepared by transfecting PT67 cells with the vector plasmid by the calcium phosphate precipitation method (44) using a CellPhect transfection kit (Pharmacia). Two days later, culture supernatants were harvested, filtered, and stored at −80°C. A Cre-expressing adenoviral vector, Adex1CAN-Cre (12, 13), was provided by Izumu Saito (Institute of Medical Science, The University of Tokyo) and propagated as described previously (26). For Adex1CAN-Cre infection, cells were added to the virus-containing fluid and incubated for 1 h at 37°C in a CO_2 incubator for adsorption. Then the virus fluid was removed and replaced with fresh culture medium.

DNA analysis. Chromosomal DNA (0.5 to 2 μg) was digested with *Tsp*509I, treated with T4 DNA ligase, and ethanol precipitated. The dried DNA pellet was used as a template for inverse PCR as described previously (19). All of the reagents for PCR were obtained from PE Biosystems. For the first-round reaction, a 50-μl reaction mixture was prepared by adding 38.5 μl of sterile distilled water, 5 μl of 10× PCR buffer, 1 μl each of dATP, dCTP, dGTP, and dTTP (final concentration, 1 mM each), 1 μl each of LTR-specific oligonucleotide primers (5′-ACTTGTGGTCTCGCTGTTCCTTGGG-3′ and 5′-ATCTGTTCCTGACC TTGATCTGAA C-3′; final concentration, 33 pM each), and 0.5 μl of AmpliTaq DNA polymerase (5 U/μl). Then, 25 cycles of PCR were carried out, with each cycle consisting of 94°C for 1 min, 55°C for 1 min, and 72°C for 2 min. Using 2 μl of the first-round PCR product as a template, 30 cycles of second-round PCR were carried with another set of primers (5′-GTCTCCTCTGAGTGATTGA C-3′ and 5′-GTTACTTAAGCTAGCTTGCC-3′). Purified PCR products were cloned in pT7Blue (Novagen) using a Perfectly Blunt Cloning Kit (Novagen) and transformed into NovaBlue *Escherichia coli* (Novagen), which allowed the blue-white screening by α-complementation (18). The nucleotide sequence of the cloned DNA was determined with an ABI Prism cycle sequencing kit (PE Biosystems) and a genetic analyzer (model 310; PE Biosystems). For PCR amplification of uninfected Rat2 cell DNA corresponding to proviral integration target sites, the following sets of oligonucleotides were used: 5′-CTTGCAACG CTAAGGTCGTT-3′ and 5′-TTCCTTACAAAGGGGCTTCA-3′ for the integration target in clone 1-32; 5′-GCCAGCCTGGTCTACATAGTG-3′ and 5′-A CAGAAAACGGTTGGAGGTG-3′ for clone 2-21; 5′-ATTGTGAGCAATGG TGAGCA-3′ and 5′-TTGTTAACTTTTCTTG-3′ for clone 3-5; 5′-TTCTATCA GTTTGCCTATAG-3′ and 5′-CATCGAGAGGTAAAATACTC-3′ for clone 3-19; 5′-GTGGCCACCTCGTGTAGTTT-3′ and 5′-CCCAACAGACCTAATG AAAGAA-3′ for clone 4-5; 5′-CAGTATGGCTGGAGACACGG-3′ and 5′-G CTTCCTTCTTGTGTCGCTT-3′ for clone 5-6; and 5′-AGAGAGGGTGGCT GAG-3′ and 5′-AAACTGGTCTCCGAATCCTG-3′ for clone 5-H1-6. Then, amplified fragments were cloned and sequenced as described above.

Sequence analysis. Database analysis of the obtained sequences was performed by using the BLAST homology search program (1).

RESULTS

Construction of proviral integration site libraries by SLIP. A general procedure for the SLIP method is depicted in Fig. 1. To construct a proviral integration site library, tk^- Rat2 cells seeded at a density of 10^5 cells per well in a six-well plate were infected with LTL-lox vector at a low multiplicity of infection so that multiple integration events in each cell could be avoided. Since LTL-lox carries the herpes simplex virus type 1 *tk* gene, the cells successfully transduced with the vector were obtained by 2 weeks of HAT selection. Five independent transduction experiments generated 23, 30, 33, 20, and 45 HAT-resistant colonies, and the colonies in each well were trypsinized and collected to make five respective pooled cultures designated as pools 1 through 5 (Table 1). Without further passage, cells of each pool were seeded at a density of 5×10^5 cells per 60-mm dish and the next day they were infected with adenoviral vector Adex1CAN-Cre (multiplicity of infection, 10), which expresses Cre recombinase (12, 13). A portion of the cells from pool 5 were passaged once or five times every 3 days under HAT selection and then infected with

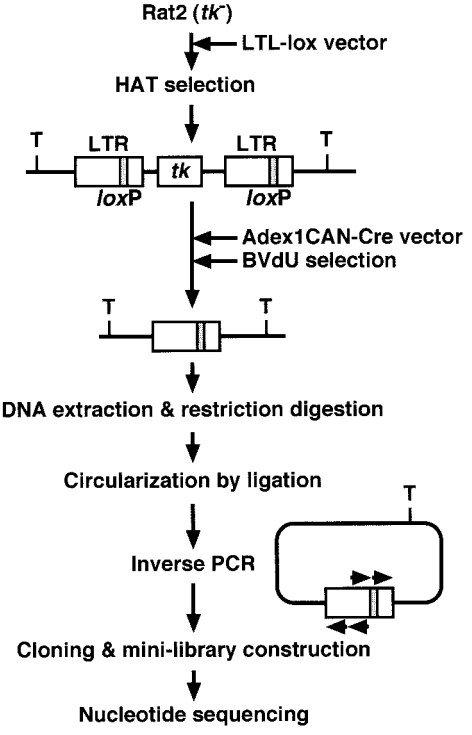

FIG. 1. Schematic representation of the protocol for SLIP. The *loxP* sequence in the LTR and *Tsp*509I cleavage sites (T) in the cellular flanking DNA are indicated. Arrowheads correspond to PCR primers. Relative sizes of the elements in the figure are arbitrarily determined and do not represent the actual ratio.

TABLE 1. Summary of SLIP analysis of LTL-lox-transduced Rat2 cell populations

Expt	No. of clones from LTL-lox-transduced cell population					
	1	2	3	4	5[a]	Total
Original HAT^r cell clones	23	33	30	20	45	151
Sequenced molecular clones	30	38	44	30	92	234
Artifact clones[b]	3	16	6	2	24	51
Integration target clones	27	22	38	28	68	183
Obtained integration targets	6	8	8	3	13	38

[a] The results include the data for pools 5, 5-H1, 5-H5, 5-B1, and 5-B5.
[b] Clones with an insert derived from PCR primers, vector plasmid, or endogenous retrovirus.

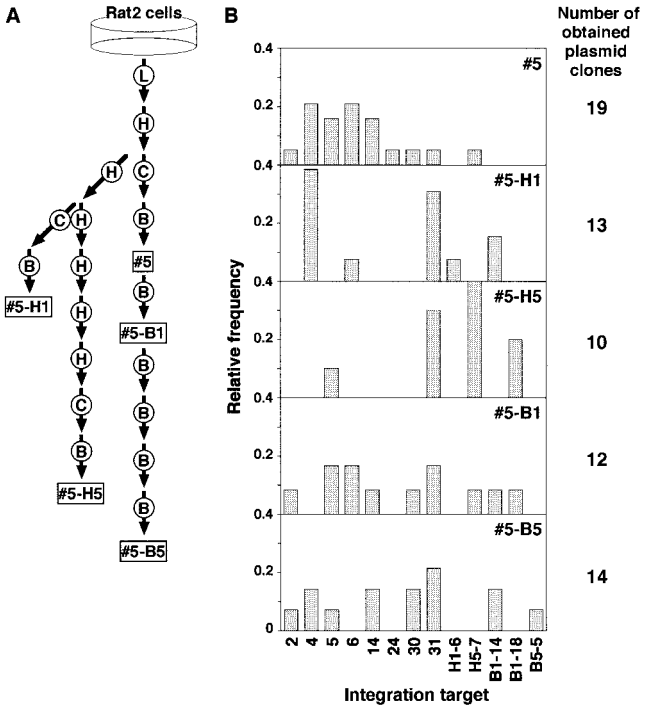

FIG. 2. Generation and analysis of sibling cultures derived from pool 5. (A) Flow chart showing the protocol for generating pools 5, 5-H1, 5-H5, 5-B1, and 5-B5. L, H, C, and B in the circle represent LTL-lox transduction, HAT selection, Cre expression, and BVdU selection, respectively. (B) Relative frequencies of various integration targets obtained from different pools. Bars indicate the ratio of the number of sister plasmid clones representing each integration target to the total number of the obtained clones shown on the right.

Adex1CAN-Cre to generate pools 5-H1 and 5-H5, respectively (Fig. 2A). As demonstrated previously (19), it was expected that Cre-*loxP* DNA recombination would excise the *tk*-bearing vector proviral DNA (Fig. 1). Therefore, BVdU (100 μM), which is toxic to *tk*-expressing cells, was added to the culture medium 48 h after Adex1CAN-Cre infection. The BVdU-resistant cells in each dish were harvested, and their chromosomal DNA was extracted. A part of the cells derived from pool 5 were passaged once or five times every 3 days in the presence of BVdU to generate pools 5-B1 and 5-B5, respectively, and then their DNA was extracted (Fig. 2A). The cellular DNA (0.5 to 2 μg) was digested with *Tsp*509I, a 4-base cutter whose recognition site does not exist in the LTR, and treated with T4 DNA ligase to circularize restriction fragments (Fig. 1). Then, nested inverse PCR with two sets of LTR-specific primers was carried out, and the products were cloned in pT7Blue and transformed into *E. coli* (NovaBlue) under conditions that several hundred colonies were formed on a 100-mm agar plate (Fig. 1). The α-complementation indicated that more than 90% of the colonies had a plasmid with an insert fragment. As a control, DNA extracted from untransduced Rat2 cells was also subjected to the same procedure.

Analysis of the integration site libraries. Of 479 white colonies on the agar plate of the integration site library for pool 1, 30 white colonies were randomly selected, and the plasmids extracted from these clones were used for nucleotide sequencing (Table 1). Similarly, 38, 44, and 30 clones derived from pools 2, 3, and 4 were sequenced. As for pool 5 and its derivatives (Fig. 2A), a total of 92 clones were sequenced (Table 1). Three clones of pool 1 had a primer- or vector-derived sequence as the insert and were most likely generated by an artifact. One clone of pool 2 had a sequence homologous to N-tropic ecotropic mouse endogenous retrovirus envelope gene (25) (Table 1). The same endogenous viral sequence was obtained when a SLIP library of uninfected Rat2 cells was analyzed. It was most likely that the PCR primers could cross-hybridize with the LTR of the endogenous retrovirus. The clones that carried an insert generated by those artifacts were discarded, and the remaining clones were defined as representatives of proper viral integration events. Some of the clones satisfying the criteria contained the same sequence, indicating that they were sister clones corresponding to same integration target. Thus, 27 clones derived from pool 1 were finally categorized to six integration sites (Table 1 and Fig. 3). Similarly, clones derived from pool 2 through 5 defined 8, 8, 3, and 13 integration sites, respectively (Table 1 and Fig. 3). Collectively, 38 integration sites of 151 integration events were identified (Table 1).

Comparison of the results obtained by SLIP analysis of pool 5 and its derivative pools suggested that prolonged HAT selection, but not BVdU selection, decreased the clonal diversity of the culture (Fig. 2B).

Aberrant sequence duplication was found at the virus-host DNA junction of more than 20% of the integration sites. Figure 3 shows the nucleotide sequences of the 38 integration sites. A major advantage of the SLIP method was that both the upstream and downstream cellular flanking sequences were determined simultaneously for individual integration sites of the cells in polyclonal population. It is generally believed that integration of Mo-MuLV generates a 4-bp duplication of the target DNA sequence at the junction of the provirus and flanking regions (38, 42). Consistently, 30 of 38 integration sites examined in this study had a 4-bp duplication (Fig. 3). In contrast, the other eight integration sites, amounting to more than 20% of the examined sites, had aberrant structures at the virus-host DNA junctions. Of these eight cases, six had a 5-bp duplication at the junction (Fig. 4A). To determine nucleotide sequences of the original integration targets, genomic DNA of uninfected Rat2 cells was amplified by PCR and sequenced, except for clone B5-5 of pool 5, whose target site was immediately flanked by a repetitive motif. The results indicated that the five duplicated nucleotides were derived from rat DNA in all cases examined. Clone 19 of pool 3 had 5′-AATCC-3′ and 5′-TTTCC-3′ at the upstream and downstream junctions, respectively (Fig. 4B). PCR amplification and nucleotide sequencing of this region in uninfected Rat2 chromosomes revealed that the original sequence at the integration target was 5′-AATCCAA-3′ (Fig. 4B). As for clone 6 of pool 5, four molecular clones were obtained. Two of them had 5′-CTAG-3′ and 5′-TTAG-3′, and the other two had 5′-CTAA-3′ and 5′-CTAG-3′ at the upstream and downstream junctions, respectively (Fig. 4C). The original rat genomic sequence at this site was 5′-CTAG-3′ (Fig. 4C).

Other molecular characteristics found at integration targets. Homology search analysis was carried out on the isolated integration target sequences with the BLAST software (1). The results indicated that more than 45% of the integration targets

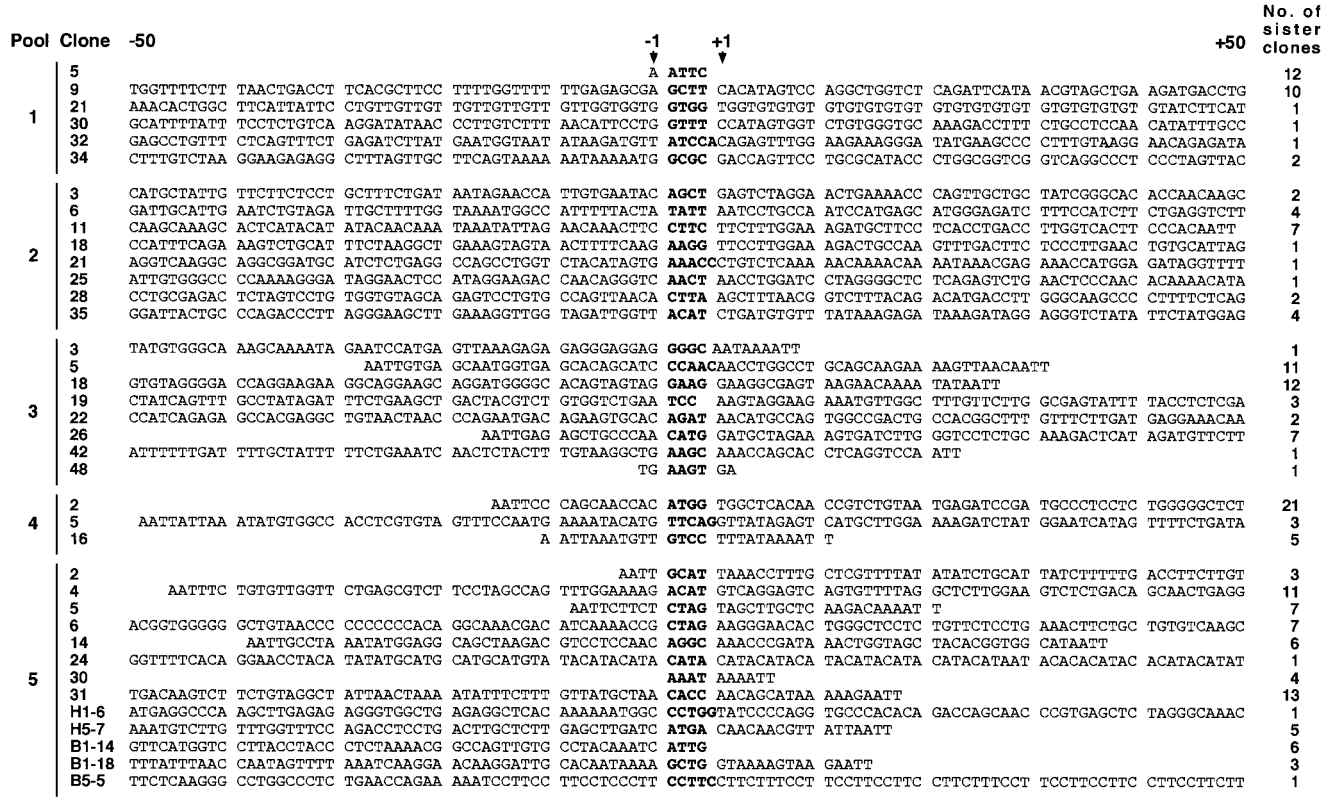

FIG. 3. Nucleotide sequences of 38 proviral integration target sequences determined in this study. Nucleotides in boldface in the center are duplicated at virus-host DNA junctions, and the deduced IN cleavage sites (arrows) are indicated. Positions of the upstream and downstream nucleotides adjacent to the duplicated residues are numbered −1 and +1, and nucleotide sequences of the region spanning from −50 to +50 are aligned. Pool and clone numbers are shown on the left, and the numbers of sister clones corresponding to each integration target are shown on the right.

were localized within or in the vicinity of repetitive sequences such as B family (Fig. 5), Alu-like sequence, and L1 (Table 2). Integration targets of clone 21 of pool 1 and clones 24 and B5-5 of pool 5 were found in the region of simple-patterned repetitive motifs such as $(G_{1\text{-}2}T_{1\text{-}2})_n$, $(CATA)_n$, and $(C_{1\text{-}3}T_{1\text{-}3})_n$, respectively (Fig. 3). The BLAST search did not unequivocally reveal homology between cloned integration targets and known functional genes.

The average GC content of the target regions shown in Fig. 3 was 42.6% and was comparable to the value for the rat genome (41.8%). However, the frequency of each nucleotide varied from position to position, and the schematic box plot analysis indicated that the AT frequencies at positions −2 and +2 and the GC frequency at position +27 were higher than those at other positions (Fig. 6).

DISCUSSION

Previous PCR-based methods, such as inverse PCR and vectorette PCR, are usually used to amplify either the upstream or downstream cellular flanking sequence at proviral integration sites, but not both at the same time. Therefore, when a polyclonal population of cells harboring provirus at different chromosomal positions is analyzed, it is difficult to determine which upstream and downstream sequences are derived from the same integration site. In contrast, the SLIP method enables simultaneous isolation of upstream and downstream cellular flanking sequences of individual proviral integration sites and it was used in the previous study to analyze monoclonal cell populations for their proviral integration targets (19). In this study, it was demonstrated that SLIP is also useful for analyzing polyclonal cell populations.

By analysis of five cell populations representing a total of 151 integration events of the Mo-MuLV-based LTL-lox vector, nucleotide sequences of 38 independent integration targets were determined. Due to the unique ability of the SLIP method to amplify the upstream and downstream flanking sequences simultaneously, we were able to directly compare the 5′ and 3′ virus-host DNA junctions of these integration sites. As shown in Fig. 7A, it is thought that Mo-MuLV generates duplication of 4-bp host sequences at the virus-host DNA junctions (38, 42). Thirty of 38 integration sites examined in this study were compatible with this canonical structure. Although no strong consensus sequence was found among those 4-bp duplications, the results appeared to be consistent with the previous study showing that the middle two positions of the 4-bp direct repeat are preferentially occupied by AA, TT, or AT dinucleotides (31). Interestingly, as many as eight integration sites, amounting to more than 20% of the obtained clones, had atypical virus-host DNA junction structures. Of these aberrant clones, six had 5-bp duplication at the junctions. Anal-

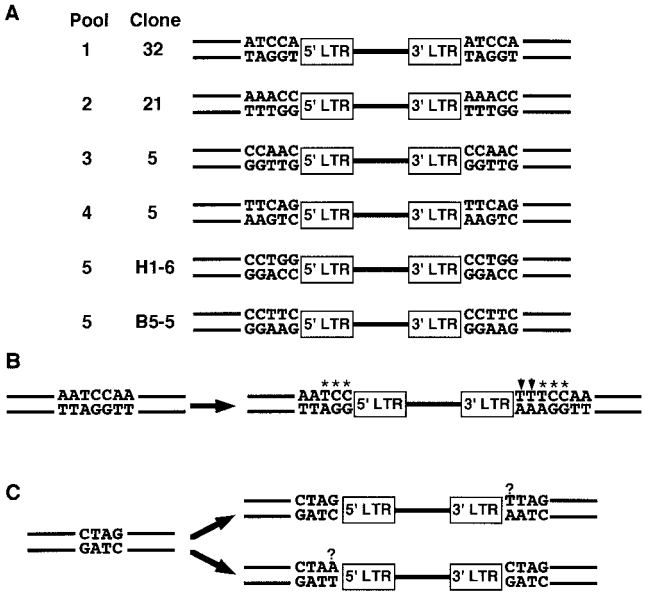

FIG. 4. Aberrant sequence duplications found at virus-host DNA junctions. (A) The 5-bp duplications found in six integration sites. Proviral DNA delineated by the 5′ and 3′ LTRs is depicted as a thick line, and parallel thin lines represent the cellular flanking DNA. Pool and clone numbers are shown on the left. (B) Aberrant duplication in clone 19 of pool 3. Insertion of proviral terminal dinucleotides (arrowheads) and the 3-bp duplication (asterisks) of the cellular target sequence may be responsible. (C) Aberrant duplication in clone 6 of pool 5. Two sets of clones bearing different nucleotide sequences were obtained. Nucleotides of unknown origin are indicated by question marks.

ysis of original rat DNA sequences corresponding to these integration targets suggested that the 5-bp duplications had been generated by introduction of staggered cuts in the host DNA at positions 5 bp apart from each other (Fig. 7B). Although a 4-bp duplication associated with incomplete removal of a viral terminal nucleotide (Fig. 7C and D) (11, 17, 19, 39) could explain 2 of the 5 bp duplications (clones 1-32 and clone 4-5), that does not appear to be the case. Clone 3-19 had 5′-AATCC-3′ and 5′-TTTCC-3′ at the upstream and downstream junctions, respectively. Since the original sequence of

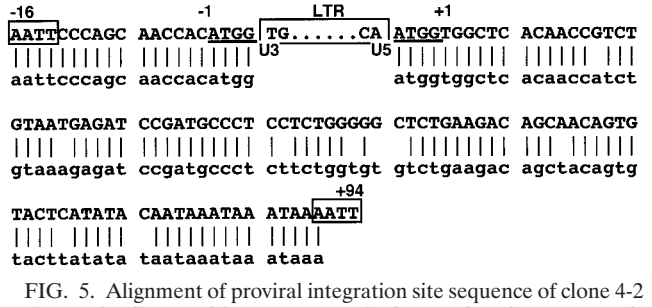

FIG. 5. Alignment of proviral integration site sequence of clone 4-2 and rat B family repetitive sequence. The integration site sequence is shown in uppercase letters with the integrated solo LTR, and nucleotide positions are numbered as described in the legend for Fig. 2. Duplicated nucleotides at virus-host DNA junctions are underlined, and *Tsp*509I recognition sites are boxed. The rat B family sequence is shown in lowercase letters, and nucleotide identity is indicated by vertical bars.

TABLE 2. Classification of the proviral integration site sequences

Sequence class	No. of sequences in class	Proportion (%)
Repetitive		
SINE[a]	8[c]	21.0
LINE/L1	3	7.9
Satellite DNA	2	5.3
Simple repeat	3[d]	7.9
Other[b]	2	5.3
Anonymous	20[e]	52.6

[a] SINE includes Alu-like sequences, B family repetitive sequences, and mammalian-wide interspersed repeat.
[b] These include an endogenous LTR sequence and ORR1A2.
[c] Aberrant clone 2-21 is included.
[d] Aberrant clone 5-B5-5 is included.
[e] The other six integration targets with aberrant junction structures are included.

this site in the rat genome was determined to be 5′-AATCCA A-3′, it is possible that this aberrant junction structure was generated by a 3-bp duplication (5′-TCC-3′) combined with failure of removal of the 3′ terminal dinucleotide of the unintegrated proviral DNA (Fig. 7E). As for clone 6 of pool 5, four molecular clones with different sequences were obtained. Two of them had 5′-CTAG-3′ and 5′-TTAG-3′, and the other two had 5′-CTAA-3′ and 5′-CTAG-3′ at the upstream and downstream junctions, respectively. The original rat genomic sequence of this site was 5′-CTAG-3′. It is possible that a base-pair mismatch caused by incomplete removal of the terminal nucleotide was retained at each end of the integrated provirus and that DNA replication of the region took place before the mismatches were repaired (Fig. 7F). Following cell division, two types of daughter cells, carrying different junction structures may have been generated. Previous studies on AKR MuLV and avian leukemia and sarcoma virus-based vectors have described unexpected virus-host DNA junction structures at the integration target revealing microhomologies with viral LTR ends (23, 43). However, similar findings were not obtained for eight aberrant cases found in this study. It has previously been shown that noncanonical virus-host DNA junction structures could be generated for Mo-MuLV with mutations at the LTR terminus and AKR MuLV (6, 7, 43). It has also been shown that mutation in conserved amino acids in the catalytic domain of the human immunodeficiency virus type 1 IN impaired IN-mediated proviral integration, causing aberrant virus-host DNA junction structures (8). In contrast, the vector virus used in this study had a Mo-MuLV LTR with the intact terminal structure and was produced by PT67 packaging cells, which express wild-type Mo-MuLV IN (21). Therefore, it is possible that the normal process of Mo-MuLV proviral integration is intrinsically more error-prone than generally thought.

Nucleotide sequence analysis showed that as many as 47% of the integration targets were found within or near a repetitive sequence. Since it is estimated that about 40% of a mammalian genome consists of repetitive sequences, the results may not necessarily indicate that repetitive sequences are preferred targets for proviral integration. A previous study on human immunodeficiency virus type 1 also revealed no strong biases either for or against integration near repetitive sequences in

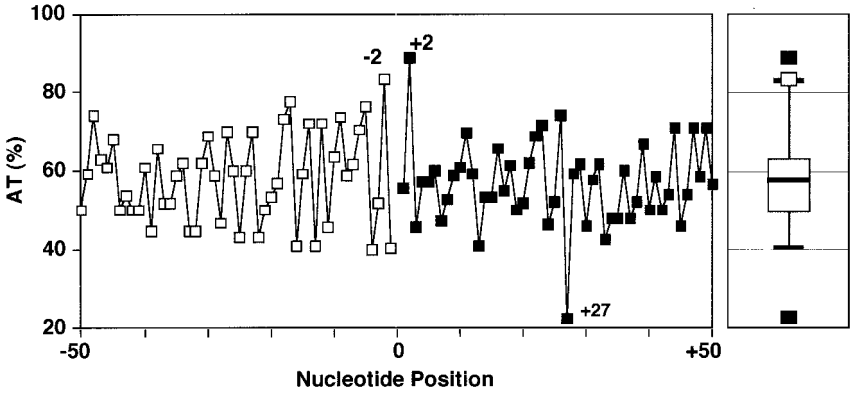

FIG. 6. Distribution of AT nucleotides at each of the first 50 positions of the upstream (open squares) and downstream (solid squares) flanking sequences shown in Fig. 2. The right panel shows the schematic box plot analysis of the AT content along the flanking sequences. The box represents the difference between the 75th percentile and the 25th percentile of the AT distribution (i.e., the interquartile range [IQR]). A thick horizontal line within the box represents the median. The whiskers above and below the box extend to the upper and lower inner fence values, respectively. The open box indicates that the percent AT at position −2 corresponds to the upper inner fence value [the largest value that does not exceed the median + (1.5 × IQR)], while the solid boxes indicate that the percentages of AT at positions +2 and +27 are high and low outliers, respectively.

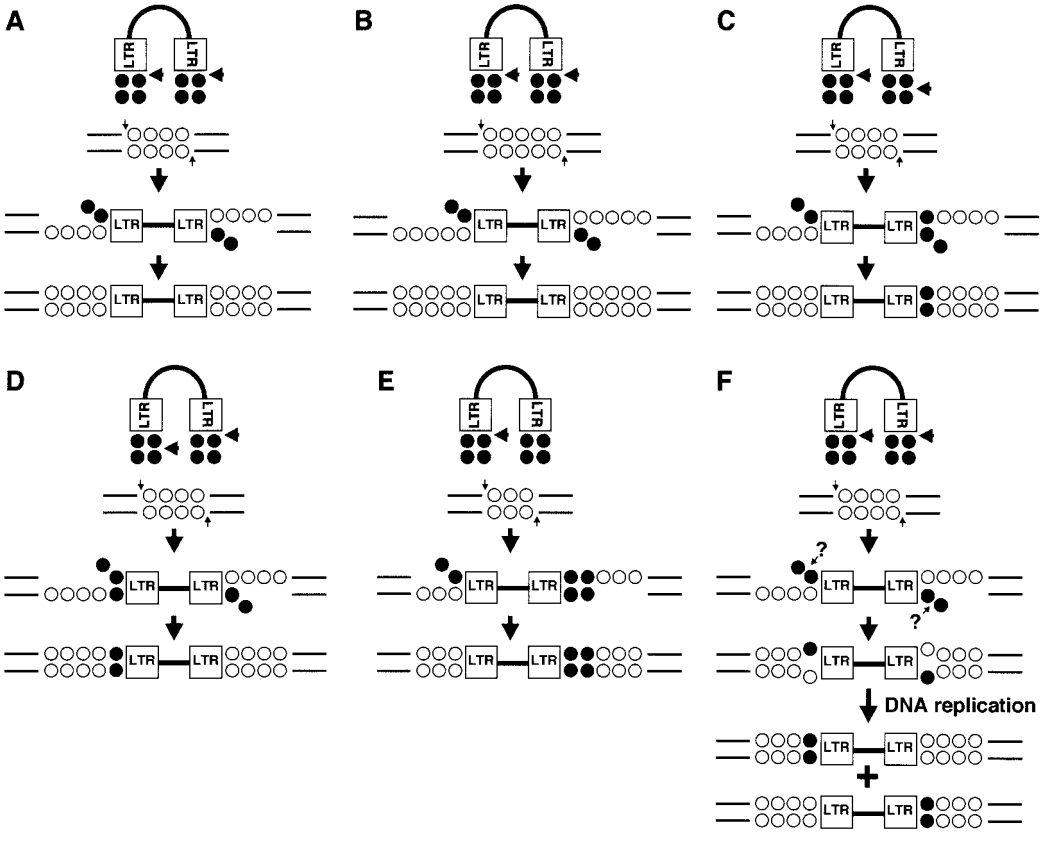

FIG. 7. Possible DNA processing mechanisms involved in Mo-MuLV proviral integration. Unintegrated linear viral DNA is depicted as a bent thick line with the 5′and 3′ LTRs. The target cellular DNA is shown as parallel thin lines. Open circles and solid circles correspond to cellular and viral nucleotides, respectively. Larger and smaller arrowheads indicate IN cleavage sites in viral and cellular DNA, respectively. (A) The IN-mediated normal mechanism removes terminal dinucleotides from each end of viral DNA, introduces a staggered cut at positions four bases away from each other, and generates a 4-bp duplication of cellular sequence at virus-host DNA junctions. (B) Introduction of a staggered cut in target DNA at positions five bases away from each other generates a 5-bp duplication of the cellular sequence. This mechanism appears to be involved in clones 1-32, 2-21, 3-5, 4-5, 5-H1-6, and possibly 5-B5-5. (C) Incomplete removal of the viral 3′ terminal nucleotide causes insertion of an additional residue at the downstream junction. (D) Incomplete removal of the viral 5′ terminal nucleotide causes insertion of an additional residue at the upstream junction. (E) Introduction of a staggered cut in target DNA at positions three bases away from each other without removal of the viral 3′ terminal dinucleotides could generate unusual junction structures, such as the one in clone 3-19. (F) Second strand joining before complete removal of viral terminal nucleotide could cause mismatched base pairs adjacent to the LTR. Arrowheads with a question mark indicate putative incorrect cleavage sites in the viral terminal nucleotides. DNA replication of the region before repair of the mismatches may result in two sets of molecular clones with different sequences (e.g., clone 5-6).

the human genome (4). In this study, several integration targets, such as clone 21 of pool 1 and clones 24 and B5-5 of pool 5, consisted of repetition of simple motifs, such as $(G_{1\text{-}2}T_{1\text{-}2})_n$, $(CATA)_n$, and $(C_{1\text{-}3}T_{1\text{-}3})_n$, respectively. Further studies are necessary to examine whether repetition of these simple motifs plays a role in constituting a preferred structure for proviral integration. Analysis of the integration target sequences also showed that the AT frequencies at positions −2 and +2 were higher than those at other positions. The results appear to be compatible with the previous data on human T-cell leukemia virus type 1 integration targets (16). The GC frequency at position +27 was also high. Further studies are necessary to examine its biological significance.

The SLIP method used in the present study depended on selection for transgene (*tk*) expression. Therefore, the collected integration sites may represent not only preferred integration targets but also favorable contexts for proviral expression. In addition, prolonged HAT selection appeared to decrease clonal diversity of the pooled culture, possibly causing different levels of effects on cell proliferation of individual clones. It should also be mentioned that the fraction of successfully cloned sites was relatively small (38 out of 151 integration events) and that some sites were cloned multiple times. Several factors, such as the biases for and against growth of certain cell clones and different efficiencies in PCR amplification of various sites, may have been responsible. To solve these problems, efforts are being made to improve the protocol of the SLIP method. For example, a DNA clone for Mo-MuLV bearing *loxP* in the R region of the 5′ and 3′ LTRs was constructed, and our preliminary data indicate that the recombinant virus could be propagated without losing *loxP* (data not shown). Using this virus, it may be possible to carry out the SLIP method without having to depend on proviral expression.

Retroviral vectors are thought to be useful tools for gene therapy. However, additional studies on the target specificity of proviral integration would be necessary for addressing the safety issue of retrovirus-mediated gene therapy (10). Characterization of integration targets by using the SLIP method may be effective for this purpose and may provide useful insights into the mechanism for retroviral integration and its target selection.

ACKNOWLEDGMENTS

We thank Izumu Saito for the Adex1 CAN-Cre vector, Hiroko Igarashi for technical assistance, and Etsuko Suzuki and Yuko Matsushita (Graduate School of Medicine, The University of Tokyo) as well as Saki Yachuuda and Yuki Shinozaki (Dokkyo University, School of Medicine) for secretarial work. We also thank Takao Masuda (Tokyo Medical and Dental University) for useful discussion.

REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
2. **Arnold, C., and I. J. Hodgson.** 1991. Vectorette PCR: a novel approach to genomic walking. PCR Methods Appl. **1:**39–42.
3. **Berger, N., A. E. Heller, K. D. Stormann, and E. Pfaff.** 2001. Characterization of chimeric enzymes between caprine arthritis-encephalitis virus, maedivisna virus and human immunodeficiency virus type 1 integrases expressed in *Escherichia coli*. J. Gen. Virol. **82:**139–148.
4. **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.
5. **Chou, K. S., A. Okayama, I. J. Su, T. H. Lee, and M. Essex.** 1996. Preferred nucleotide sequence at the integration target site of human T-cell leukemia virus type I from patients with adult T-cell leukemia. Int. J. Cancer **65:**20–24.
6. **Colicelli, J., and S. P. Goff.** 1988. Isolation of an integrated provirus of Moloney murine leukemia virus with long terminal repeats in inverted orientation: integration utilizing two U3 sequences. J. Virol. **62:**633–636.
7. **Colicelli, J., and S. P. Goff.** 1985. Mutants and pseudorevertants of Moloney murine leukemia virus with alterations at the integration site. Cell **42:**573–580.
8. **Gaur, M., and A. D. Leavitt.** 1998. Mutations in the human immunodeficiency virus type 1 integrase D,D(35)E motif do not eliminate provirus formation. J. Virol. **72:**4678–4685.
9. **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 leukosis viruses. J. Virol. **61:**2489–2498.
10. **Holmes-Son, M. L., R. S. Appa, and S. A. Chow.** 2001. Molecular genetics and target site specificity of retroviral integration. Adv. Genet. **43:**33–69.
11. **Ju, G., and A. M. Skalka.** 1980. Nucleotide sequence analysis of the long terminal repeat (LTR) of avian retroviruses: structural similarities with transposable elements. Cell **22:**379–386.
12. **Kanegae, Y., G. Lee, Y. Sato, M. Tanaka, M. Nakai, T. Sakaki, S. Sugano, and I. Saito.** 1995. Efficient gene activation in mammalian cells by using recombinant adenovirus expressing site-specific Cre recombinase. Nucleic Acids Res. **23:**3816–3821.
13. **Kanegae, Y., K. Takamori, Y. Sato, G. Lee, M. Nakai, and I. Saito.** 1996. Efficient gene activation system on mammalian cell chromosomes using recombinant adenovirus producing Cre recombinase. Gene **181:**207–212.
14. **Katz, R. A., K. Gravuer, and A. M. Skalka.** 1998. A preferred target DNA structure for retroviral integrase in vitro. J. Biol. Chem. **273:**24190–24195.
15. **Kitamura, Y., Y. M. 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.
16. **Leclercq, I., F. Mortreux, M. Cavrois, A. Leroy, A. Gessain, S. Wain-Hobson, and E. Wattel.** 2000. Host sequences flanking the human T-cell leukemia virus type 1 provirus in vivo. J. Virol. **74:**2305–2312.
17. **Majors, J. E., and H. E. Varmus.** 1981. Nucleotide sequences at host-proviral junctions for mouse mammary tumour virus. Nature **289:**253–258.
18. **Maniatis, T., J. Sambrook, and E. F. Fritsch.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Masuda, M., H. Igarashi, M. Kano, and H. Yoshikura.** 1998. Effects of procollagen C-proteinase enhancer protein on the growth of cultured rat fibroblasts revealed by an excisable retroviral vector. Cell Growth Differ. **9:**381–391.
20. **Mielke, C., K. Maass, M. Tummler, and J. Bode.** 1996. Anatomy of highly expressing chromosomal sites targeted by retroviral vectors. Biochemistry **35:**2239–2252.
21. **Miller, A. D., and F. Chen.** 1996. Retrovirus packaging cells based on 10A1 murine leukemia virus for production of vectors that use multiple receptors for cell entry. J. Virol. **70:**5564–5571.
22. **Milot, E., A. Belmaaza, E. Rassart, and P. Chartrand.** 1994. Association of a host DNA structure with retroviral integration sites in chromosomal DNA. Virology **201:**408–412.
23. **Moreau, K., C. Torne-Celer, C. Faure, G. Verdier, and C. Ronfort.** 2000. In vivo retroviral integration: fidelity to size of the host DNA duplication might be reduced when integration occurs near sequences homologous to LTR ends. Virology **278:**133–136.
24. **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.
25. **Portis, J. L., S. Perryman, and F. J. McAtee.** 1991. The R-U5–5′ leader sequence of neurovirulent wild mouse retrovirus contains an element controlling the incubation period of neurodegenerative disease. J. Virol. **65:**1877–1883.
26. **Precious, B., and W. C. Russell.** 1985. Growth, purification, and titration of adenoviruses, p. 193–206. *In* B. W. J. Mahy (ed.) Virology: a practical approach. IRL Press, Oxford, United Kingdom.
27. **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.
28. **Pruss, D., R. Reeves, F. D. Bushman, and A. P. Wolffe.** 1994. The influence of DNA and nucleosome structure on integration events directed by HIV integrase. J. Biol. Chem. **269:**25031–25041.
29. **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.
30. **Pryciak, P. M., A. Sil, and H. E. Varmus.** 1992. Retroviral integration into minichromosomes in vitro. EMBO J. **11:**291–303.
31. **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.
32. **Robinson, H. L., and G. C. Gagnon.** 1986. Patterns of proviral insertion and deletion in avian leukosis virus-induced lymphomas. J. Virol. **57:**28–36.

33. **Rohdewohld, H., H. Weiher, W. Reik, R. Jaenisch, and M. Breindl.** 1987. Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. J. Virol. **61:**336–343.
34. **Rynditch, A. V., S. Zoubak, L. Tsyba, N. Tryapitsina-Guley, and G. Bernardi.** 1998. The regional integration of retroviral sequences into the mosaic genomes of mammals. Gene **222:**1–16.
35. **Scherdin, U., K. Rhodes, and M. Breindl.** 1990. Transcriptionally active genome regions are preferred targets for retrovirus integration. J. Virol. **64:**907–912.
36. **Shibagaki, Y., and S. A. Chow.** 1997. Central core domain of retroviral integrase is responsible for target site selection. J. Biol. Chem. **272:**8361–8369.
37. **Shih, C. C., J. P. Stoye, and J. M. Coffin.** 1988. Highly preferred targets for retrovirus integration. Cell **53:**531–537.
38. **Shoemaker, C., S. Goff, E. Gilboa, M. Paskind, S. W. Mitra, and D. Baltimore.** 1980. Structure of a cloned circular Moloney murine leukemia virus DNA molecule containing an inverted segment: implications for retrovirus integration. Proc. Natl. Acad. Sci. USA **77:**3932–3936.
39. **Shoemaker, C., J. Hoffman, S. P. Goff, and D. Baltimore.** 1981. Intramolecular integration within Moloney murine leukemia virus DNA. J. Virol. **40:**164–172.
40. **Silver, J., and V. Keerikatte.** 1989. Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. J. Virol. **63:**1924–1928.
41. **Topp, W. C.** 1981. Normal rat cell lines deficient in nuclear thymidine kinase. Virology **113:**408–411.
42. **van Beveren, C., J. G. Goddard, A. Berns, and I. M. Verma.** 1980. Structure of Moloney murine leukemia viral DNA: nucleotide sequence of the 5′ long terminal repeat and adjacent cellular sequences. Proc. Natl. Acad. Sci. USA **77:**3307–3311.
43. **van Beveren, C., E. Rands, S. K. Chattopadhyay, D. R. Lowy, and I. M. Verma.** 1982. Long terminal repeat of murine retroviral DNAs: sequence analysis, host-proviral junctions, and preintegration site. J. Virol. **41:**542–556.
44. **van der Eb, A. J., and F. L. Graham.** 1980. Assay of transforming activity of tumor virus DNA. Methods Enzymol. **65:**826–839.
45. **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.
46. **Weidhaas, J. B., E. L. Angelichio, S. Fenner, and J. M. Coffin.** 2000. Relationship between retroviral DNA integration and gene expression. J. Virol. **74:**8382–8389.
47. **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.