

# Effect of Distance between Homologous Sequences and 3' Homology on the Frequency of Retroviral Reverse Transcriptase Template Switching

KRISTA A. DELVIKS<sup>1,2</sup> AND VINAY K. PATHAK<sup>2,3\*</sup>

*Department of Genetics and Developmental Biology,<sup>1</sup> Mary Babb Randolph Cancer Center,<sup>2</sup> and Department of Biochemistry,<sup>3</sup> West Virginia University, Morgantown, West Virginia 26506*

Received 11 March 1999/Accepted 17 June 1999

**Deletion of direct repeats in retroviral genomes provides an in vivo system for analysis of reverse transcriptase (RT) template switching. The effect of distance between direct repeats on the rate of deletion was determined for 16 murine leukemia virus (MLV)-based vectors containing a 701-bp direct repeat of overlapping fragments of the herpes simplex virus thymidine kinase gene (HTK). The direct repeats were separated by spacer fragments of various lengths (0.1 to 3.5 kb). Southern analysis of infected cells after one replication cycle indicated that all vectors in which the distance between homologous sequences was >1,500 bp deleted at very high rates (>90%). In contrast, vectors containing <1,500 bp between homologous sequences exhibited lower frequencies of deletion (37 to 82%). To analyze the pattern of locations at which RT switched templates, restriction site markers were introduced to divide the downstream direct repeat into five regions. RT switched templates within all five regions of the 701-bp direct repeat and the frequency of template switching was greater within the 5' regions in comparison to the 3' regions. The probability of RT switching templates within the 5' regions doubled when the MLV packaging sequence ( $\Psi$ ) was placed between the 701-bp direct repeats. However,  $\Psi$  did not increase the rate of template switching for shorter direct repeats. These results indicate that linear distance between homologous sequences increases the rate of template switching and suggest that duplex formation between nascent DNA and homologous template sequences 3' of RT promote template switching.**

Retroviral reverse transcriptases (RTs) convert single-stranded retroviral RNA into double-stranded viral DNA (1, 3, 45). The process of reverse transcription involves two obligatory template-switching events, designated minus- and plus-strand DNA transfer, which require that RT dissociate from the template at one location and reassociate with a homologous sequence at another location (3). Because RT is required to dissociate from the template, it is postulated that RT evolved to possess low template affinity and processivity (43). The genetic consequences of RT's low template affinity are that additional template switching events frequently occur during the process of reverse transcription. Intermolecular template switching events between the two copackaged viral RNAs can lead to homologous and nonhomologous recombination (15, 23, 44), whereas intramolecular template switching events (within the same template RNA) can lead to mutations such as deletions, deletions with insertions, insertions, and duplications (33, 35). The low template affinity and low processivity of RT may also significantly contribute to the high rate of substitution and frameshift mutations during reverse transcription (34).

Several in vivo and in vitro studies have analyzed minus-strand and plus-strand transfer events in an effort to elucidate the mechanism of RT template switching (9, 10, 20, 37, 48, 53–55). Retroviral vectors containing directly repeated homologous sequences constitute a powerful in vivo model system to analyze RT template switching during one cycle of retroviral

replication (6, 22, 33). Directly repeated sequences in retroviral genomes are unstable and are frequently deleted from the integrated proviruses (2, 4, 6, 17, 23, 31, 33–35, 47, 49). Direct repeats composed of the neomycin phosphotransferase gene (*neo*) and the herpes simplex virus thymidine kinase gene (HTK) have been shown to delete accurately and at high frequencies during both spleen necrosis virus and murine leukemia virus (MLV) replication (6, 22). These and other in vivo studies have shown that the frequency of template switching is dependent on the length of homology (55). Using direct repeats as a model system, it was shown that template switching events leading to direct repeat deletions are primarily intramolecular (15). Additionally, it was shown that the frequency of RT template switching is equal during minus- and plus-strand DNA synthesis (2). Other in vivo studies have also shown that the dimer linkage structure constitutes a hot spot for intermolecular template switching events (27–29).

Several in vitro assays have been used to analyze the template switching properties of RT. These studies have shown that pausing of RT promoted by template secondary structure, depletion of nucleotide pools, or specific nucleotide sequences increases the rate of template switching (7, 11, 12, 26, 32, 40–42, 50). Specific pause sites present within retroviral genomes have also been identified and are implicated in promoting template switching (25, 51).

Our previous studies indicated that a 701-bp direct repeat composed of overlapping fragments of the HTK gene deleted at a high frequency (6). When the direct repeats were adjacent to one another, they deleted at a frequency of 57% during a single replication cycle. Interestingly, the direct repeats separated by an 818-bp spacer fragment composed of the MLV packaging signal ( $\Psi$ ), including the dimer linkage structure, deleted at a frequency of 91%. The increase in the frequency

\* Corresponding author. Mailing address: Mary Babb Randolph Cancer Center, West Virginia University, Morgantown, WV 26506. Phone: (304) 293-0495. Fax: (304) 293-4667. E-mail: VPATHAK@WVU.edu.

of template switching could be caused by either a specific effect of the MLV  $\Psi$  on the secondary and tertiary structure of the viral RNA, or an effect of the increased linear distance between the direct repeats. In this study, we demonstrate that increasing the linear distance between homologous sequences increases the frequency of RT template-switching events.

#### MATERIALS AND METHODS

**Definitions and plasmid construction.** The presence of a "p" preceding a retroviral vector name refers to the plasmid, whereas the absence of a "p" refers to the viruses or proviruses derived from these plasmids. All vectors were constructed by using standard cloning procedures (38). A detailed description of all cloning steps is available upon request. Briefly, MLV-based vectors pKD-HTnotTK and all vectors containing spacers of various lengths were derived from the previously described vector pKD-HTTK (6). Vector pKD-HTpTK was also previously described (6). Each vector contains *neo* from Tn5 expressed from the encephalomyocarditis virus internal ribosomal entry site (IRES) (18, 19, 21). A *NotI* linker was inserted between the HT and TK fragments of pKD-HTTK to create pKD-HTnotTK. Various restriction fragments from the murine Na<sup>+</sup>-K<sup>+</sup> ATPase gene that confers resistance to ouabain (*ouabain*) (24), the bacterial  $\beta$ -galactosidase gene (*lacZ*) (pSV $\beta$ ; Clontech), or the hygromycin phosphotransferase B gene (*hygro*) (13) were isolated and inserted into the *NotI* site of pKD-HTnotTK. The numbers within the vector names refer to the length of the spacer fragment in base pairs between the HT and TK fragments.

PCR amplification of the TK portions of HTK were used to create pKD-HTp1TK, pKD-HTp2TK, pKD-HTp4TK, pKD-HT1TK, pKD-HT2TK, and pKD-HT4TK. The structures of the vectors were verified by DNA sequencing analysis to confirm that no new mutations were introduced (data not shown).

To create pKD-HTT4\*K and pKD-HTpT4\*K, site-directed mutagenesis (Chameleon double-stranded mutagenesis kit; Stratagene) was performed to introduce four unique restriction enzyme sites (*NdeI*, *HincII*, *Eco47III*, and *NotI*) within the downstream T portion of the direct repeat without changing any of the HTK amino acid sequence. No new mutations other than those intended were introduced into the vectors as was verified by DNA sequencing analysis (data not shown).

**Cells, transfections, and infections.** PG13 (obtained from the American Type Culture Collection) is a helper cell line expressing the MLV *gag-pol* and the gibbon ape leukemia virus envelope (30). The absence of the gibbon ape leukemia virus receptor on murine cells prevents reinfection of the PG13 helper cells. The target 143B cells (obtained from the American Type Culture Collection) are a thymidine kinase-deficient human osteosarcoma cell line. Cells were maintained in Dulbecco's modified Eagle's medium (ICN Biomedicals) supplemented with penicillin (50 U/ml; Gibco), streptomycin (50  $\mu$ g/ml; Gibco), and bovine calf serum (10% for PG13 and 6% for 143B; HyClone Laboratories).

PG13 helper cells were transfected with 10  $\mu$ g of each retroviral vector by the previously described CaPO<sub>4</sub> method (38). Cells were subjected to G418 (an analog of neomycin) selection at a final concentration of 600  $\mu$ g/ml (0.87 mM; Gibco). At least 2,000 G418-resistant colonies derived from each vector were pooled and expanded. For each vector, 2.5  $\times$  10<sup>6</sup> transfected G418-resistant cells were plated on 100-mm-diameter dishes, and the culture medium was changed 24 h later. For infections involving vectors KD-HTT4\*K and KD-HTpT4\*K, 5.0  $\times$  10<sup>6</sup> cells were plated. Virus was harvested 24 h later, serially diluted, and used to infect 143B target cells plated at 2  $\times$  10<sup>5</sup> cells per 60-mm-diameter dish in the presence of Polybrene (50  $\mu$ g/ml) as previously described (16). Infected 143B cells were subjected to either G418 (400  $\mu$ g/ml; 0.58 mM) or hypoxanthine-aminopterin-thymidine (HAT; as specified by Boehringer Mannheim) selection 1 day postinfection. Drug-resistant colonies were counted, and viral titers were determined from two to six independent experiments.

**Southern analysis.** Genomic DNA was isolated, and proviral DNA was analyzed by Southern blot hybridization using standard procedures from pools of at least 2,000 HAT- or G418-resistant colonies (38). A 1.2-kb IRES-*neo* DNA fragment was used to generate a probe with [ $\alpha$ -<sup>32</sup>P]dCTP (specific activity, >10<sup>9</sup> cpm/ $\mu$ g; ICN Biomedicals) by using a Random Priming DNA-labeling kit (Boehringer Mannheim). The membrane was exposed to X-ray film (Kodak) and PhosphorImager cassette (Molecular Dynamics). Quantitation of bands was performed with the ImageQuant program (Molecular Dynamics).

**PCR analysis of KD-HTT4\*K and KD-HTpT4\*K.** After HAT selection of the 143B-infected cells, single cell colonies were isolated from separate dishes, expanded, and lysed (14). For each clone, the T portion of HTK was PCR amplified under standard conditions (46) with primers T for (5'-GACGATATCGTCTACGTACCCGAGCCC-3') and TrevB (5'-AGACGTGCATGGAACGAGCGCTTTGGCC-3'). The amplified T portions were then subjected to restriction enzyme digestions separately and analyzed on a 1% agarose gel.

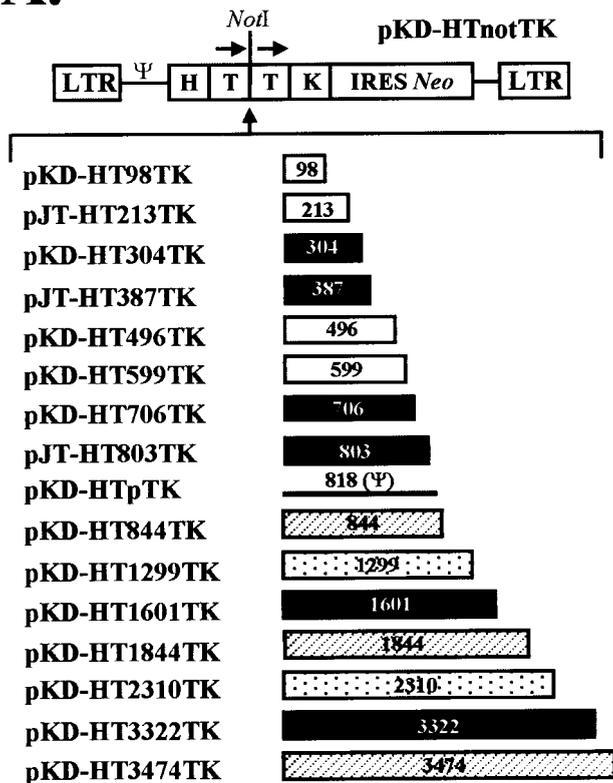
#### RESULTS

**Construction of MLV-based vectors containing direct repeats separated by spacer fragments of various lengths.** To elucidate the effect of distance between direct repeats on tem-

plate switching by RT, we constructed a series of MLV-based retroviral vectors that contained a 701-bp directly repeated sequence composed of the middle T portion of HTK (Fig. 1). Each vector contained the *cis*-acting elements necessary for viral replication in addition to *neo* expressed from IRES. In the parental vector pKD-HTnotTK, the overlapping HT and TK fragments were separated by a *NotI* restriction site. Spacer fragments that varied in length from 98 to 3,474 bp were inserted into the *NotI* site to generate a total of 16 vectors. To minimize sequence-specific effects of the spacer fragments on the frequency of RT template switching, the spacer fragments were derived from unrelated genes *ouabain*, *lacZ*, and *hygro*. Retroviral vectors pKD-HT304TK, pJT-HT387TK, pKD-HT706TK, pJT-HT803TK, pKD-HT1601TK, and pKD-HT3322TK contain spacer fragments derived from *ouabain* (Fig. 1). Retroviral vectors pKD-HT844TK, pKD-HT1844TK, and pKD-HT3474TK contain spacer fragments derived from *lacZ*. Retroviral vectors pKD-HT98TK, pJT-HT213TK, pKD-HT496TK, and pKD-HT599TK contain spacer fragments derived from *hygro*. Retroviral vectors pKD-HT1299TK and pKD-HT2310TK were derived by insertion of fragments from both *hygro* and *ouabain* (not shown). Finally, the previously described vector pKD-HTpTK contains an 818-bp spacer fragment that contains the MLV packaging signal  $\Psi$  (6). The approximate locations of the spacer fragments relative to *ouabain*, *lacZ*, and *hygro* are shown in Fig. 1B. Some of the spacer fragments derived from the same gene do not share significant sequence homology; for example, the 98- and the 213-bp spacer fragments were derived from different locations in *hygro* and are distinct in sequence. Other spacer fragments share significant sequence homology; for example, the 844-bp spacer fragment is present within the 3,474-bp spacer fragment derived from *lacZ*. However, there is no simple relationship between sequence homology and the length of the spacer fragments. Therefore, the effects of spacer fragments on the frequency of direct repeat deletions are likely to be due to the increase in distance rather than the primary sequences of the spacers.

**Functional reconstitution of HTK during a single cycle of retroviral replication.** The results of infections performed with virus derived from all 16 vectors are summarized in Table 1. The vectors were transfected into MLV-based packaging cell line PG13. Pools of G418-resistant PG13 cells containing each vector were expanded, and virus was harvested. The virus produced from each vector was used to infect 143B, a human thymidine kinase-deficient cell line. The infected cells were selected for resistance to either G418 or HAT. All cells infected with a virus were expected to express *neo* and confer resistance to G418. The results of experiments with vectors containing only the HT or the TK fragment did not confer HAT resistance (data not shown). Therefore, only cells containing proviruses that underwent direct repeat deletion and functionally reconstituted HTK were expected to confer resistance to HAT. The high virus titers observed after selection for HAT indicated that direct repeat deletions occurred at a high rate during reverse transcription and functionally reconstituted HTK. The G418 and HAT titers obtained for each vector were within twofold of each other. A twofold difference in virus titer is not statistically significant unless the experiments are repeated several times. However, for some vectors, the HAT titers were higher than the G418 titers. We have previously shown that the HAT titers are approximately 1.7-fold higher than the G418 titers when 143B cells are used for infection (6). The higher HAT titers probably reflect differential toxicities of the two selection procedures to the 143B cells.

**A.**



**B.**

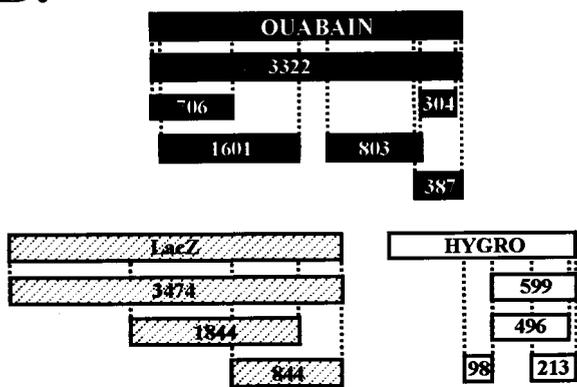


FIG. 1. Structures of MLV-based retroviral vectors containing direct repeats separated by various spacers. (A) All vectors were based on the parental vector pKD-HTnotTK, which contains *neo* expressed from IRES, and two overlapping fragments of HTK separated by a *NotI* restriction site linker. Boxes labeled T (arrows above boxes) represent a 701-bp direct repeat derived from the middle portion of HTK. Each of the 16 vectors shown contains a spacer fragment between the direct repeats ranging in size from 98 to 3,474 bp (shown within each box) cloned into the *NotI* site of pKD-HTnotTK. The spacer fragments used to generate the vectors were derived from *ouabain* (black boxes), *lacZ* (hatched boxes), *hygro* (white boxes), or multiple fragments derived from *ouabain* and *hygro* (dotted boxes). The 818-bp fragment containing the MLV packaging sequence ( $\Psi$ ) is shown as a thick line. LTR (here and in Fig. 2, 4, 5, and 6), long terminal repeat. (B) Approximate locations of fragments used as spacers are shown for fragments derived from *ouabain*, *lacZ*, and *hygro*.

TABLE 1. Virus titers after infection with vectors containing 114- to 701-bp direct repeats separated by spacer fragments of various lengths

Virus	Distance between homologous sequences (bp) <sup>a</sup>	Titer (CFU/ml) <sup>b</sup>	
		G418	HAT
KD-HTnotTK	701	$7.3 \times 10^4$	ND <sup>c</sup>
KD-HT98TK	799	$3.7 \times 10^3$	$3.9 \times 10^3$
JT-HT213TK	914	$5.0 \times 10^3$	$5.4 \times 10^3$
KD-HT304TK	1,005	$5.2 \times 10^3$	$5.5 \times 10^3$
JT-HT387TK	1,088	$3.1 \times 10^3$	$7.0 \times 10^3$
KD-HT496TK	1,197	$6.6 \times 10^3$	$8.7 \times 10^3$
KD-HT599TK	1,300	$3.3 \times 10^3$	$4.6 \times 10^3$
KD-HT706TK	1,407	$1.1 \times 10^3$	$2.0 \times 10^3$
JT-HT803TK	1,504	$6.3 \times 10^3$	$9.6 \times 10^3$
KD-HTpTK <sup>d</sup>	1,519	$6.0 \times 10^3$	$1.1 \times 10^4$
KD-HT844TK	1,545	$9.3 \times 10^4$	$1.0 \times 10^5$
KD-HT1299TK	2,000	$5.5 \times 10^3$	ND
KD-HT1601TK	2,302	$5.4 \times 10^3$	ND
KD-HT1844TK	2,545	$3.6 \times 10^3$	$4.6 \times 10^3$
KD-HT2310TK	3,011	$7.2 \times 10^3$	ND
KD-HT3322TK	4,023	$4.4 \times 10^3$	$3.8 \times 10^3$
KD-HT3474TK	4,175	$2.1 \times 10^5$	$2.4 \times 10^5$
KD-HTT4*K	701	ND	$2.7 \times 10^5$
KD-HTpT4*K	1,519	ND	$1.8 \times 10^4$
KD-HT1TK	114	$1.6 \times 10^5$	$1.1 \times 10^4$
KD-HT2TK	225	$6.7 \times 10^4$	$1.8 \times 10^4$
KD-HT4TK	349	$9.5 \times 10^4$	$1.6 \times 10^4$
KD-HTp1TK	932	$4.4 \times 10^4$	$2.6 \times 10^3$
KD-HTp2TK	1,043	$3.2 \times 10^4$	$1.2 \times 10^3$
KD-HTp4TK	1,167	$1.9 \times 10^5$	$4.7 \times 10^3$
KD-HTTK3474	701	$4.1 \times 10^1$	$1.2 \times 10^2$
KD-neoHTTK	701	$4.3 \times 10^3$	ND

<sup>a</sup> Length of one direct repeat plus length of the spacer fragment between the direct repeats.

<sup>b</sup> Viral titers shown are the Average of two to six independent experiments.

<sup>c</sup> ND, not determined.

<sup>d</sup> Data from reference 6.

**Deletion frequencies of vectors in which the direct repeats are separated by spacer fragments of various lengths.** After infection of 143B cells with virus derived from each vector, G418-resistant and HAT-resistant colonies were pooled and expanded. Genomic DNAs were isolated from the pools of infected cells for Southern analysis. The genomic DNA was digested with *XbaI*, which cuts in both the 5' and 3' long terminal repeats of the proviral DNA (Fig. 2A). Proviruses that did not undergo direct repeat deletion were expected to generate an *XbaI* band that is 5.0 kb plus the size of the spacer fragment. Proviruses that underwent direct repeat deletion were expected to generate a 4.3-kb band, regardless of the size of the spacer fragment.

Southern analyses of pools of G418-resistant colonies derived from infected 143B cells were performed (Fig. 2B and C). Since the G418-resistant pools contained a population of both undeleted and deleted proviruses, the deletion frequency could be determined by comparing the intensities of the deleted and undeleted bands. The deletion frequencies for each vector were obtained from at least two independent infections and at least three different Southern blots. The average deletion frequency during reverse transcription of each vector, determined by comparison of the intensities of the undeleted and deleted bands, is shown (Fig. 2B and C). The Southern analyses of vectors in which the homologous sequences (spacer fragment

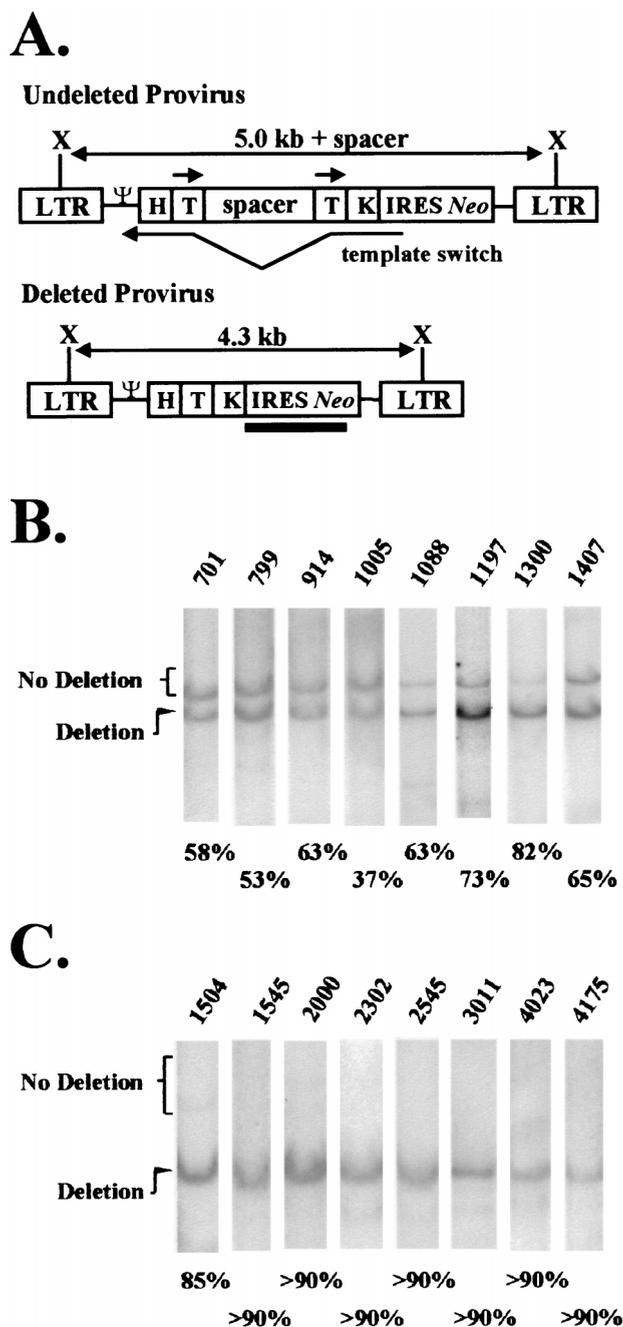


FIG. 2. Southern analysis of genomic DNAs from pools of infected cells. (A) Restriction digestion with *Xba*I (labeled X) of proviral DNAs that did not undergo direct repeat deletion (undeleted provirus) is expected to yield an undeleted band that is 5.0 kb plus the size of the spacer fragment inserted between the direct repeats. The same *Xba*I digestion of proviral DNAs that underwent a direct repeat deletion (deleted provirus) is expected to yield a 4.3-kb band. The black bar below the deleted provirus indicates the 1.2-kb IRES-neo probe used for Southern analysis. (B) Southern analysis of proviral DNAs from pools of G418-resistant 143B cells infected with vectors in which the homologous sequences were separated by less than 1,500 bp (length of spacer fragment plus 701 bp). Numbers above each lane refer to the distance between homologous sequences in each vector. Quantitative analysis of the bands representing the deleted and undeleted proviruses provides a frequency of direct repeat deletion (shown below each lane). (C) Southern analysis of proviral DNAs from pools of G418-resistant 143B cells infected with vectors in which the homologous sequences were separated by greater than 1,500 bp. The average deletion frequencies shown in B and C are also adjusted for deletions that occur during transfection of the vectors into packaging cells (average, 5%).

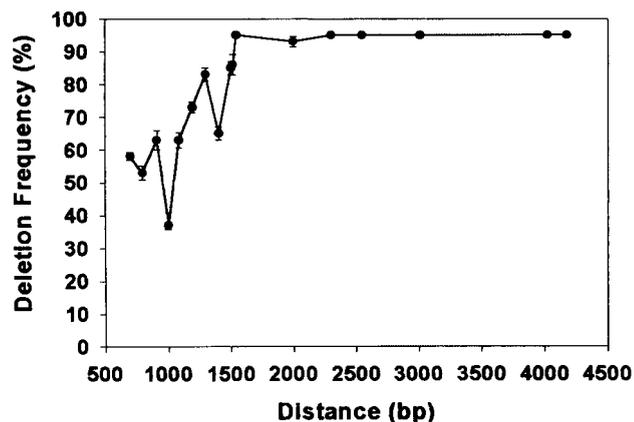


FIG. 3. Effect of distance between homologous sequences on frequency of direct repeat deletion. The distance between homologous sequences is shown in base pairs, and the deletion frequency is shown as the percentage of proviruses that underwent direct repeat deletion. The error bars ( $\pm$  standard error) represent the average of at least two independent infections and three to nine Southern analyses. The standard error for each vector ranged from 0 to  $\pm 3.2\%$ . The average deletion frequencies are also adjusted for deletions that occurred during transfection of the vectors into packaging cells (average, 5%).

+701 bp; [Table 1]) were separated by less than 1,500 bp are summarized in Fig. 2B. The deletion frequencies ranged from 37 to 82%. It was also important to determine the frequency of direct repeat deletion during transfection of PG13 helper cells. Southern analyses of genomic DNAs from pools of transfected PG13 cells indicated that approximately 5% of the viral DNAs underwent deletions during the process of transfection (data not shown). Therefore, the observed deletion frequencies were adjusted by subtraction of 5% to accurately reflect the frequency of direct repeat deletion during one cycle of reverse transcription.

Southern analyses of vectors in which the homologous sequences were separated by greater than 1,500 bp are summarized in Fig. 2C. For these vectors, the deletion frequencies ranged from 85% to greater than 90%. The band representing the undeleted provirus was not detectable for vectors in which the homologous sequences were separated by greater than 1,545 bp. Based on the sensitivity of the Southern blots, we estimate that at least 90% of the proviruses underwent direct repeat deletion during reverse transcription. Southern analyses of genomic DNAs from pools of HAT-resistant cells were also performed. As expected, only the deleted 4.3-kb proviral band was detected (data not shown).

**Effect of distance between homologous sequences on deletion frequency.** The deletion frequencies obtained from Southern analyses were plotted with respect to the distance between homologous sequences in Fig. 3. As the error bars indicate, the deletion frequencies observed were highly reproducible in independent experiments. The results show that in general, the frequency of direct repeat deletions increased as the distance between homologous sequences increased from 701 to 1,504 bp. The deletion frequencies for pKD-HT304TK and pKD-HT706TK were lower than for some vectors in which the homologous sequences were closer to each other. For vectors in which the distance between homologous sequences was between 1,545 to 4,175 bp, a greater than 90% frequency of direct repeat deletion was observed (Fig. 3). The results show that a high plateau of direct repeat deletion was reached when the distance between homologous sequences was increased to greater than 1,500 bp.

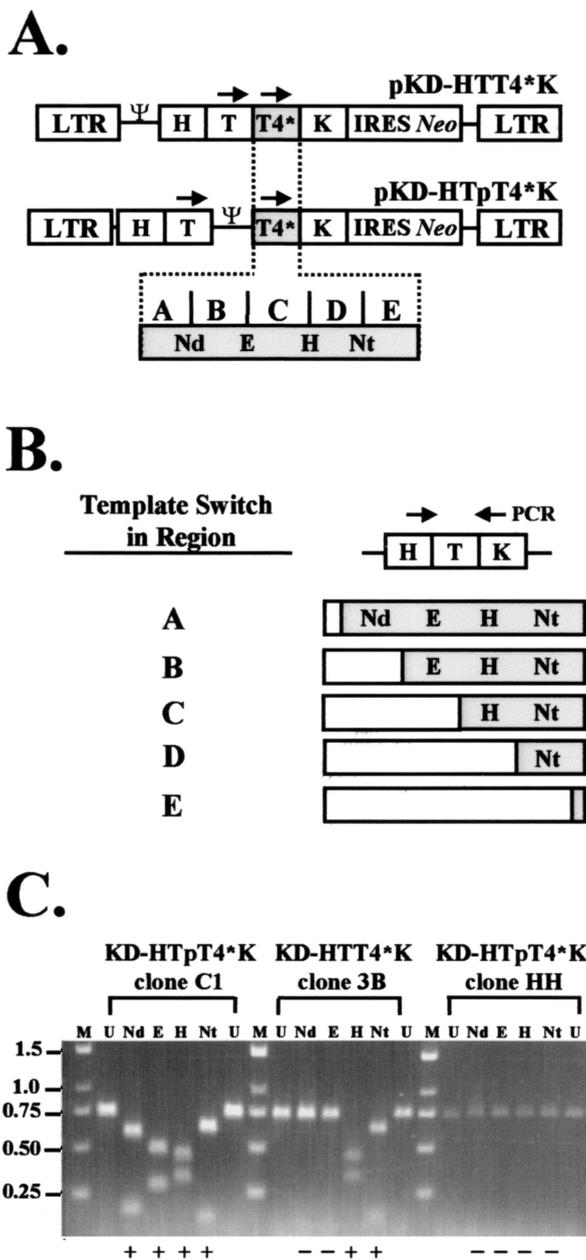


FIG. 4. Analysis of positions at which RT switches templates within the direct repeats. (A) Structures of MLV-based retroviral vectors KD-HTT4\*K and KD-HTpT4\*K. Each vector contains the 701-bp direct repeat derived from the HTK gene (open box labeled T and shaded box labeled T4\*, with arrows above both boxes). The directly repeated sequences are adjacent to one another in the vector pKD-HTT4\*K and are separated by a 818-bp spacer fragment containing the MLV Ψ in the vector pKD-HTpT4\*K. The 3' copy of the repeat (shaded boxes) in both vectors contains four restriction sites that are not present in the 5' copy of the repeat. Restriction sites: Nd, *Nde*I; E, *Eco*47III; H, *Hinc*II; Nt, *Not*I. The presence of these restriction sites divides the 3' direct repeat into 5 regions labeled A, B, C, D, and E. (B) Expected structures of proviral clones after direct repeat deletion. Amplification of the T portion of the reconstituted HTK using PCR is expected to yield products that can be digested with various restriction enzymes. The observed pattern of restriction sites reflects the region in which RT switched templates. For example, an RT template switch in region A is expected to generate a product containing *Nde*I, *Eco*47III, *Hinc*II, and *Not*I restriction sites. (C) Restriction digests of PCR products derived from three representative proviral cell clones. Cell clones C1 and HH were derived from KD-HTpT4\*K, and cell clone 3B was derived from KD-HTT4\*K. Sizes of molecular weight marker fragments (lanes M) are shown in kilobases on the left. Lanes U represent undigested PCR product. The plus or minus sign below each lane indicates that the PCR product was or was not cleaved by the restriction enzyme.

TABLE 2. Frequencies of RT template switching in different regions of the direct repeats adjacent to one another or separated by a spacer fragment of 818 bp

Region of direct repeat <sup>a</sup> (length [bp])	KD-HTT4*K			KD-HTpT4*K		
	No. of clones	% RT template switching		No. of clones	% RT template switching	
		Per region <sup>b</sup>	Per 100 bp <sup>c</sup>		Per region	Per 100 bp
A (153)	11	30	20	7	70	46
B (110)	7	16	15	5	33	30
C (158)	8	15	9	10	40	25
D (180)	4	7	4	2	7	4
E (100)	4	7	7	2	7	7
Total	34			26		

<sup>a</sup> See Fig. 4 for definition of regions.

<sup>b</sup> The percent RT template switching per region was adjusted to the length of each region to estimate the percent RT template switching per 100 bp by using the following formula: (% RT template switching per region/length of region [bp]) × 100 bp. For example, the % RT template switching per 100 bp for region A of KD-HTT4\*K was (30%/153 bp) × 100 bp = 20%.

<sup>c</sup> The percent RT template switching per 100 bp was calculated as follows. First, the fraction of RTs that underwent a template switch and reconstituted a functional HTK gene (57% for KD-HTTK and 91% for KD-HTpTK) was used to estimate the theoretical population of undeleted and deleted clones analyzed. Specifically, 34 clones of KD-HTT4\*K underwent direct repeat deletion, which represents 57% of 60 clones. Similarly, 26 clones of KD-HTpT4\*K underwent direct repeat deletion, which represents 91% of 29 clones. Next, the number of RTs that copied each region was estimated by subtraction of the fraction of RTs that switched templates 3' to the region by using the following formula: (number of clones that switched templates in region)/(theoretical population - number of clones that switched templates in regions 3' to the region) × 100. For example, the number of RTs that copied region A of KD-HTT4\*K was 11/(60 - 23) × 100% = 30%. This calculation indicates that 30% of RTs that copied region A of KD-HTT4\*K underwent a template switch.

The frequency of template switching is higher in the 5' regions of the direct repeat relative to the 3' regions. To analyze the region(s) within the downstream 701-bp repeat in which template switching events occurred, vectors pKD-HTT4\*K and pKD-HTpT4\*K were analyzed (Fig. 4A). Each vector contained four restriction sites that were present in only the 3' copy of the direct repeat. These restriction sites were generated by introduction of silent mutations. The presence of these four restriction sites divided the 3' direct repeat into five regions (A through E) ranging in size from 100 to 180 bp. An RT template switch in each of the five regions is expected to generate a deleted provirus that contains a different pattern of restriction sites (Fig. 4B). A template switch in region A is expected to generate a provirus containing *Nde*I, *Eco*47III, *Hinc*II, and *Not*I restriction sites. A template switch in region B is expected to generate a provirus containing *Eco*47III, *Hinc*II, and *Not*I restriction sites. A template switch in region C is expected to generate a provirus containing *Hinc*II and *Not*I restriction sites. A template switch in region D is expected to generate a provirus containing the *Not*I restriction site. Finally, a template switch in region E is expected to generate a provirus that does not contain any of the four restriction sites.

Each vector was transfected into PG13 helper cells, and over 2,500 G418-resistant colonies were pooled and expanded for each experiment. Virus was harvested and used to infect 143B target cells. The cells were subjected to HAT selection, and the resulting single cell clones were isolated and expanded. The viral titers for KD-HTT4\*K and KD-HTpT4\*K (Table 1) were comparable to the titers of similar vectors lacking the four restriction sites (KD-HTTK and KD-HTpTK in reference 6). The T portion of the provirus in each clone was PCR amplified

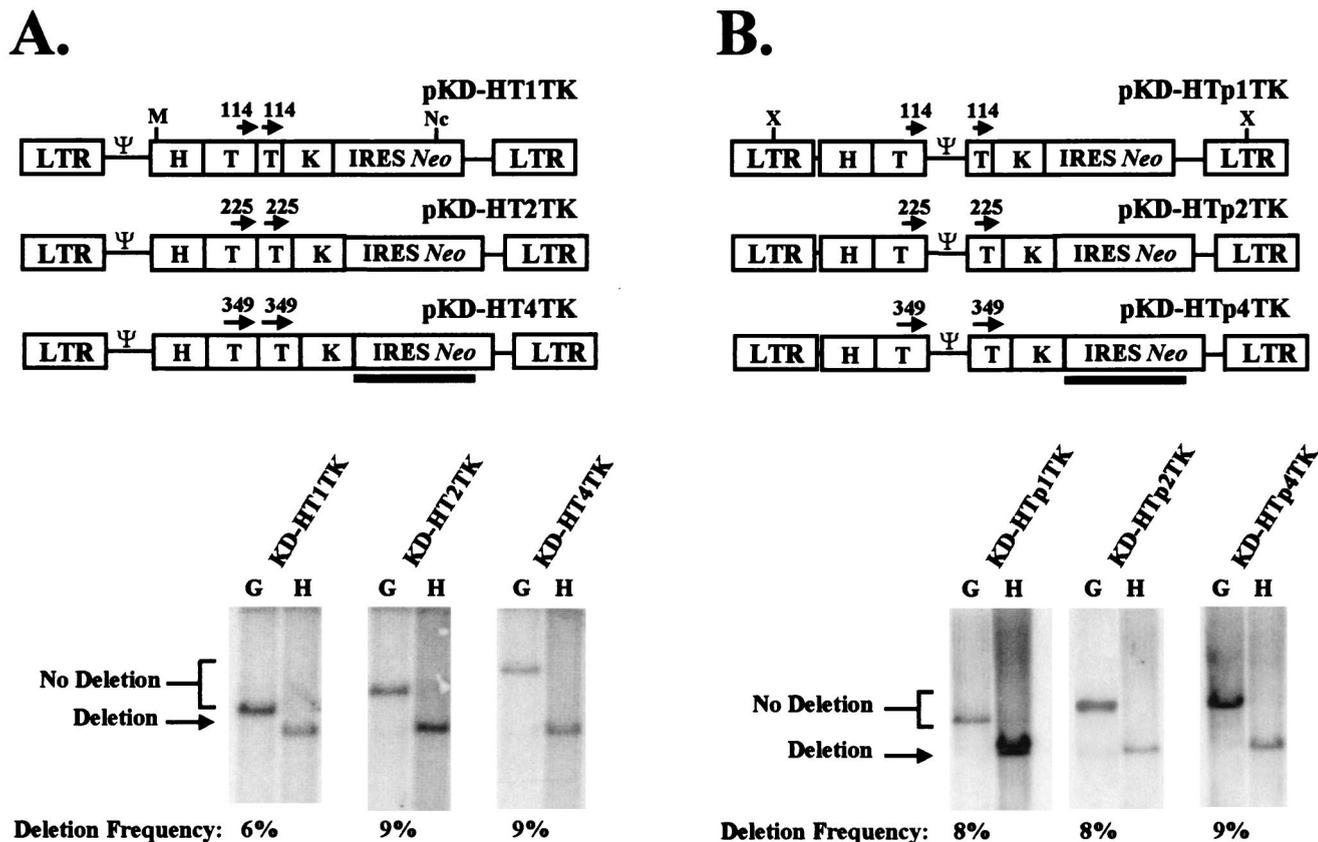


FIG. 5. Southern analysis of vectors containing shorter direct repeats. (A) Structures of MLV-based retroviral vectors pKD-HT1TK, pKD-HT2TK, and pKD-HT4TK. Each vector contains overlapping fragments of HTK generating a 114-bp (pKD-HT1TK), 225-bp (pKD-HT2TK), or 349-bp (pKD-HT4TK) direct repeat. Proviral DNAs from KD-HT1TK, KD-HT2TK, and KD-HT4TK were digested with *Mlu*I (M) and *Nco*I (Nc) to generate 2.4-, 2.5-, and 2.7-kb undelleted bands, respectively. The black bar below pKD-HT4TK represents the 1.2-kb IRES-*neo* probe. The Southern blot shown below the vectors is representative of analyses performed to determine the frequency of direct repeat deletion. Lanes G represent analyses of genomic DNAs derived from pools of G418-resistant infected cells; lanes H represent analyses of genomic DNAs derived from pools of HAT-resistant infected cells. The G418-resistant cells are expected to generate undelleted bands (No Deletion) as well as 2.3-kb deleted bands (Deletion). The HAT-resistant cells are expected to generate only the 2.3-kb deleted bands. The frequencies of deletion are shown below lanes G and represent averages of three independent experiments. (B) Structures of MLV-based retroviral vectors pKD-HTp1TK, pKD-HTp2TK, and pKD-HTp4TK. Each vector contains overlapping fragments of HTK generating a 114-bp (pKD-HTp1TK), 225-bp (pKD-HTp2TK), or 349-bp (pKD-HTp4TK) direct repeat separated by the 818-bp  $\Psi$ . Proviral DNAs from KD-HTp1TK, KD-HTp2TK, and KD-HTp4TK were digested with *Xba*I (X) to generate 4.4-, 4.5-, and 4.7-kb undelleted bands, respectively. In the Southern analysis shown below the vectors, the G418-resistant cells are expected to generate undelleted bands (No Deletion) as well as 3.5-kb deleted bands (Deletion). The HAT-resistant cells are expected to generate only the 3.5-kb deleted band. The frequencies of deletion are shown below lanes G and represent averages of three independent experiments.

and subjected to restriction enzyme digestion with *Nde*I, *Eco*47III, *Hinc*II, or *Not*I. A representative agarose gel is shown in Fig. 4C. The PCR product derived from KD-HTpT4\*K clone C1 was cleaved by all four enzymes, indicating that a template switch had occurred in region A. The PCR product derived from KD-HTT4\*K clone 3B was cleaved by *Hinc*II and *Not*I, indicating that a template switch had occurred in region C. On the other hand, the PCR product derived from KD-HTpT4\*K clone HH was not cleaved by any of the four enzymes, indicating that a template switch occurred in region E.

Results of restriction analysis of PCR products derived from 34 HAT-resistant cell clones containing KD-HTT4\*K proviruses and 26 cell clones containing KD-HTpT4\*K proviruses are summarized in Table 2. Analysis of KD-HTT4\*K proviruses indicated that 11, 7, 8, 4, and 4 proviruses underwent an RT template switch in regions A, B, C, D, and E, respectively. Analysis of KD-HTpT4\*K proviruses indicated that 7, 5, 10, 2, and 2 proviruses underwent an RT template switch in regions A, B, C, D, and E, respectively. These data were used to determine the rate of RT template switching per 100 bp. The

observed frequencies of direct repeat deletions were normalized for the length of each region as well as the fraction of RTs expected to reverse transcribe each region. It is important to note that the fraction of RTs that copy a region is dependent on the fraction of RTs that undergo a template switch 3' to the region (see Table 2, footnote c). The frequencies of RT template switching per 100 bp observed for KD-HTT4\*K were 20, 15, 9, 4, and 7% for regions A, B, C, D, and E, respectively. In comparison, the frequencies of RT template switching per 100 bp observed for KD-HTpT4\*K were 46, 30, 25, 4, and 7% for regions A, B, C, D, and E, respectively. These results indicated that in both KD-HTT4\*K and KD-HTpT4\*K vectors, RT switched templates in all five regions of the direct repeat. However, the frequency of RT template switching per 100 bp was not equal in all five regions; the probability of RT switching templates was increased as RT reverse transcribed the template from the 3' (region E) to the 5' (region A) ends of the direct repeat. When the direct repeats were separated by the 818-bp spacer (KD-HTpT4\*K), the frequency of RT template switching per 100 bp was increased approximately twofold in

the 5' half of the direct repeat (regions A, B, and C) relative to the same regions in KD-HTT4\*K.

**The 818-bp  $\Psi$  sequence does not increase the rate of template switching for shorter direct repeats.** To determine whether the increase in template switching observed between KD-HTTK (57%) and KD-HTpTK (91%) could also be observed with shorter direct repeats, we constructed a series of vectors containing shorter direct repeats (Fig. 5). Vectors pKD-HT1TK, pKD-HT2TK, and pKD-HT4TK contain direct repeats of 114, 225, and 349 bp, respectively, adjacent to one another (Fig. 5A). All vectors lack regions A, B, and the 5' half of region C as defined in Fig. 4B. Specifically, the 349-bp repeat contains region E, D, and 69 bp of region C. The 225-bp repeat contains region E plus 125 bp of region D. The 114-bp repeat contains region E plus 14 bp of region D. Each vector was separately transfected into PG13 helper cells, and over 2,000 G418-resistant colonies were pooled and expanded. Virus was harvested, serially diluted, and used to infect 143B target cells. The target cells were subjected to either G418 or HAT selection, and viral titers were determined from the resistant colonies. As shown in Table 1, similar titers were obtained after both G418 and HAT selection.

Genomic DNAs were also isolated from pools of G418- and HAT-resistant colonies. To analyze the structure of the proviral DNA for KD-HT1TK, KD-HT2TK, and KD-HT4TK, genomic DNA was cut with *Mlu*I and *Nco*I to yield 2.4-, 2.5-, and 2.7-kb undeleated bands, respectively, and a 2.3-kb deleted band in Southern analyses (Fig. 5A). As expected, the HAT-resistant pools of colonies generated only the deleted proviral band (Fig. 5A, lanes H), and the G418-resistant pools generated both the deleted and undeleated proviral bands (Fig. 5A, lanes G). As before, the intensities of the deleted and undeleated proviral bands in the G418-resistant pools of cells were compared to determine the deletion frequency. The average deletion frequencies from three independent experiments were 6% ( $\pm 5\%$ ) for KD-HT1TK, 9% ( $\pm 2\%$ ) for KD-HT2TK, and 9% ( $\pm 4\%$ ) for KD-HT4TK. Thus, the shorter direct repeats exhibited a significantly lower frequency of RT template switching in comparison to the 701-bp direct repeat in KD-HTnotTK (Fig. 2B).

Southern analysis was also performed for KD-HTp1TK, KD-HTp2TK, and KD-HTp4TK, in which the shorter direct repeats were flanking the 818-bp  $\Psi$  sequence (Fig. 5B). Genomic DNAs isolated from G418-resistant pools of cells were digested with *Xba*I to generate 4.4-, 4.5-, and 4.7-kb undeleated bands (Fig. 5B, lanes G) and a 3.5-kb deleted band (Fig. 5B, lanes H). Genomic DNAs from HAT-resistant pools of cells generated only the 3.5-kb deleted band. The deletion frequencies obtained from G418-resistant pools of cells were 8% ( $\pm 5\%$ ) for KD-HTp1TK, 8% ( $\pm 3\%$ ) for KD-HTp2TK, and 9% ( $\pm 9\%$ ) for KD-HTp4TK. Thus, the presence of the 818-bp  $\Psi$  sequence did not significantly increase the rate of RT template switching for shorter direct repeats. These low deletion frequencies observed for vectors containing shorter direct repeats were consistent with the observation that only 4 and 7% of the RTs switched templates per 100 bp in regions D and E, respectively (Table 2).

**Effect of viral RNA length and location of direct repeat on the frequency of direct repeat deletion.** Most of the vectors tested in these studies generated RNAs that were shorter than the 8.3-kb wild-type MLV RNA (39). To rule out the possibility that the deletion frequencies were affected by the size of the packaged RNA, we constructed pKD-HTTK3474, which contained the same direct repeat as KD-HTnotTK; however, a 3,474-bp *lacZ* fragment was also placed 3' of *neo* (Fig. 6A). The overall length of the vector RNA was 8.9 kb, similar to that of wild-type MLV RNA. Vector pKD-HTTK3474 was trans-

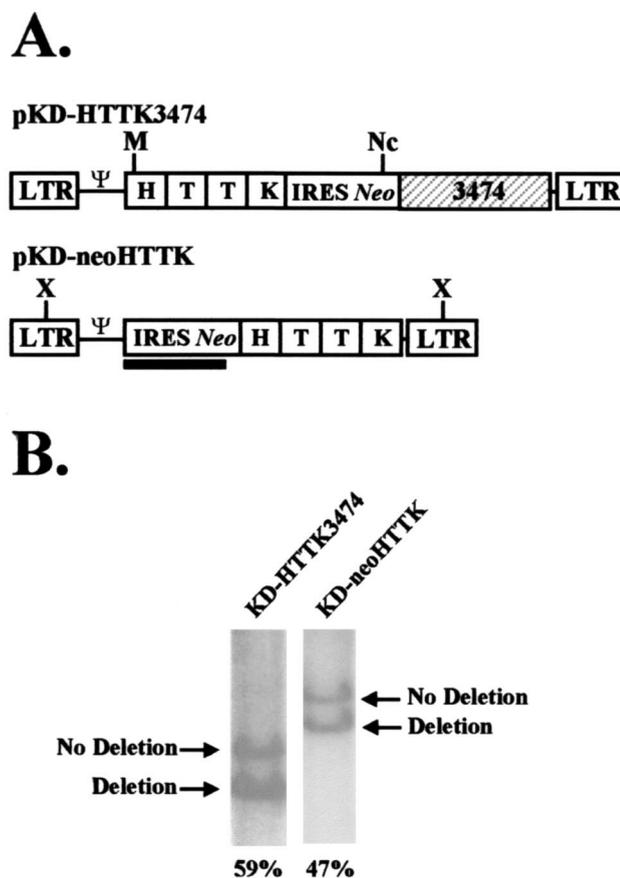


FIG. 6. Southern analysis of a direct repeat vector that generates a full-length RNA and a vector that contains the direct repeat at a 3' location. (A) Structures of vectors pKD-HTTK3474 and pKD-neoHTTK. The vector pKD-HTTK3474 is similar to pKD-HTnotTK except that a 3,474-bp fragment encoding *lacZ* was inserted 3' of *neo* (hatched box). pKD-neoHTTK was derived by insertion of the HTTK cassette containing the 701-bp direct repeat 3' of *neo*. The black bar below pKD-neoHTTK represents the 1.2-kb IRES-*neo* probe. Restriction sites: M, *Mlu*I; Nc, *Nco*I; X, *Xba*I. (B) Southern analysis of proviral DNAs from pools of G418-resistant cells. Proviral DNAs derived from KD-HTTK3474 and digested with *Mlu*I and *Nco*I are expected to generate a 3.2-kb undeleated band (No Deletion) and a 2.5-kb deleted band (Deletion). Proviral DNAs derived from KD-neoHTTK and digested with *Xba*I are expected to generate a 5.0-kb undeleated band (No Deletion) and a 4.3-kb deleted band (Deletion). The average deletion frequencies from two independent experiments (adjusted for deletions that occurred during transfection; average, 5%) are shown below each lane.

ected into PG13 helper cells, and G418-resistant colonies were pooled and expanded. Virus was harvested and used to infect 143B cells which were subjected to either G418 or HAT selection. The resulting titers are shown in Table 1. For Southern analysis, genomic DNAs were isolated from pools of G418-resistant 143B colonies and digested with both *Mlu*I and *Nco*I to generate a 3.2-kb undeleated proviral band and a 2.5-kb deleted proviral band. As shown in Fig. 6B, the G418-resistant pools for KD-HTTK3474 exhibited a deletion frequency of 59%, which was nearly identical to the deletion frequency observed for KD-HTnotTK (58% [Fig. 2B]). Thus, the size of the viral RNA did not affect the frequency of direct repeat deletion.

It was also possible that the location of the direct repeat in the viral RNA affected the frequency of deletion. To test the effect of location on the deletion frequency, we constructed pKD-neoHTTK, in which the HTTK cassette containing the 701-bp direct repeat was placed 3' of *neo* (Fig. 6A). Vector

pKD-neoHTTK was transfected into PG13 helper cells, and G418-resistant colonies were pooled and expanded. Virus was harvested and used to infect 143B cells, which were then selected for G418 resistance. The virus titers obtained are shown in Table 1. Genomic DNAs were isolated from the G418-resistant 143B cells and digested with *Xba*I to generate a 5.0-kb undeleted proviral band and a 4.3-kb deleted proviral band. As shown in Fig. 6B, KD-neoHTTK resulted in a deletion frequency of 47%, which was comparable to the deletion frequency obtained for KD-HTnotTK (Fig. 2B). Thus, the location of the direct repeats relative to the 5' or 3' ends of the viral RNA did not have a drastic effect on the frequency of deletion.

## DISCUSSION

**Increasing distance between homologous sequences increases the frequency of RT template switching.** In this study, we observed that two vectors in which the directly repeated sequences were separated by unrelated sequences of approximately the same length as  $\Psi$  (pKD-HT803TK and pKD-HT844TK) also displayed a higher frequency of RT template switching. Therefore, the results of these studies do not support the previous hypothesis that a sequence-specific characteristic of  $\Psi$  affects the frequency of RT template switching within adjacent direct repeats to a greater extent than unrelated spacers of the same length could not be excluded since a greater than 90% deletion frequency was achieved with the unrelated spacers. These results are not inconsistent with previous reports of an increase in template switching at the MLV dimerization linkage structure (27–29). These previous reports indicated that the location of intermolecular RT template switching was within  $\Psi$ , whereas in this study the location of RT template switching events, which are most likely to be intramolecular events (15), involved directly repeated sequences adjacent to  $\Psi$ .

The general increase in RT template switching with increasing distance between homologous sequences was unlikely to be a result of the effect of spacer fragments on the secondary structure of the template for the following reasons. First, most of the spacer fragments were derived from unrelated genes and did not share any sequence homology. Second, the vectors that shared sequence homology did not correlate with the distance between homologous sequences. Third, the G/C content of the spacer fragments ranged from 48 to 62% and did not correlate with the frequency of deletion (data not shown). Fourth, an analysis of the predicted RNA secondary structures of KD-HTT4\*K and KD-HTpT4\*K performed with the mfold program showed no drastic alterations in the RNA secondary structure (56). Since sequences up to 3-kb in length can be folded by using mfold, we folded 2,862 nucleotides for the KD-HTT4\*K vector, beginning at the 5' end of the  $\Psi$  fragment and ending at the 3' end of the K portion of HTK. For the KD-HTpT4\*K vector, we folded 2,862 nucleotides beginning with the 5' end of the H portion of HTK and ending at the 3' end of the K portion of HTK. Although small alterations in the RNA secondary structures were observed, it was not possible to attribute the increase in template switching to specific differences in the RNA secondary structures (reference 56 and data not shown).

**A model for RT template switching.** We have developed a model to explain the observation that distance between homologous sequences increases the frequency of direct repeat deletion (Fig. 7). It is proposed that homologous interactions

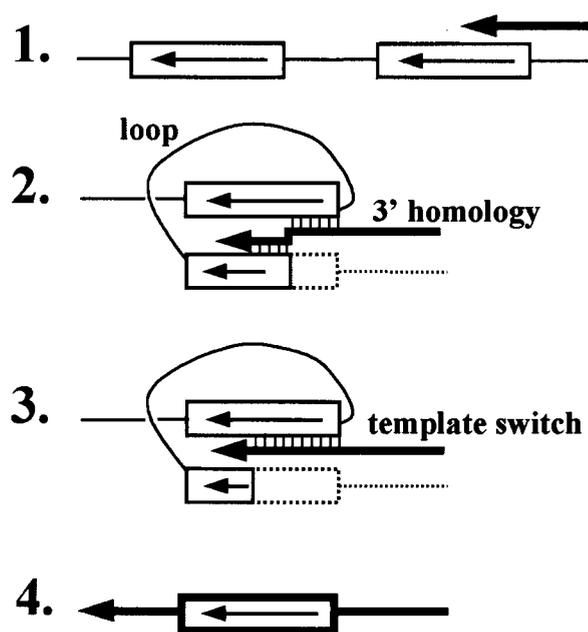


FIG. 7. Model for RT template switching. Open boxes with arrows represent direct repeats. Thin lines represent template RNA, and thick lines represent nascent DNA. (Step 1) During minus-strand synthesis, RT begins copying the 3' copy of the direct repeat. (Step 2) The template RNA undergoes a conformational rearrangement to allow nascent DNA 3' to the RT to base pair with homologous sequences in the 5' copy of the direct repeat in the template (labeled 3' homology). The conformational rearrangement is depicted as formation of a loop. (Step 3) RT dissociates from the 3' direct repeat and reassociates with the homologous sequence in the 5' direct repeat, resulting in a template switch. (Step 4) Final product of reverse transcription after the RT template switch.

between newly synthesized DNA sequences 3' of the RT with complementary sequences on the template (labeled "3' homology"), increases the probability of RT switching templates. As shown in step 1, RT begins to reverse transcribe the 3' copy of the direct repeat. RNase H is expected to degrade the template RNA 3' of the RT. In step 2, conformational rearrangements allow hydrogen bonding between the newly synthesized DNA and the complementary sequences in the 5' copy of the direct repeat. The conformational rearrangements that permit the interaction between the nascent DNA and the template are depicted as a formation of a loop in the template. In step 3, RT dissociates from the 3' direct repeat and reassociates with the homologous sequence in the 5' direct repeat. The proposed homologous interactions of template and nascent DNA 3' to the RT serve to bring the homologous acceptor template in close proximity to the RT and the growing end of the DNA, which increases the probability of a template switch. The proposed interactions also increase the probability that RT reassociates with the 5' direct repeat rather than the 3' direct repeat because of duplex formation and branch migration. After the template switch, RT continues to copy the 5' direct repeat. In step 4, the final product of reverse transcription is shown after the template switch has deleted one copy of the direct repeat and sequences between the direct repeats. The model depicts the deletion of direct repeats occurring intramolecularly; however, it is possible that the deletions can occur intermolecularly with the copackaged RNA (not shown). Previous results have shown that direct repeat deletions primarily involve intramolecular template-switching events (15). This model, revised from the previously proposed model (22), pro-

vides for the first time a possible mechanism for the observed increase of RT template switching with increasing distance.

A minus-strand exchange model was previously proposed as a mechanism for intermolecular template switching and recombination between copackaged viral genomes (3). This model also proposed that the nascent DNA and the copackaged RNA template form a duplex 3' of the RT, which promotes intermolecular template switching. Our *in vivo* analysis of KD-HTT4\*K and KD-HTpT4\*K supports the notion that increasing the length of homologous sequences 3' to the RT was associated with an increase in RT template switching.

The model also provides an explanation for the results obtained with vectors in which the homologous sequences were separated by spacer fragments of various lengths. It is possible that when the distance between homologous sequences is greater than 1,500 bp, a structural constraint associated with the template is relieved (depicted as formation of a loop) so that the nascent DNA can interact with the homologous sequences in the 5' repeat. As a result, direct repeat deletion may occur at a high rate. Conversely, when the distance between homologous sequences is less than 1,500 bp, structural constraints may be present that inhibit or prevent interactions between the nascent DNA and the 5' direct repeat, resulting in a lower frequency of RT template switching. Although RNA is believed to be very flexible, its association with the viral nucleocapsid protein and possibly other viral proteins may reduce the flexibility and result in the structural constraints (5). On the other hand, the presence of nucleocapsid protein also has been shown *in vitro* to promote template switching involving strand transfer events (5, 8, 36, 52). Therefore, the role of NC in direct repeat deletion is unknown.

The hypothesis that a minimum distance between homologous sequences is needed for efficient duplex formation 3' to the RT is consistent with the observation that the  $\Psi$  spacer fragment did not increase the frequency of deletion of shorter direct repeats. In vectors containing 114-, 225-, and 349-bp direct repeats, the presence of  $\Psi$  resulted in increasing the distance between homologous sequences to 932, 1,043, and 1,167 bp, respectively. It is possible that since the distance between homologous sequences was not increased to greater than 1,500 bp, we did not observe an increase in the rate of RT template switching. It is also possible that the shorter direct repeats did not have as much 3' homology as the larger direct repeats, and the reduced length of 3' homology could have resulted in a low deletion frequency. Finally, since the deletion frequencies were low with shorter direct repeats, the sensitivity of the Southern blotting analysis may have precluded detection of an increase in the deletion frequency.

The frequencies of deletions obtained with the shorter direct repeats were lower than previously observed (22, 33). A 388-bp direct repeat separated by the spleen necrosis virus packaging signal provided a deletion frequency of approximately 40%, whereas in this study a 349-bp repeat provided a deletion frequency of 9% (22). These results suggest that the sequence of the direct repeat, as well as the size of the repeat and the distance between the repeats, may influence the rate of template switching.

The model proposed here suggests that degradation of the template RNA by RNase H is necessary for duplex formation 3' to RT. Previous data from our laboratory indicated that RT switched templates at equal rates during RNA- and DNA-dependent DNA synthesis, suggesting that RNase H activity was not required for RT template switching (2). It is conceivable that RNase H degradation of the template increases the rate of RT template switching during minus-strand DNA synthesis, whereas other mechanisms, such as displacement syn-

thesis, increase the rate of RT template switching during plus-strand DNA synthesis.

The results of these studies suggest that the dynamic structure of the viral RNA influences RT template-switching events. We have utilized the fact that large distances do not decrease the rate of direct repeat deletion to develop MLV-based self-inactivating and self-activating retroviral vectors that delete the selectable marker and  $\Psi$  at greater than 99% efficiency in one cycle of retroviral replication (6a). Further studies are under way to test the model that 3' homology plays a major role in promoting RT template-switching events.

#### ACKNOWLEDGMENTS

We especially thank Wei-Shau Hu for intellectual input throughout the project, and we thank Jennifer Tordilla and Yegor Voronin for technical assistance. We also thank Jeffrey Anderson, Ben Beasley, Sara Cheslock, Que Dang, Elias Halvas, Carey Hwang, and Evgenia Svarovskaia for helpful comments on the manuscript and for valuable discussion of the results.

This work was supported by the Public Health Service grant CA58875 from the National Institutes of Health and the American Cancer Society grant VM84706 to V.K.P.

#### REFERENCES

- Baltimore, D. 1970. RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**:1209-1211.
- Bowman, R. R., W. S. Hu, and V. K. Pathak. 1998. Relative rates of retroviral reverse transcriptase template switching during RNA- and DNA-dependent DNA synthesis. *J. Virol.* **72**:5198-5206.
- Coffin, J. M. 1996. Retroviridae and their replication, p. 1437-1500. *In* B. N. Fields et al. (ed.), *Fields virology*, vol. 2. Lippincott-Raven Publishers, Philadelphia, Pa.
- Czernilofsky, A. P., A. D. Levinson, H. E. Varmus, J. M. Bishop, E. Tischer, and H. M. Goodman. 1980. Nucleotide sequence of an avian sarcoma virus oncogene (src) and proposed amino acid sequence for gene product. *Nature* **287**:198-203.
- Darlix, J. L., M. Lapadat-Tapolsky, H. de Rocquigny, and B. P. Roques. 1995. First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *J. Mol. Biol.* **254**:523-537.
- Delviks, K. A., W. S. Hu, and V. K. Pathak. 1997.  $\Psi^-$  vectors: murine leukemia virus-based self-inactivating and self-activating retroviral vectors. *J. Virol.* **71**:6218-6224.
- Delviks, K. A., and V. K. Pathak. 1999. Development of murine leukemia virus-based self-activating vectors that efficiently delete the selectable drug resistance gene during reverse transcription. *J. Virol.* **73**:8837-8842.
- DeStefano, J., J. Ghosh, B. Prasad, and A. Raja. 1998. High fidelity of internal strand transfer catalyzed by human immunodeficiency virus reverse transcriptase. *J. Biol. Chem.* **273**:1483-1489.
- DeStefano, J. J. 1996. Interaction of human immunodeficiency virus nucleocapsid protein with a structure mimicking a replication intermediate. Effects on stability, reverse transcriptase binding, and strand transfer. *J. Biol. Chem.* **271**:16350-16356.
- DeStefano, J. J. 1994. Kinetic analysis of the catalysis of strand transfer from internal regions of heteropolymeric RNA templates by human immunodeficiency virus reverse transcriptase. *J. Mol. Biol.* **243**:558-567.
- DeStefano, J. J., R. A. Bambara, and P. J. Fay. 1994. The mechanism of human immunodeficiency virus reverse transcriptase-catalyzed strand transfer from internal regions of heteropolymeric RNA templates. *J. Biol. Chem.* **269**:161-168.
- DeStefano, J. J., R. G. Buiser, L. M. Mallaber, P. J. Fay, and R. A. Bambara. 1992. Parameters that influence processive synthesis and site-specific termination by human immunodeficiency virus reverse transcriptase on RNA and DNA templates. *Biochim. Biophys. Acta* **1131**:270-280.
- DeStefano, J. J., L. M. Mallaber, L. Rodriguez-Rodriguez, P. J. Fay, and R. A. Bambara. 1992. Requirements for strand transfer between internal regions of heteropolymer templates by human immunodeficiency virus reverse transcriptase. *J. Virol.* **66**:6370-6378.
- Gritz, L., and J. Davies. 1983. Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*. *Gene* **25**:179-88.
- Higuchi, R. 1989. Simple and rapid preparation of samples for PCR, p. 31-38. *In* H. A. Erlich (ed.), *PCR technology: principles and applications for DNA amplification*. Stockton Press, New York, N.Y.
- Hu, W. S., E. H. Bowman, K. A. Delviks, and V. K. Pathak. 1997. Homologous recombination occurs in a distinct retroviral subpopulation and exhibits high negative interference. *J. Virol.* **71**:6028-6036.
- Hu, W. S., and H. M. Temin. 1990. Genetic consequences of packaging two

- RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc. Natl. Acad. Sci. USA* **87**:1556–1560.
17. **Hughes, S., and E. Kosik.** 1984. Mutagenesis of the region between env and src of the SR-A strain of Rous sarcoma virus for the purpose of constructing helper-independent vectors. *Virology* **136**:89–99.
  18. **Jang, S. K., M. V. Davies, R. J. Kaufman, and E. Wimmer.** 1989. Initiation of protein synthesis by internal entry of ribosomes into the 5' nontranslated region of encephalomyocarditis virus RNA in vivo. *J. Virol.* **63**:1651–1660.
  19. **Jang, S. K., H. G. Krausslich, M. J. Nicklin, G. M. Duke, A. C. Palmenberg, and E. Wimmer.** 1988. A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J. Virol.* **62**:2636–2643.
  20. **Jones, J. S., R. W. Allan, B. Seuffer, and H. M. Temin.** 1994. Copackaging of different-sized retroviral genomic RNAs: little effect on retroviral replication or recombination. *J. Virol.* **68**:4097–4103.
  21. **Jorgensen, R. A., S. J. Rothstein, and W. S. Reznikoff.** 1979. A restriction enzyme cleavage map of Tn5 and location of a region encoding neomycin resistance. *Mol. Gen. Genet.* **177**:65–72.
  22. **Julias, J. G., D. Hash, and V. K. Pathak.** 1995. E<sup>-</sup> vectors: development of novel self-inactivating and self-activating retroviral vectors for safer gene therapy. *J. Virol.* **69**:6839–6846.
  23. **Katz, R. A., and A. M. Skalka.** 1990. Generation of diversity in retroviruses. *Annu. Rev. Genet.* **24**:409–445.
  24. **Kent, R. B., J. R. Emanuel, Y. Ben Neriah, R. Levenson, and D. E. Housman.** 1987. Ouabain resistance conferred by expression of the cDNA for a murine Na<sup>+</sup>, K<sup>+</sup>-ATPase alpha subunit. *Science* **237**:901–903.
  25. **Kim, J. K., C. A. Palaniappan, W. Wu, P. J. Fay, and R. A. Bambara.** 1997. Evidence for a unique mechanism of strand transfer from the transactivation response region of HIV-1. *J. Biol. Chem.* **272**:16769–16777.
  26. **Klarmann, G. J., C. A. Schaubert, and B. D. Preston.** 1993. Template-directed pausing of DNA synthesis by HIV-1 reverse transcriptase during polymerization of HIV-1 sequences in vitro. *J. Biol. Chem.* **268**:9793–9802. (Erratum, 268:13764.)
  27. **Mikkelsen, J. G., A. H. Lund, M. Duch, and F. S. Pedersen.** 1998. Recombination in the 5' leader of murine leukemia virus is accurate and influenced by sequence identity with a strong bias toward the kissing-loop dimerization region. *J. Virol.* **72**:6967–6978.
  28. **Mikkelsen, J. G., A. H. Lund, K. Dybkaer, M. Duch, and F. S. Pedersen.** 1998. Extended minus-strand DNA as template for R-U5-mediated second-strand transfer in recombinational rescue of primer binding site-modified retroviral vectors. *J. Virol.* **72**:2519–2525.
  29. **Mikkelsen, J. G., A. H. Lund, K. D. Kristensen, M. Duch, M. S. Sorensen, P. Jorgensen, and F. S. Pedersen.** 1996. A preferred region for recombinational patch repair in the 5' untranslated region of primer binding site-impaired murine leukemia virus vectors. *J. Virol.* **70**:1439–1447.
  30. **Miller, A. D., J. V. Garcia, N. von Suhr, C. M. Lynch, C. Wilson, and M. V. Eiden.** 1991. Construction and properties of retrovirus packaging cells based on gibbon ape leukemia virus. *J. Virol.* **65**:2220–2224.
  31. **Omer, C. A., K. Pogue-Geile, R. Guntaka, K. A. Staskus, and A. J. Faras.** 1983. Involvement of directly repeated sequences in the generation of deletions of the avian sarcoma virus *src* gene. *J. Virol.* **47**:380–382.
  32. **Palaniappan, C., M. Wisniewski, W. Wu, P. J. Fay, and R. A. Bambara.** 1996. Misincorporation by HIV-1 reverse transcriptase promotes recombination via strand transfer synthesis. *J. Biol. Chem.* **271**:22331–22338.
  33. **Pathak, V. K., and H. M. Temin.** 1990. Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: deletions and deletions with insertions. *Proc. Natl. Acad. Sci. USA* **87**:6024–6028.
  34. **Pathak, V. K., and H. M. Temin.** 1990. Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. *Proc. Natl. Acad. Sci. USA* **87**:6019–6023.
  35. **Pathak, V. K., and W.-S. Hu.** 1997. "Might as well jump!" Template switching by retroviral reverse transcriptase, defective genome formation, and recombination. *Semin. Virol.* **8**:141–150.
  36. **Peliska, J. A., S. Balasubramanian, D. P. Giedroc, and S. J. Benkovic.** 1994. Recombinant HIV-1 nucleocapsid protein accelerates HIV-1 reverse transcriptase catalyzed DNA strand transfer reactions and modulates RNase H activity. *Biochemistry* **33**:13817–13823.
  37. **Peliska, J. A., and S. J. Benkovic.** 1992. Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. *Science* **258**:1112–1118.
  38. **Sambrook, J. E., E. F. Fritsch, and T. Maniatis.** 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
  39. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukaemia virus. *Nature* **293**:543–548.
  40. **Suo, Z., and K. A. Johnson.** 1998. DNA secondary structure effects on DNA synthesis catalyzed by HIV-1 reverse transcriptase. *J. Biol. Chem.* **273**:27259–67.
  41. **Suo, Z., and K. A. Johnson.** 1997. Effect of RNA secondary structure on the kinetics of DNA synthesis catalyzed by HIV-1 reverse transcriptase. *Biochemistry* **36**:12459–12467.
  42. **Suo, Z., and K. A. Johnson.** 1997. RNA secondary structure switching during DNA synthesis catalyzed by HIV-1 reverse transcriptase. *Biochemistry* **36**:14778–14785.
  43. **Temin, H. M.** 1993. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proc. Natl. Acad. Sci. USA* **90**:6900–6903.
  44. **Temin, H. M.** 1991. Sex and recombination in retroviruses. *Trends Genet.* **7**:71–74.
  45. **Temin, H. M., and S. Mizutani.** 1970. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**:1211–1213.
  46. **Thilly, W. G.** 1995. Specificity, efficiency and fidelity of PCR, p. 37–51. *In* C. W. Dieffenbach and G. A. Dveksler (ed.), *PCR primer: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
  47. **Tikhonenko, A. T., and M. L. Linial.** 1993. Transforming variants of the avian *myc*-containing retrovirus FH3 arise prior to phenotypic selection. *J. Virol.* **67**:3635–3638.
  48. **Topping, R., M. A. Demoitie, N. H. Shin, and A. Telesnitsky.** 1998. Cis-acting elements required for strong stop acceptor template selection during Moloney murine leukemia virus reverse transcription. *J. Mol. Biol.* **281**:1–15.
  49. **Varela-Echavarría, A., C. M. Prorock, Y. Ron, and J. P. Dougherty.** 1993. High rate of genetic rearrangement during replication of a Moloney murine leukemia virus-based vector. *J. Virol.* **67**:6357–6364.
  50. **Wu, W., B. M. Blumberg, P. J. Fay, and R. A. Bambara.** 1995. Strand transfer mediated by human immunodeficiency virus reverse transcriptase in vitro is promoted by pausing and results in misincorporation. *J. Biol. Chem.* **270**:325–332.
  51. **Wu, W., L. E. Henderson, T. D. Copeland, R. J. Gorelick, W. J. Bosche, A. Rein, and J. G. Levin.** 1996. Human immunodeficiency virus type 1 nucleocapsid protein reduces reverse transcriptase pausing at a secondary structure near the murine leukemia virus polypurine tract. *J. Virol.* **70**:7132–7142.
  52. **You, J. C., and C. S. McHenry.** 1994. Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription. *J. Biol. Chem.* **269**:31491–31495.
  53. **Yu, H., A. E. Jetzt, Y. Ron, B. D. Preston, and J. P. Dougherty.** 1998. The nature of human immunodeficiency virus type 1 strand transfers. *J. Biol. Chem.* **273**:28384–28391.
  54. **Zhang, J., and H. M. Temin.** 1993. Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication. *Science* **259**:234–238.
  55. **Zhang, J., and H. M. Temin.** 1994. Retrovirus recombination depends on the length of sequence identity and is not error prone. *J. Virol.* **68**:2409–2474.
  56. **Zuker, M.** 1994. Prediction of RNA secondary structure by energy minimization. *Methods Mol. Biol.* **25**:267–294.