

Effects of Homology Length in the Repeat Region on Minus-Strand DNA Transfer and Retroviral Replication

QUE DANG^{1,2} AND WEI-SHAU HU^{2*}

Department of Microbiology and Immunology, School of Medicine, West Virginia University, Morgantown, West Virginia 26506,¹ and HIV Drug Resistance Program, National Cancer Institute, Frederick Cancer Research and Development Center, Frederick, Maryland 21702²

Received 13 July 2000/Accepted 25 October 2000

Homology between the two repeat (R) regions in the retroviral genome mediates minus-strand DNA transfer during reverse transcription. We sought to define the effects of R homology lengths on minus-strand DNA transfer. We generated five murine leukemia virus (MLV)-based vectors that contained identical sequences but different lengths of the 3' R (3, 6, 12, 24 and 69 nucleotides [nt]); 69 nt is the full-length MLV R. After one round of replication, viral titers from the vector with a full-length downstream R were compared with viral titers generated from the other four vectors with reduced R lengths. Viral titers generated from vectors with R lengths reduced to one-third (24 nt) or one-sixth (12 nt) that of the wild type were not significantly affected; however, viral titers generated from vectors with only 3- or 6-nt homology in the R region were significantly lower. Because expression and packaging of the RNA were similar among all the vectors, the differences in the viral titers most likely reflected the impact of the homology lengths on the efficiency of minus-strand DNA transfer. The molecular nature of minus-strand DNA transfer was characterized in 63 proviruses. Precise R-to-R transfer was observed in most proviruses generated from vectors with 12-, 24-, or 69-nt homology in R, whereas aberrant transfers were predominantly used to generate proviruses from vectors with 3- or 6-nt homology. Reverse transcription using RNA transcribed from an upstream promoter, termed read-in RNA transcripts, resulted in most of the aberrant transfers. These data demonstrate that minus-strand DNA transfer is homology driven and a minimum homology length is required for accurate and efficient minus-strand DNA transfer.

Retroviruses are RNA viruses that replicate through a DNA intermediate (49). Most retroviral particles contain viral RNA; upon infection of the target cells, viral RNA is copied into DNA by the viral enzyme reverse transcriptase (RT) (2, 27, 51) and then integrates into the target cell genome to form a provirus (49). Host cell RNA polymerase II transcribes the provirus to generate viral RNA transcripts; the full-length viral RNA is packaged into viral particles to serve as genetic material for the next round of viral infection (7).

Retroviruses have evolved to adapt to the dual phase of the life cycle. Two of the adapted features are the genome structure and the mechanism by which the viral DNA is synthesized. The proviral genome contains two long terminal repeats (LTRs), one at each end of the viral DNA sequences (7, 50). The LTR is composed of three sections, unique 3' (U3), repeat (R), and unique 5' (U5) regions; each plays important roles during the viral life cycle (7). The U3 region contains the promoter from which viral RNA transcripts are expressed (7, 36). R and U5 are both important in the process of reverse transcription of the viral RNA into DNA, U5 for the initiation and R for the extension of viral DNA synthesis (7, 26, 48). The viral RNA transcript, which includes sequences from the upstream R to the end of downstream R, is generated by the host cell RNA polymerase II and is shorter than the provirus (7, 36, 50). It is advantageous for the viral DNA to contain an active

promoter at the 5' end of the sequences to ensure active RNA transcription; however, the upstream U3 is missing from the viral RNA, which is the template for viral DNA synthesis. Therefore, retroviruses undergo two strand transfer steps to first regenerate the LTR by joining the U3 and R-U5 sequences and then to duplicate the LTR and place it at both ends of the viral DNA.

Reverse transcription is initiated near the 5' end of the viral RNA by using a tRNA primer that has hybridized to the primer binding site (PBS) adjacent to the U5 sequence (7, 11). Because viral RNA is plus-sense, the first strand of DNA synthesized is referred to as minus-strand DNA. Minus-strand DNA synthesis copies U5 and R, in a step termed minus-strand DNA transfer, and reverse transcription switches to using the 3' viral RNA as a template and continues DNA synthesis (7, 11, 48). This step joins together the U3, R, and U5 sequences. Minus-strand DNA synthesis continues to copy the RNA template, including the U3 and adjacent polypurine tract (PPT) sequences. Once minus-strand DNA synthesis passes the PPT, RT makes a specific cut in the RNA strand at the 3' end of the PPT and uses the RNA as a primer for plus-strand DNA synthesis (5), which copies U3, R, U5, and the first 18 nucleotides (nt) of the tRNA sequences (7). This plus-strand DNA is then transferred to the 5' end of the viral DNA to duplicate the LTR. Once DNA synthesis is completed, the resulting viral DNA has two LTRs, one on each end of the viral genome (11).

Minus-strand DNA transfer joins U3 with R-U5 to regenerate the LTR during DNA synthesis. Although it is a critical step of reverse transcription, the exact mechanism of minus-strand DNA transfer remains unclear. The current hypothesis

* Corresponding author. Mailing address: HIV Drug Resistance Program, National Cancer Institute, FCRDC, Room 336, Building 535, Frederick, MD 21702. Phone: (301) 846-1250. Fax: (301) 846-6013. E-mail: whu@ncicrf.gov.

is that minus-strand DNA synthesis copies U5 and R to form minus-strand strong-stop DNA and that the RNase H function of RT degrades the RNA template in the DNA-RNA hybrid (5, 47, 52). The single-stranded minus-strand strong-stop DNA contains an R region that is complementary to the R region at the 3' viral RNA. Using this complementarity, the nascent DNA can hybridize to the 3' viral RNA, and the reverse transcription complex can continue copying the U3 sequences (7).

This hypothesis predicts that minus-strand DNA transfer depends on the hybridization of the two R regions; thus, the RNase H activity of RT and the homology in the R region are critical for minus-strand DNA transfer. It has been shown that abolishing the RNase H activity of RT prevents minus-strand DNA transfer (5, 47), which supports this hypothesis. However, little is known about the requirement of R homology during minus-strand DNA transfer.

The R regions of different retroviruses vary in size, from 15 nt in mouse mammary tumor virus (MMTV) to 247 nt in human T-cell leukemia virus type 2 (HTLV-2) (33, 42). Premature minus-strand DNA transfer has been observed in murine leukemia virus (MLV), spleen necrosis virus (SNV), and human immunodeficiency virus type 1 (HIV-1) at low frequencies, with R regions of 69, 82, and 97 nt, respectively (24, 25, 28, 37, 55). These events indicated that partial R sequences are able to mediate strand transfer in these viruses. The effect of R homology length in HIV-1 was examined by truncation of the 3' R; delayed viral replication kinetics were observed when the 3' R was shortened to 30 or 15 nt (3). The effect of removing R homology was examined by using a chimeric vector containing R regions from MLV and SNV that did not have significant homology (55). The chimeric vector replicated at a reduced titer compared with its counterpart containing two highly homologous R sequences; molecular analyses indicated that most of the minus-strand DNA transfer events were mediated by short stretches of homologous sequences. Taken together, these data indicated that most retroviruses probably could use a shorter R sequence to mediate minus-strand DNA transfer. However, the effect of R homology length on the efficiency of minus-strand DNA transfer is unclear.

In this report, we explored the impact of R homology length on the efficiency and accuracy of minus-strand DNA transfer. Using MLV-derived vectors and a system that allowed only one round of viral replication, we examined the ability of 3, 6, 12, 24, and 69 nt of R homology to mediate minus-strand DNA transfer. The molecular nature of the strand transfer and transfer junctions was also analyzed.

MATERIALS AND METHODS

Plasmid construction and definition. Retroviral vectors were constructed using standard cloning techniques (40). All primer and linker sequences are available upon request. In this report, the p in the vector name (e.g., pMMQD3) refers to the plasmid, whereas the vector name without the p (e.g., MMQD3) refers to virus derived from this plasmid.

Briefly, pLXSN (32) was digested with *Hind*III (located downstream of the simian virus 40 [SV40] promoter), treated with Klenow fragment of the DNA polymerase, digested with *Xba*I (located in the downstream U3 region), and then treated with calf intestinal alkaline phosphatase to produce a 4.6-kb DNA fragment containing the SV40 promoter and the plasmid backbone. Plasmid pAR2 (55) was digested with *Sal*I, located upstream of the hygromycin phosphotransferase B gene (*hygro*) (12), treated with Klenow fragment, and then digested with *Xba*I, located in the downstream U3 region. This procedure generated a 1.4-kb DNA fragment containing *hygro*. Ligation of the *hygro* DNA fragment to the

digested pLXSN backbone generated pMSM2, an MLV-based vector plasmid containing the SV40 promoter and *hygro*. Next, pMSM2 was digested with *Clal* plus *Nde*I and treated with calf intestinal alkaline phosphatase to excise the downstream LTR. To construct pMSM3, a linker containing the MLV PPT, the attachment site (*att*), and a *Bgl*II site was ligated to the digested pMSM2. A 101-bp linker containing the SNV R region (82 nt) was inserted into the unique *Bgl*II restriction site in pMSM3 to generate pMSM4. Various linkers generated from annealed synthesized oligonucleotides were inserted into the *Bgl*II and *Mlu*I sites of pMSM4 to generate pMSM6, pMSM12, pMSM24, and pMSM69. These linkers contained different lengths of MLV R sequences; each linker also contained a unique restriction enzyme site marker. These restriction enzyme sites were *Bst*1107, *Hind*III, *Eco*RI, and *Clal*, for pMSM6, pMSM12, pMSM24, and pMSM69, respectively. It was later found that SNV R sequences do not provide an adequate signal for the cleavage of RNA transcription in murine cells. To introduce a functional polyadenylation signal for these constructs, a 0.4-kb DNA fragment containing the SV40 polyadenylation signal was inserted into these plasmids. The SV40 DNA fragment was isolated from pJD220svhygro (8), which had been digested with *Clal* and *Bst*1107, followed by treatment with Klenow fragment, and was inserted into plasmids digested with *Sal*I and treated with Klenow fragment. This insertion generated pMMQD3, pMMQD6, pMMQD12, pMMQD24, and pMMQD69 from pMSM4, pMSM6, pMSM12, pMSM24, and pMSM69, respectively.

Cells, transfections, and virus propagations. D17 and PG13 cells were obtained from the American Type Culture Collection. D17 is a dog osteosarcoma cell line that is permissive to infection by MLV (38). PG13 is a murine cell line that expresses MLV *gag-pol* and gibbon ape leukemia virus *env* (31). Both cells were grown in Dulbecco's modified Eagle's medium supplemented with 6% (D17) or 10% (PG13) calf serum (HyClone Laboratories, Inc.). Penicillin (50 U/ml, Gibco) and streptomycin (50 µg/ml, Gibco) were also added to the medium. Cells were maintained at 37°C with 5% CO₂. Hygromycin (Calbiochem) selection was performed at 120 µg/ml for D17 cells and 240 µg/ml for PG13 cells.

Viral vectors were transfected into PG13 cells by the calcium phosphate precipitation method (40). Transfected PG13 cells were pooled, expanded, and plated at a density of 5×10^6 cells per 100-mm-diameter dish. Fresh medium was added to the cells 24 h before the virus was harvested. Virus-containing cell culture medium was centrifuged at $6,000 \times g$ for 10 min to pellet cellular debris. The supernatant was then aliquoted for use in the infections, RT assay, and isolation of cell-free virion RNA.

Tenfold serial dilutions of each virus were used to infect D17 cells. Viral infections were performed in the presence of Polybrene (50 µg/ml) for 4 h at 37°C with 5% CO₂. The infected cells were subjected to hygromycin selection, and viral titers were determined by the number of hygromycin-resistant D17 cell colonies. Individual D17 cell clones containing proviruses were isolated for analysis of minus-strand DNA transfer junctions.

RNA isolation and analysis. PG13 cells were plated at a density of 5×10^6 cells per 100-mm-diameter dish. Cellular RNA was isolated 48 h later using Trizol (Gibco/BRL) as specified by the manufacturer. The integrity of the cellular RNA was examined by gel electrophoresis and the inspection of the rRNA bands.

To obtain cell-free virion RNA, viruses were harvested as described above and concentrated by centrifugation at 25,000 rpm for 90 min in a Beckman SW28 rotor. Viral pellets were resuspended in diethyl pyrocarbonate-treated water. This preparation was divided into two aliquots, one for the RT assay (described below) and the other for the RNA analysis. To monitor for any RNA loss during the isolation procedure, an aliquot of SNV (CG4) (10) was added as an internal control to the viral preparation used for RNA analysis. Sodium dodecyl sulfate (0.1%, final concentration) and tRNA (200 µg/ml, final concentration) were added to lyse the mixture of viruses. Phenol-chloroform extractions were performed and the viral RNA was precipitated with ethanol by standard methods. Viral RNA was resuspended in diethyl pyrocarbonate-treated water. Alternatively, cell-free virion RNA was isolated using the QIAamp viral RNA kit (Qiagen).

RNase protection assay. RNase protection assay analyses were performed on the cellular and cell-free virion RNA as specified by the manufacturer (Ambion); these results were quantified using the ImageQuant program on a PhosphorImager (Molecular Dynamics).

RT assay. Viruses were harvested and concentrated as described above. RT assays were performed using standard procedures (2, 13, 51). Briefly, an aliquot of each virus was added to a reaction mixture containing 50 mM Tris (pH 8.0), 0.6 mM MnCl₂, 60 mM NaCl, 0.5% IgePal CA-630 (Sigma), 1 U of RNasin (Boehringer Mannheim Biochemicals) per µl, 0.05 µg of oligo(dT)/µl, 0.1 µg of poly(A)/µl, 80 µM dTTP, 10 mM dithiothreitol, and 10 µCi of [³H]dTTP (72 Ci/mmol; ICN). The reaction mixture was incubated at 37°C for 1 h and then precipitated with 1 ml of 10% ice-cold trichloroacetic acid for 1 h on ice. The

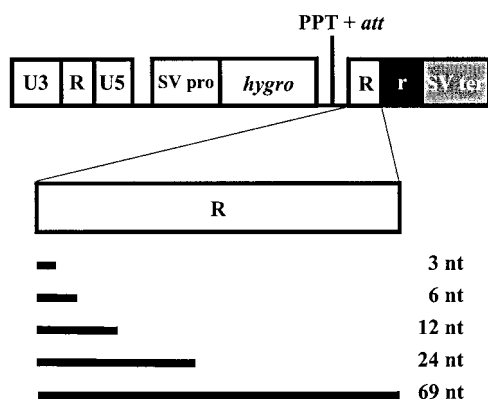


FIG. 1. Structures of the MLV-based vectors used to examine the effects of different homology lengths in the R region on minus-strand DNA transfer. R, MLV R; SV pro, SV40 promoter; *hygro*, hygromycin phosphotransferase B gene; PPT, polypurine tract; *att*, attachment site; r in black box, SNV R region; SV ter, SV40 polyadenylation signal. Various vectors contained different lengths of the downstream MLV R sequences as indicated.

viral mixture was filtered through a 0.45- μ m-pore-size GN6 Metrical membrane (Gelman Sciences) and washed twice with 10% trichloroacetic acid. Radioactivity incorporated into the newly synthesized DNA by RT was measured using a scintillation counter.

PCR and DNA sequencing. Hygromycin-resistant D17 cell clones were isolated. Each cell clone was expanded and then lysed (16). The lysate was incubated at 60°C for 1 h and then at 95°C for 10 min. Primers located at the 3' end of *hygro* and in U5 were used to amplify a segment of proviral DNA by PCR. The U5 primer was biotinylated and was used to isolate the PCR product using Streptavidin magnetic beads (DynaL Inc.) as specified by the manufacturer. After denaturation of the DNA, the biotinylated single-stranded DNA was collected for direct sequencing using an AutoRead kit (Pharmacia). Alternatively, PCR products were sequenced directly using a BigDye termination cycle sequencing ready reaction kit (PE Applied Biosystems).

RESULTS

Retroviral vectors used to study minus-strand DNA transfer during reverse transcription. A series of five vectors was utilized to examine the effects of R homology length on minus-strand DNA transfer; the vector structures are illustrated in Fig. 1. These vectors were derived from the pLN series of plasmids (32) and contained *cis*-acting elements from Moloney murine leukemia virus (MoMLV) and/or Moloney murine sarcoma virus, a virus derived from MoMLV. The *cis*-acting sequences from MoMLV and Moloney murine sarcoma virus contain high homology; for simplicity, these sequences are referred to as MLV sequences in this report. The five vectors contain identical sequences except for the downstream R regions. Each vector contains the SV40 promoter, *hygro* (12), and MLV-derived *cis*-acting elements, such as the upstream LTR, PBS, extended packaging signal (Ψ^+), PPT, and *att*. The downstream LTR of these vectors was replaced with the 3' *att*, a variable stretch of homology to MLV R, SNV R, and the polyadenylation signal from SV40. The SNV R and SV40 polyadenylation signals were added to replace the signal located in MLV R. SNV R was first inserted with the intention of using the polyadenylation signal within these sequences. It was later found that SNV R was not sufficient in providing the polyadenylation signal within this sequence context in murine cells (data not shown). A DNA fragment containing the SV40 poly-

adenylation signal was then inserted into these plasmids. It had been shown previously that the function of the SV40 polyadenylation signal was not affected by SNV R when the two sequences were present in tandem (18–20).

Different lengths of the MLV R sequences were inserted at the 3' end of the viral genome within the vectors pMMQD3, pMMQD6, pMMQD12, pMMQD24, and pMMQD69, which contained 3, 6, 12, 24, and 69 nt of the downstream R sequences, respectively. All of the various lengths inserted at the downstream R started with the 5' end of the R; for example, pMMQD3 contained the first 3 nt of R. The viral RNAs generated from these vectors would have complete R and U5 regions at the upstream sequences, and the downstream sequences would contain 5' portions of R or the entire R region. During reverse transcription, RT used a tRNA primer and copied U5 and R to form minus-strand strong-stop DNA. The 3' end of the nascent DNA was complementary to the downstream R sequences in the viral RNA; the length of this complementarity varied from 3 to 69 nt. If the length of R between the nascent DNA and RNA was sufficient to mediate minus-strand DNA transfer, then it was expected that the virus would replicate efficiently and most of the minus-strand DNA transfer would be from R to R (R-to-R transfer). In contrast, if the length of R complementarity was not sufficient to mediate minus-strand DNA transfer, then it was expected that the virus would have reduced replication efficiency and the minus-strand DNA transfer events would be aberrant in nature.

MMQD69 viral RNA had two full-length MLV R regions, mimicking the structure of the wild-type viral RNA, and thus served as a positive control for the viral replication and minus-strand DNA transfer efficiency. The full-length R should mediate efficient and precise R-to-R minus-strand DNA transfer. The RNA transcripts from all the other vectors would terminate using the polyadenylation signal in SV40 sequences. To ensure that MMQD69 had the same 3' sequences in its RNA transcripts as the transcripts from the other vectors, the AAT AAA sequences in the 3' MLV R were changed to AAGGAA, which had been previously demonstrated to abolish the polyadenylation signal (15, 44, 45). The change of 2 nt in these sequences was not expected to affect minus-strand DNA transfer for the following reasons. First, these 2 nt were located far from the transfer junction, 49 and 50 nt away from the 5' end of R. Second, vectors with two R regions containing few mismatches in the internal regions have been reported to replicate efficiently (32, 55).

The viral RNA of MMQD24 had 24 nt of homology in the R region; this was approximately one-third the homology length in wild-type MLV. The viral RNA of MMQD12 had 12-nt homology in the R region; this was close to the shortest naturally occurring R in MMTV (15 nt) (7, 33). It had been previously observed that 6-nt homology could be used to mediate minus-strand DNA transfer, although with unknown efficiency (55). The viral RNA of MMQD6 had 6-nt homology between the two R regions; this vector allowed us to test the efficiency of minus-strand DNA transfer mediated by 6-nt homology. The viral RNA of MMQD3 had 3-nt homology between the two R regions; this vector was used to examine whether homology shorter than 6 nt could be used to mediate minus-strand DNA transfer.

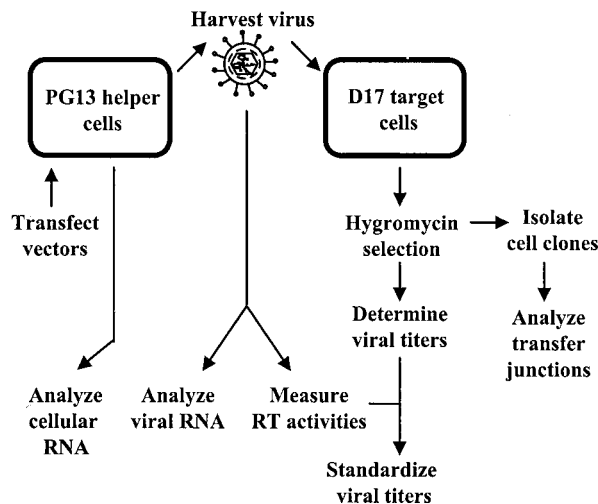


FIG. 2. Experimental protocol used to study minus-strand DNA transfer in one round of retroviral replication.

Experimental protocol used to examine the effect of R homology length on minus-strand DNA transfer. The protocol used in this study is outlined in Fig. 2. Each retroviral vector was separately transfected into PG13 helper cells and the transfected cells were subjected to hygromycin selection. Hygromycin-resistant cells transfected with each vector were separately pooled; all of the cell pools used in these experiments contained at least 600 colonies. From the viruses harvested from these transfected cell pools, one portion was used for the measurement of RT activity, a second portion was used for the isolation of cell-free virion RNA, and the third portion was used to infect D17 target cells. Infected D17 cells were subjected to hygromycin selection, and the number of resistant colonies was used to determine the viral titer. Viral titers were standardized to RT activity. In addition, hygromycin-resistant D17 target cell clones were isolated. Portions of the proviral genomes were amplified by PCR and characterized by DNA sequencing to determine the molecular nature of minus-strand DNA transfer.

Cellular RNA and cell-free virion RNA were isolated from the transfected PG13 cells and the viruses harvested from these cells, respectively. The integrity and concentration of the cellular RNA were determined by visualization of the rRNA bands in agarose gels and by spectrophotometer readings, respectively. To monitor the recovery of virion RNA isolation, an aliquot of wild-type SNV was added to the cell-free virions prior to the concentration of the virus particles by ultracentrifugation as an internal control. The amounts of vector RNA in the transfected cells and released virions were analyzed by RNase protection assay.

Effect of R homology length on viral replication. Three independent sets of transfection and infection experiments were performed using different DNA sets for each experiment. Viral titers were determined and were standardized to the RT activity. A summary of the standardized viral titer is shown in Fig. 3. The positive control of the experiment, MMQD69, had two full-length copies of MLV R; this vector generated titers with a mean of 4.3×10^2 CFU/ml. MMQD24 and MMQD12 generated titers with means of 3.3×10^2 and 2.7×10^2 CFU/ml,

respectively. The titers generated by these three vectors were not significantly different ($P = 0.46$ for MMQD69 and MMQD24, $P = 0.27$ for MMQD69 and MMQD12, and $P = 0.559$ for MMQD24 and MMQD12; two-sample t test). This similarity in the viral titers indicated that the ability of the virus to replicate was not significantly affected by reducing the R homology from 69 nt to 24 or 12 nt, approximately one-third or one-sixth of the length of wild-type R.

In contrast, MMQD6 generated titers with a mean of 0.34×10^2 CFU/ml; these titers were approximately 13-fold lower than those generated by MMQD69. The differences between the titers of MMQD6 and those from the three vectors with longer R homology were statistically significant ($P < 0.03$, $P < 0.02$, and $P < 0.02$ when MMQD6 was compared with MMQD69, MMQD24, and MMQD12, respectively; two-sample t test). MMQD3 generated titers with an average of 0.29×10^2 CFU/ml. These titers were also significantly different from those of the first three vectors ($P < 0.03$, $P < 0.01$, and $P < 0.02$ when MMQD3 was compared with MMQD69, MMQD24, and MMQD12, respectively; two-sample t test). The titers generated by MMQD6 and MMQD3 were not significantly different from each other ($P = 0.58$; two-sample t test).

These data indicated that a minimum length of R homology was needed for the retrovirus to mediate minus-strand DNA transfer to allow efficient viral replication. When the viral genome had more than 12 nt of homology in R, the viral titers remained similar to those from virus containing 69 nt of homology. In contrast, when the R homology length was reduced to 6 or 3 nt, the viral titers dropped significantly. The alteration in the viral titers among these constructs most likely reflected the efficiencies of minus-strand DNA transfer during reverse transcription. However, we could not rule out the possibility that the differences among the viral titers in these constructs were caused by other factors, such as the amounts of RNA in

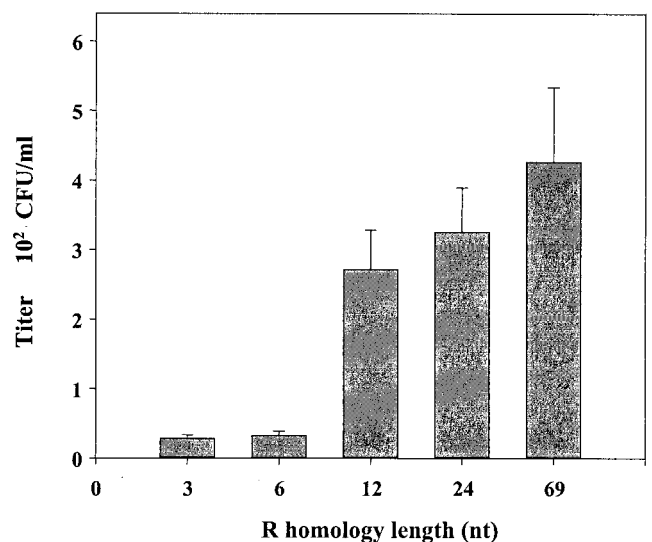


FIG. 3. Standardized viral titer. The number of hygromycin-resistant colonies was used to calculate the viral titer. The RT activity of each viral sample used for infection was measured and utilized to standardize the viral titer. Error bars represent the standard errors of the means from three independent experiments.

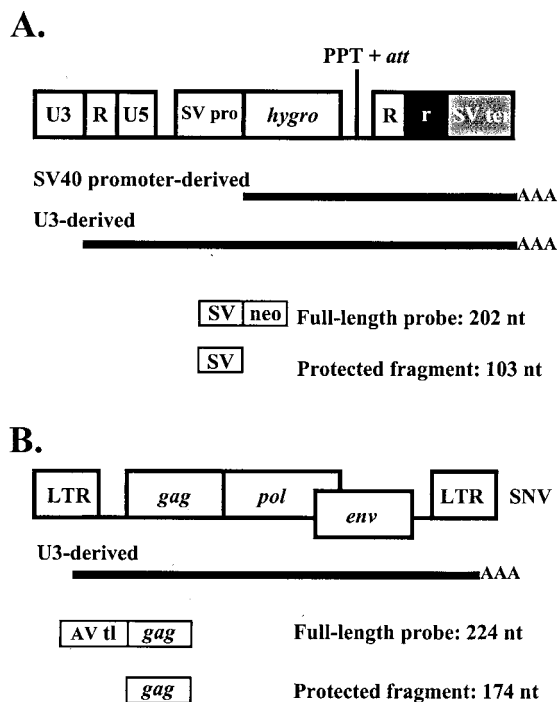


FIG. 4. Strategy and probes used to detect vector RNA levels in transfected cells and cell-free virions by RNase protection assays. All abbreviations are the same as in Fig. 1. (A) Detection of vector RNA. Two types of transcripts were expected to be generated from the vector: those derived from the SV40 promoter and those derived from the upstream U3. The SV-*neo* probe, generated from a DNA fragment amplified from pLXSN, was used to detect U3-derived transcripts. This probe contained sequences from both the SV40 promoter and the neomycin phosphotransferase gene (*neo*). Only the SV40 promoter sequences were protected in the RNase protection assay. (B) Detection of internal control RNA. The probe was generated from an amplified DNA fragment, using pRD136 as the template, and contained a portion of adenovirus tripartite leader sequence (AV tl) and a portion of SNV *gag*. Only the *gag* portion of the probe was protected in the RNase protection assay.

the transfected cells or the viral RNA packaged in the virions. To address these possibilities, we examined the amount of cellular RNA and cell-free virion RNA.

Similar expression and packaging of vector RNA among all constructs. To examine the level of viral RNA expression in cells transfected with various constructs, total cellular RNAs were isolated from transfected cell pools. Equal amounts of RNA isolated from each cell pool were used in RNase protection assays to determine the amount of virus-specific RNA. The strategy and probes used in these analyses are shown in Fig. 4. Two types of RNA transcripts were expected to be generated from these constructs, one from the U3 promoter and the other from the SV40 promoter (Fig. 4A). Only the U3-derived transcripts contained Ψ^+ and were expected to be packaged efficiently for viral replication; therefore, the amounts of U3-derived transcripts were measured. U3-derived RNA transcripts should also contain SV40 promoter sequences, which were lacking in SV40 promoter-derived RNA transcripts. A 202-nt RNA probe containing the 3' end of the SV40 promoter and the 5' end of the neomycin phosphotransferase gene (*neo*) (21) was generated by in vitro transcription

using a DNA fragment amplified from pLXSN (32) as a template. Viral transcripts from all constructs driven by the U3 promoter should contain the SV40 promoter sequences but not *neo* (Fig. 4A). Therefore, it is expected that the SV40 promoter sequences but not the *neo* sequences should be protected during the RNase digestion and generate a 103-nt protected fragment. A representative RNase protection assay of the cellular RNA from transfected cell pools is shown in Fig. 5A. The levels of U3-derived viral RNA expression in cells transfected with different constructs were similar (Fig. 5A and data not shown).

To quantify the amount of RNA in the cell-free virions released by the transfected cells, viruses were harvested from these transfected cells. Prior to isolation of the viral RNA, each virus sample was mixed with an aliquot of SNV, which served as an internal control to monitor the loss of the viral RNAs during the isolation procedure and RNase protection assay. The amount of cell-free virion RNA from each sample was measured by RNase protection assay with the same SV40-*neo* probe described for the cellular RNA analyses. The amount of the SNV RNA (internal control) in each sample was measured by RNase protection assay using a 224-nt RNA probe generated from pRD136 (30), which partially hybridized to SNV *gag* (Fig. 4B) to produce a 174-nt fragment. SNV and MLV are distantly related viruses; they do not contain significant homology in *gag* at the nucleotide sequence level. Representative RNase protection assay analyses of vector viral RNA and internal control SNV RNA are shown in Fig. 5B and C, respectively. These data showed that the cell pools transfected with different vectors produced similar amounts of cell-free virion RNA (Fig. 5B and C and data not shown).

These experiments demonstrated that cells transfected with different vectors had similar RNA expression and that the vector RNAs were packaged and released at similar levels. Therefore, the differences in the viral titers generated by these constructs were not caused by RNA expression or packaging but most likely reflected the impact of the R homology lengths on minus-strand DNA transfer.

Characterization of the molecular nature of minus-strand DNA transfer events using different lengths of R homology. To study the molecular nature of minus-strand DNA transfer events, individual cell clones containing proviruses were isolated from five independent infection experiments and from separate cell culture dishes. These cell clones were generated with low multiplicities of infection (<0.001) such that most of the cells should only contain one provirus (probability of double infection, <0.00001). DNA lysates were generated from these cell clones and used to amplify a portion of the proviral structures. If minus-strand DNA transfer occurred from the R of the minus-strand DNA to the R of the 3' viral RNA (R-to-R transfer), the resulting viral DNA should contain *hygro*, PPT, *att*, R, and U5 at the 3' portion of the genome (Fig. 6A). Using primers near the 3' end of *hygro* and in U5, the PCR-amplified DNA fragment should be 0.45 kb in length (Fig. 6A). These primers should also be able to amplify proviral genomes that did not use R-to-R transfer. Reverse transcription of all of these vectors should be initiated from the PBS and copy U5 and then R prior to strand transfer. Therefore, regardless of the acceptor template used for minus-strand DNA transfer, all of these proviruses should have U5 sequences at the 3' end of

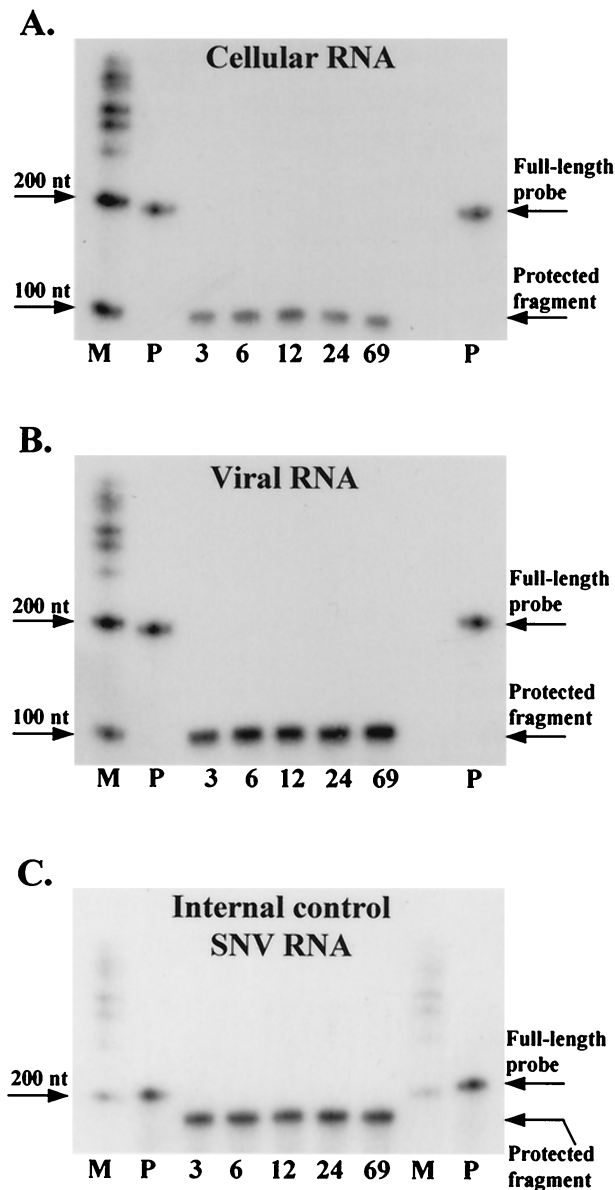


FIG. 5. RNase protection assay analyses of the expression and packaging of vector RNAs. Lanes: M, RNA molecular weight markers; P, full-length probe; 3, 6, 12, 24, and 69, samples from cells or viral samples derived from vectors pMMQD3, pMMQD6, pMMQD12, pMMQD24, and pMMQD69, respectively. The intensities of the bands were quantified by PhosphorImager analysis (ImageQuant program). (A) Vector RNA expression in cells transfected with different vectors. (B) Vector RNA from cell-free virions harvested from the vector-transfected cells. (C) SNV cell-free virion RNA used as an internal control to monitor the yield of virion RNA isolation. The probe shown in Fig. 4A was used to detect vector RNA shown here in panels A and B, whereas the probe shown in Fig. 4B was used to detect internal control RNA shown here in panel C.

the viral DNA. The infected cell clones were selected based on the hygromycin-resistant phenotypes; consequently, the proviruses from these cell clones should contain *hygro*. Therefore, proviruses using aberrant minus-strand DNA transfer during viral replication should also have *hygro* and U5 sequences to allow amplification of the viral genome. However, the size of

the DNA fragments amplified by PCR should not be 0.45 kb but should vary depending on the locations of the donor and acceptor templates.

A representative agarose gel containing PCR-amplified DNA products from different cell clones is shown in Fig. 6B. Most of the PCRs using lysates infected with vectors having at least 12-nt homology (MMQD12, MMQD24, MMQD69) gave rise to the expected 0.45-kb DNA fragment (Fig. 6B and data not shown); 37 of 40 proviruses had the 0.45-kb fragment, whereas the other three proviruses generated DNA fragments of other sizes. In contrast, PCR analyses from proviruses generated by vectors with 3- or 6-nt homology produced very different results. Only 3 of 12 MMQD6 proviruses had the 0.45-kb DNA fragments, whereas the other 9 proviruses produced DNA fragments of other sizes (Fig. 6B and data not shown). Of the 11 proviruses generated from MMQD3, none had the 0.45-kb DNA fragment, and all had DNA fragments of different sizes (Fig. 6B and data not shown).

The molecular nature of minus-strand DNA transfer from these proviruses was characterized by directly sequencing the

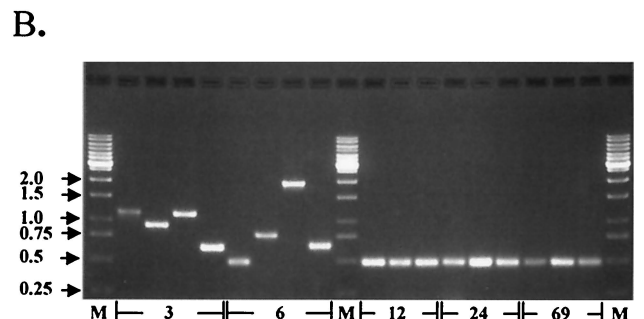
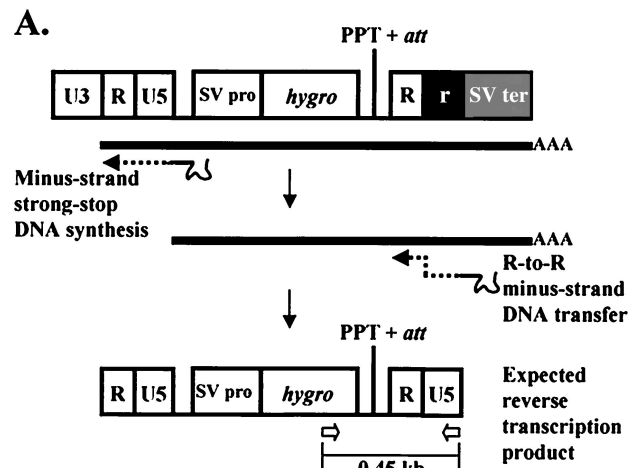


FIG. 6. Strategy used to analyze minus-strand DNA transfer junctions. (A) Expected minus-strand DNA transfer and predicted PCR product size. Thick lines, viral RNA; dotted lines with arrows, nascent DNA with arrows pointing toward the direction of DNA synthesis; open arrows, PCR primers; all other abbreviations are the same as in Fig. 1. (B) Representative agarose gel showing the products of PCR amplification from cell clones infected with various vectors. Lanes: M, molecular size markers in kilobases; 3, 6, 12, 24, and 69, lysates used for the PCR that were derived from cell clones infected with MMQD3, MMQD6, MMQD12, MMQD24, and MMQD69, respectively.

TABLE 1. Analysis of minus-strand DNA transfer junctions

Homology length (nt)	No. of proviruses with R-to-R transfers	No. of proviruses with transfers to other regions		No. of proviruses analyzed
		U3 read-in	Minus-strand strong-stop DNA	
3	0	10	1	11
6	3	9	0	12
12	10	1	0	11
24	15	0	0	15
69	12	2	0	14
Total	40	22	1	63

PCR products. These data are summarized in Table 1. Of the 40 proviruses that produced the 0.45-kb DNA fragments, 37 were produced by vectors with at least 12-nt homology and 3 were from MMQD6 (Table 1, column 2); all 40 of these proviruses were generated by R-to-R minus-strand DNA transfer. Moreover, all of the 40 transfer junctions were precise; we did not observe misincorporation, deletion, or addition of sequences. Sequencing analyses were also performed on the PCR DNA fragments generated from the 23 proviruses yielding different sizes that varied from 0.43 to 2.0 kb. Two transfer patterns emerged from these analyses based on the usage of donor and acceptor templates: misaligned minus-strand strong-stop DNA transfer and minus-strand DNA transfer using "read-in" RNA transcripts (see below).

The first pattern, misaligned minus-strand strong-stop DNA transfer, was observed only in one provirus generated from MMQD3; a summary of the transfer pattern is shown in Fig. 7. The PCR fragment from this provirus was 0.51 kb and contained most of the 3' end of *hygro*, a portion of PBS, a portion of Ψ^+ , the entire R, and U5. From these sequences, we deduced that the DNA was generated by two successive transfer events: one from R to Ψ^+ during minus-strand DNA transfer, and the other from PBS to *hygro* (Fig. 7A). Sequence alignment indicated that there was a 5-nt homology between the end of R and the region of Ψ^+ that served as an acceptor template; these 5 nt were probably used to mediate minus-strand DNA transfer (Fig. 7B). Sequence alignment also showed a 9-nt homology between the PBS and the 3' *hygro* that bridged the two sequences, indicating that it was used to mediate the second strand transfer event (Fig. 7B). Figure 7C illustrates the possible events that generated this provirus. During reverse transcription, minus-strand DNA transfer used the 5-nt homology to transfer to the Ψ^+ region; RT proceeded to copy 105 nt of the Ψ^+ sequence and 11 nt of the PBS. RT then used the 9-nt homology to switch the template to the 3' *hygro* and continued the DNA synthesis.

Minus-strand DNA transfer from read-in RNA transcripts. DNA sequencing analyses of 22 proviruses with aberrant PCR fragment sizes indicated that all of them contained portions of U3 sequences or the entire U3 region. Generally, viral transcripts are directed by upstream U3 sequences. These U3-derived RNA transcripts should begin with R, contain sequences between the two LTRs and then the downstream U3, and end with the downstream R prior to the addition of poly(A). Because the U3 sequences from the downstream LTR were deleted in these vectors, U3 sequences were not expected

to be present in the viral transcript (Fig. 8A). However, we observed many proviruses containing U3 sequences; furthermore, all of these proviruses had precise U3-R junctions and most of them used sequences from U3 or the plasmid backbone upstream of U3 as donors for strand transfer. This result indicated that these proviruses were generated from RNA transcripts containing the U3 sequences located in the upstream LTR. Therefore, we propose that these proviruses were generated from RNA transcripts expressed from promoters

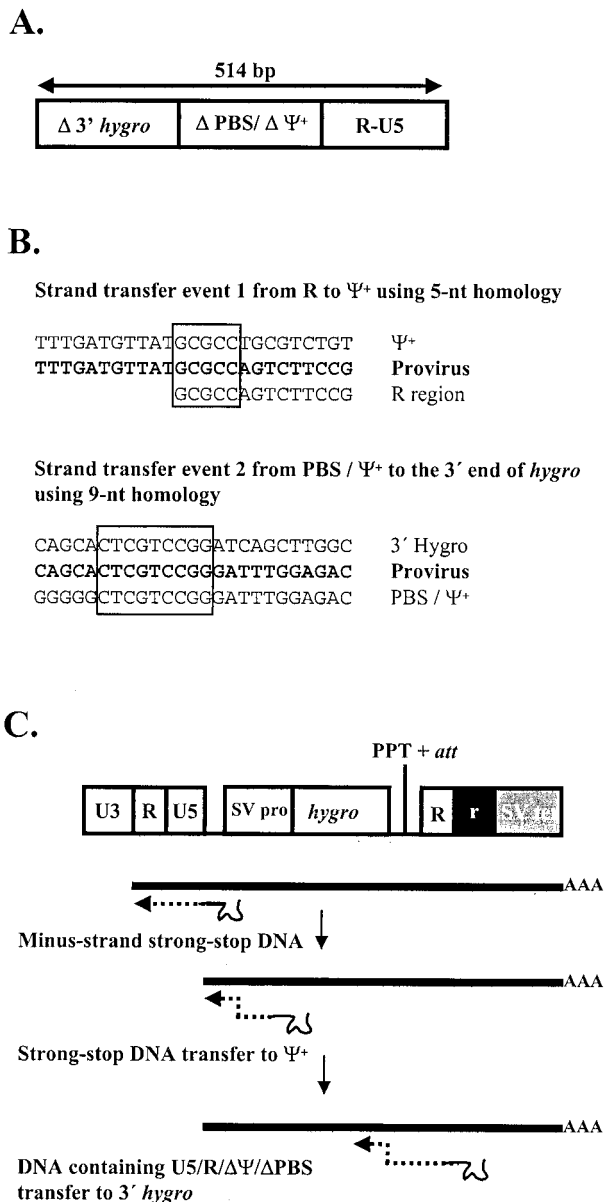


FIG. 7. Misaligned minus-strand DNA transfer using strong-stop DNA. PBS, primer binding site; Ψ^+ , extended packaging signal. (A) Illustration of the structure of an amplified PCR fragment from MMQD3 provirus. (B) Determination of transfer junctions by DNA sequence analyses and sequence alignment. Bold letters represent provirus sequences, whereas regular letters represent vector sequences. Open boxes indicate homology used for transfer. (C) Proposed strand transfer events used to generate the provirus.

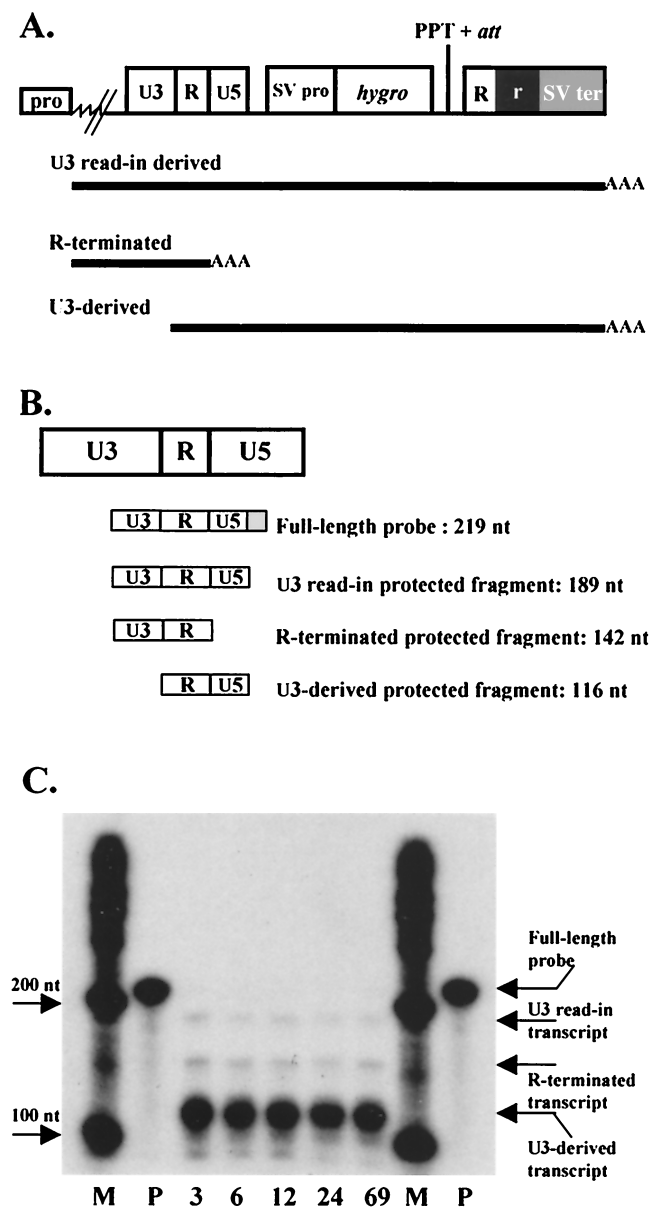


FIG. 8. Detection of read-in RNA transcripts by RNase protection assay. All abbreviations and symbols are the same as those in Fig. 4 and 5. (A) Three types of RNA transcripts that can be detected by RNase protection assay. (B) Strategy and probe used to detect read-in RNA transcripts. The full-length probe contains viral sequences (open boxes) corresponding to R, parts of U3 and U5, and a 30-nt nonviral sequence (gray box). The expected protected fragments are shown below the probe. (C) Results from a representative RNase protection assay examining total cellular RNA from transfected cells. Three distinct bands are observed, corresponding to the three different transcripts. In this example, the amounts of the read-in and R-terminated transcripts are 3 to 5% and 6 to 8% of the U3-derived RNA transcripts, respectively. RNA quantification was performed using a PhosphorImager and the ImageQuant program.

upstream of U3 and were termed “read-in” RNA transcripts (Fig. 8A). We hypothesize that during DNA transfection of the vector into the helper cells, vector DNA integrated into the helper cell genome randomly, with some of the DNA integrating close to other promoters. The transcript expressed by these

promoters could read into the vector sequences and generate RNAs containing all or a portion of the U3 sequences, followed by R, U5, and the rest of the vector sequences.

To examine whether read-in transcripts could be detected in transfected cells, RNase protection assays were performed with cellular RNA from transfected cells (Fig. 8A). Three types of viral transcripts might be present in the cellular RNA that could be detected by the RNase protection assay: the U3-derived transcript and two types of transcripts from the upstream promoters. One type of transcript, an R-terminated transcript, would read into U3 and terminate at the upstream R; these RNA transcripts would lack the rest of the viral *cis*-acting elements and were not expected to be packaged and contribute to provirus formation. The other type of RNA transcript, a read-in transcript, would read into U3, read through the poly(A) signal in R, and contain the rest of the viral sequences. Read-in transcripts were expected to be the precursors of the proviruses with U3 sequences.

A 219-nt probe was generated to hybridize to portions of the upstream LTR shared by all of the vectors (Fig. 8B). Covering the junctions of U3, R, and U5, this probe hybridized to the 3' 73 nt of U3, the entire R (69 nt), and the 5' 47 nt of U5. In addition, the last 30 nt of this probe contained sequences that did not hybridize to the viral genome and should be degraded during the assay. In an RNase protection assay, the U3-derived transcript would contain R and U5 sequences and should protect 116 nt of the probe (69 nt of R and 47 nt of U5). R-terminated transcripts derived from the upstream promoter would contain U3 and R and should protect 142 nt of the probe (73 nt of U3 and 69 nt of R). The read-in transcript would contain U3, R, and U5 and thus should protect 189 nt of the probe (73 nt of U3, 69 nt of R, and 47 nt of U5). For all of the vectors used in this study, most of the sequences from the downstream LTR, including the U3 and U5, were deleted; therefore, this assay would not be complicated by the U3 or U5 sequence near the 3' end of the transcripts.

A representative gel from an RNase protection assay is shown in Fig. 8C. In this and other analyses, three bands were detected from each cellular RNA sample, with sizes corresponding to the U3-derived, R-terminated, and read-in transcripts. In all samples, the U3-derived transcripts were the dominant RNA species; read-in RNA transcripts and R-terminated transcripts were approximately 3 to 8% and 6 to 11% of the U3-derived transcripts, respectively.

Minus-strand DNA transfer junctions generated from read-in RNA transcripts. In the read-in RNA transcripts, R regions were no longer at the 5' end of the RNA. DNA sequencing analyses revealed that during reverse transcription, DNA synthesis did not terminate at the end of the R but continued to copy the U3 sequences; in some cases, RT copied the entire U3 and went into the plasmid backbone adjacent to the U3 sequences. These U3 or plasmid backbone sequences, instead of R, were used as donor sequences for minus-strand DNA transfer.

In most of these transfer reactions, various regions of the 3' end of *hygro* or sequences downstream of *hygro* were used as acceptors (Fig. 9A), probably because the function of *hygro* was selected in these experiments. A summary of the locations of the donor and acceptor sites is shown in Fig. 9B. Sequence alignments indicated that 15 proviruses were generated by a

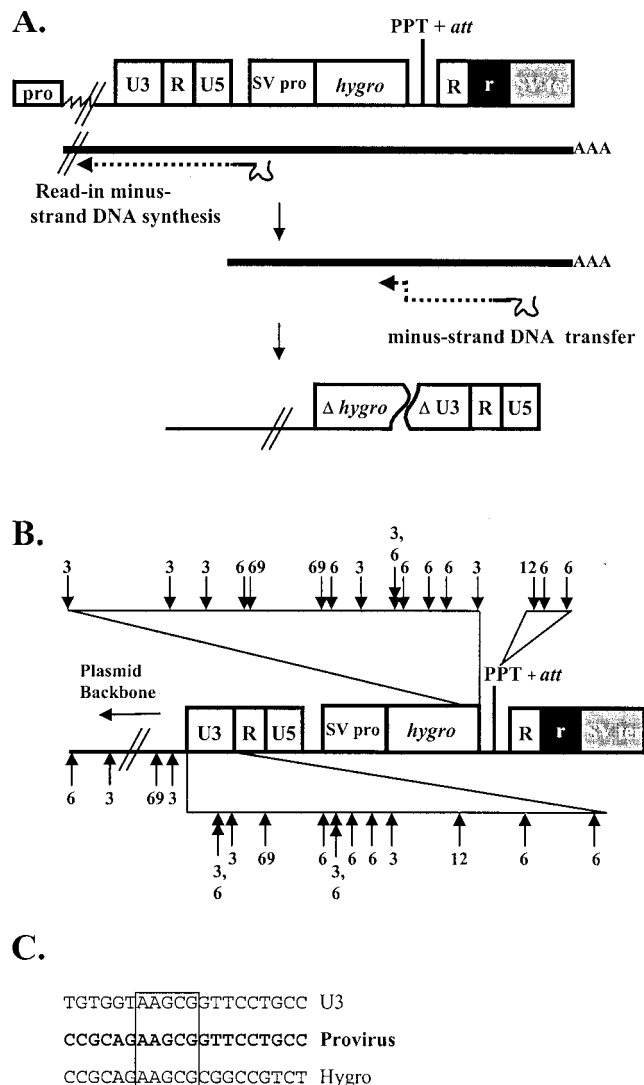


FIG. 9. Minus-strand DNA transfer of DNA synthesized from read-in RNA transcripts and locations of the donor and acceptor templates used for transfer. (A) Proposed model for the generation of proviruses containing U3 sequences by reverse transcription of read-in RNA transcripts. Pro, upstream promoter; zigzag line, host-cell genome. (B) Summary of locations of the donor and acceptor templates used for minus-strand DNA transfer. Locations of donor templates are shown below the vector structure, whereas locations of the acceptor templates are shown above the vector structure. Numbers above or below the arrows indicate the vector from which the provirus was derived: 3, pMMQD3; 6, pMMQD6; 12, pMMQD12; and 69, pMMQD69. (C) Example of a provirus generated from MMQD6 using 5 nt of homology between the donor template (U3) and acceptor template (3' hygro) to mediate minus-strand DNA transfer. Bold letters represent the proviral sequences, whereas regular letters represent vector sequences. The open box indicates homology used for the transfer.

single transfer event, using U3 or other upstream sequences as donors and 3' hygro as acceptors. An example of this type of junction is shown in Fig. 9C; in this portion of the provirus structure, 3' hygro sequences were joined to U3 sequences. Sequence alignment showed that 5-nt homology existed at the junction between the U3 and hygro sequences; the minus-strand DNA transfer event was presumably mediated by these 5-nt sequences.

Three proviruses were generated by more than one transfer event. Two of these proviruses underwent two transfers to join U3 to 3' hygro with all the sequences originating from the viral vectors. In the third provirus, five transfer events were observed to join U3 and hygro with the incorporation of sequences from an endogenous MLV isolated from *Mus musculus* (clone Mxv4) (53). This provirus is likely to be generated from copackaging of endogenous MLV followed by subsequent recombination events.

The exact number of transfers could not be determined in four proviruses because a small stretch of non-vector-derived sequences was identified between U3 and hygro. Homology searches of these sequences using the GenBank database identified one of the sequences as containing high homology to a submitted sequence from *Candida albicans* (accession number AL033501). It is highly probable that these stretches of non-vector-derived sequences were from the host cell genome adjacent to the integrated transfected DNA and thus were included in the RNA transcripts. During reverse transcription, RT copied the vector-derived sequences, continued to copy some of the host sequences, and then used these host sequences to mediate minus-strand DNA transfer. If this was the case, then only one transfer event was used to join the minus-strand DNA to the 3' hygro sequences.

The donor and acceptor sequences of 24 transfer events were identified during the analyses. Among these, 5 transfer events were performed without any observable homology and 19 transfer events were mediated by short homology. Of the 19 transfer events, 16 were mediated by homology shorter than 6 nt, 4 were mediated by 1-nt homology, 4 were mediated by 2-nt homology, 3 were mediated by 3-nt homology, 3 were mediated by 4-nt homology, and 2 were mediated by 5-nt homology. In addition, three transfer events were mediated by more than 11-nt homology (data not shown).

DISCUSSION

The efficiency and accuracy of minus-strand DNA transfer directly affect the ability of the virus to replicate and to preserve the genome structure. In this report, we delineate the role of R homology length in minus-strand DNA transfer.

Impact of R homology length on efficiency and accuracy of minus-strand DNA transfer. In these experiments, viral titers measured the efficiency of minus-strand DNA transfer, whereas the molecular characterization examined the accuracy of the transfer events by quantifying the frequencies of R-to-R transfer events. Our data indicated that removing more than 80% of the length of R in MLV, from 69 to 12 nt did not significantly affect the efficiency or accuracy of minus-strand DNA transfer. However, both the efficiency and accuracy of minus-strand DNA transfer were significantly diminished when the length of R was further reduced to 3 or 6 nt. Considering that most of the proviruses produced by vectors with 3- or 6-nt homology were generated from read-in RNA transcripts, the efficiencies of accurate minus-strand strong-stop DNA transfer in these vectors should have been even lower than the titers indicated. Taken together, these data indicate that a minimum amount of sequence homology is needed to perform efficient and accurate minus-strand DNA transfer.

In our MLV-based system, the vector with 12-nt R homology

could replicate with an efficiency close to that of the vector containing two full-length R regions. The effect of reducing the R homology in HIV-1 has been previously studied (3). Truncations of the downstream R from 97 nt in wild type to 37, 30, and 15 nt were generated; decreasing the downstream R to 30 or 15 nt reduced viral replication kinetics. Because multiple rounds of viral replications were allowed and the amounts of capsid-p24 released were used to plot the replication kinetics, it was difficult to directly correlate the difference between replication kinetics and efficiency of minus-strand DNA transfer. In our system, only one round of viral replication was allowed and the viral titers were generated and then used for statistical analyses. Although there seemed to be a slight decrease in the viral titers when the downstream R was reduced to 12 or 24 nt, statistical analyses indicated that the difference was not significant. In addition, because the R regions of HIV-1 and MLV are different, we could not rule out the possibility that the primary sequence or secondary structure of the R region may affect the requirement for the homology length.

In our experiment, the downstream R was manipulated, which may affect other events prior to minus-strand DNA transfer. However, it was shown that deletion of the 3' R did not affect viral titer when other homologous sequences were supplied for minus-strand DNA transfer (6). It was also shown that viral RNA synthesis was not affected by abolishing the polyadenylation signal in the downstream R (45). These data indicated that it is unlikely that the 3' R sequences affect the initiation of reverse transcription and synthesis of R-U5 DNA. We have shown experimentally that truncation of 3' R does not affect virus production or viral RNA packaging. Therefore, the differences in titers that we measured are most likely to be solely attributed to the efficiency and accuracy of the minus-strand DNA transfer.

Proposed mechanism for the requirement of R homology and the function of R. We propose that a minimum length of R homology is needed to align the newly synthesized DNA and the 3' RNA in the correct manner during minus-strand DNA transfer; reduction of homology length to a certain level can result in misalignment and possible dead-end reverse transcription products.

Although similar viral titers were generated by MMQD6 and MMQD3, accurate minus-strand DNA transfer was observed in MMQD6 (3 of 12) but not in MMQD3 (0 of 11). These data agree with our previous observation (55) that 6-nt homology could be used to mediate minus-strand DNA transfer. However, 6-nt homology was not sufficient to mediate minus-strand DNA transfer with the efficiency and accuracy of the wild-type virus.

If 6-nt homology is not sufficient to mediate efficient minus-strand DNA transfer, whereas 12-nt homology in R generates a wild-type phenotype, what about the lengths between 6 and 12 nt? By probability, a 6-nt homology occurs every 4,096 nt; a 10-kb viral genome is likely to contain other copies of the 6-nt sequence. By this calculation, a 7-nt homology occurring once in 16,384 nt is likely to be unique to the R sequences. However, additional lengths would reduce the probability of the minus-strand DNA to misalign with sequences having slightly less homology, such as those with 5- or 6-nt homology. Therefore, it is possible that the length of homology can be slightly re-

duced but probably not much from 12 nt before the efficiency and accuracy of the transfers are compromised. Indeed, the shortest naturally occurring R regions are 15 nt (MMTV) and 21 nt (Rous sarcoma virus) (7, 33, 46).

If the virus needs only 12 nt to mediate minus-strand DNA transfer, why do most retroviruses have longer R regions? In these experiments, we only examined the role of homology length in reverse transcription. However, most of the R regions contain *cis*-acting elements for other functions as well. For example, the R regions in many viruses contain the polyadenylation signal for RNA transcription (7). HIV-1 R also has the *trans*-activator response region (14, 39), which allows the binding of Tat and Tat-associated cellular proteins for transcription activation (29, 36). The R regions in HTLV-1 and HTLV-2 contain the Rex responsive element (41), which allows the binding of Rex protein to regulate the level of full-length and singly spliced RNA (17, 36). MLV R appears to be important in increasing cytoplasmic levels of full-length viral RNA (54). SNV R also serves as part of a translation enhancer (4). Therefore, the ability of the R sequences to mediate minus-strand DNA transfer is most likely not the only factor that impacts on the evolution of the R lengths; other aspects of viral replication also have strong influences on R lengths as well.

Minus-strand DNA transfer can use sequences other than R to mediate the transfer. During reverse transcription of wild-type retroviruses, R sequences are used for minus-strand DNA transfer. It had been proposed that R sequences evolved to facilitate template switching and that non-R sequences could not mediate minus-strand DNA transfer (1). However, in our experiments with the proviruses that were generated from the read-in RNA transcripts, U3 sequences, plasmid backbone sequences, and possibly cellular sequences were observed to mediate minus-strand DNA transfer. Furthermore, in a separate set of experiments, we also found that nonviral sequences could be used to efficiently mediate minus-strand DNA transfer (6). Taken together, these data indicate that the homology, not the sequence context of R, mediates minus-strand DNA transfer.

Short GC-rich sequences were used to mediate aberrant minus-strand DNA transfers. Previously, we proposed that GC-rich sequences were used in minus-strand DNA transfer junctions. However, minus-strand strong-stop DNA was predominantly the donor template and only two other junctions had been analyzed (55). In this report, some of the analyzed proviruses were generated from read-in RNA transcripts and used sequences other than the strong-stop DNA for minus-strand DNA transfer. This result allowed us to examine the content of the homology used for the transfer. We analyzed the homology sequences from 16 minus-strand DNA transfer events using 1 to 5 nt of homology; of the 43 nt that were utilized, 29 were C/G (14 C and 15 G) and 14 were A/T (6 A and 8 T). These data parallel the observation that GC-rich sequences were used to mediate deletions in MLV and SNV (22, 23, 34, 35).

We also compared R sequences from 10 different retroviruses (data not shown) and found that in general, R sequences were not GC-rich (ranging from 43 to 64%). However, the first nucleotide of the R region in all 10 viruses was G or C, and the second nucleotide of the R in 8 of 10 viruses was G or C as well. Together, these data suggest that GC-rich sequences may

be preferred at the junctions of the template-switching events during reverse transcription.

Minus-strand DNA transfer events are precise and occur without frequent nontemplate addition of sequences. In this report, we analyzed a total of 64 minus-strand DNA transfer junctions: 40 R-to-R transfer junctions and 24 non-R transfer junctions. We found that all of the transfer junctions were precise and we did not observe any nontemplate G additions or misincorporation of bases in the transfer junctions. In a separate experiment examining the sequence requirement of minus-strand DNA transfer, we also characterized 31 minus-strand DNA transfer junctions by DNA sequencing and found that all of the 31 transfer junctions were precise (6). Taken together, a total of 95 minus-strand DNA transfer junctions have been analyzed and all were precise.

It was previously reported that nontemplate addition of G occurred approximately 10% of the time during MLV replication (25). At the 10% frequency, we should have observed the G addition in approximately 9 transfer events among the 95 events that were analyzed; however, we did not observe any provirus with a G addition. This result led us to conclude that nontemplate G addition does not occur at a 10% frequency during minus-strand DNA transfer; rather, minus-strand DNA transfers are precise events.

Read-in RNA transcripts and reverse transcription products generated from read-in RNA transcripts. In these experiments, we observed RNA transcripts generated from upstream promoters that were termed read-in RNA transcripts; these are the first demonstrations of read-in RNA transcripts and proviruses generated from these RNAs. Several factors contributed strongly to the observation of read-in RNA transcripts and the resulting proviruses. First, the U3 viral promoter was most likely weakened by the promoter interference from the internal SV40 promoter (9); this effect may have strengthened the upstream promoter and increased the probability of generating read-in RNA transcripts. Indeed, the weakened U3 promoter was reflected in the relatively low viral titers generated by these vectors; the vectors with at least 12-nt R homology consistently generated titers between 2×10^2 and 6×10^2 CFU/ml (Fig. 3 and data not shown) rather than the usual 10^3 to 10^4 CFU/ml from PG13-transfected pools (data not shown). Second, removal of the R homology reduced the efficiency of R-to-R transfer and enhanced our ability to observe proviruses that used other mechanisms for minus-strand DNA transfer. Third, the removal of the downstream U3 allowed the observation of a large number of proviruses that would have otherwise gone undetected. If read-in transcription occurred in a wild-type provirus, both U3 sequences would be in the RNA. During reverse transcription, minus-strand DNA can use the U3 sequences to mediate transfer; the resulting provirus would have the same genotype as those that used the R sequences to mediate minus-strand DNA transfer.

This observation raised the questions of whether read-in RNA transcription occurs in nature and what would be the genetic consequences of such an event. If read-in RNA transcription occurs in the infection of wild-type viruses, it is likely to be less frequent than the events observed in this experimental system for the reasons described above. In most cases, it would have little evolutionary impact because the proviruses generated from these transcripts are likely to have the same

genotypes as those that were generated from U3-derived transcripts. However, it is possible that in rare instances, read-in RNA transcription can serve as a mechanism to incorporate host sequences in viral RNA. Similar to the RNA read-through transcription (15, 44, 45), these host sequences in viral RNA can be incorporated into the proviral genome through high-frequency recombination events during reverse transcription. In some of the proviruses analyzed in this set of experiments, nonvector sequences were found in the strand transfer junctions. One of these sequences had high homology to reported sequences in *C. albicans*, suggesting that they were probably derived from the host genome. These data suggest that read-in RNA transcription is a possible mechanism for oncogene transduction.

In summary, our findings demonstrate that there is a requirement for a minimum homology length for accurate and efficient minus-strand DNA transfer and retroviral replication. Although the length required for minus-strand DNA transfer is significantly shorter than the R lengths in most retroviruses, many retroviruses have evolved to contain longer R regions for other important functions affecting the regulation of viral replication. Many of the features of minus-strand DNA transfer observed in this study parallel other template-switching events during reverse transcription. For example, we found that minus-strand DNA transfer can be mediated by short stretches of homology, specific sequences are not required for the transfer, and transfer junctions are precise and occur without addition of sequences. These features are very similar to those observed in deletions and nonhomologous recombination (22, 23, 34, 35, 56). Therefore, these data support that minus-strand DNA transfer occurs via a mechanism similar to those used in other template-switching events during reverse transcription, which was hypothesized in a recently proposed "dynamic copy-choice model" (43).

ACKNOWLEDGMENTS

We thank Mithu Molla for the construction of many of the precursor plasmids needed for these experiments and Vinay K. Pathak for continuous intellectual input and discussions. We also thank Krista Delviks, William Fu, Vineet KewalRamani, Vinay K. Pathak, Dexter Poon, and Terence Rhodes for critical reading of the manuscript and Anne Arthur for expert editorial help. We thank John Coffin for discussions, suggestions, and coining the term "read-in RNA transcript." We also thank Jacqueline Dudley for help with clarifying the length of MMTV R.

This work was partially supported by extramural grants from the National Institutes of Health (R29 CA 58345-01 to W.-S.H.) and American Cancer Society (RPG MBC-97322 to W.-S.H.) and intramural funding from the HIV Drug Resistance Program, National Cancer Institute.

REFERENCES

- Allain, B., J. B. Rasclé, H. de Rocquigny, B. Roques, and J. L. Darlix. 1998. Cis elements and trans-acting factors required for minus strand DNA transfer during reverse transcription of the genomic RNA of murine leukemia virus. *J. Mol. Biol.* **277**:225–235.
- Baltimore, D. 1970. RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**:1209–1211.
- Berkhout, B., J. van Wamel, and B. Klaver. 1995. Requirements for DNA strand transfer during reverse transcription in mutant HIV-1 virions. *J. Mol. Biol.* **252**:59–69.
- Butsch, M., S. Hull, Y. Wang, T. M. Roberts, and K. Boris-Lawrie. 1999. The 5' RNA terminus of spleen necrosis virus contains a novel posttranscriptional control element that facilitates human immunodeficiency virus Rev/RRE-independent Gag production. *J. Virol.* **73**:4847–4855.
- Champoux, J. J. 1993. Roles of ribonuclease H in reverse transcription, p.

- 103–118. *In* A. M. Skalka and S. P. Goff (ed.), Reverse transcriptase. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
6. Cheslock, S. R., J. A. Anderson, C. K. Hwang, V. K. Pathak, and W.-S. Hu. 2000. Utilization of nonviral sequences for minus-strand DNA transfer and gene reconstitution during retroviral replication. *J. Virol.* **74**:9571–9579.
 7. Coffin, J. M. 1996. Retroviridae: the viruses and their replication, p. 1767–1847. *In* B. N. Fields, D. M. Knipe, and P. M. Howley (ed.), Fields virology, 3rd ed., vol. 2. Lippincott-Raven Publishers, Philadelphia, Pa.
 8. Dougherty, J. P., and H. M. Temin. 1987. A promoterless retroviral vector indicates that there are sequences in U3 required for 3' RNA processing. *Proc. Natl. Acad. Sci. USA* **84**:1197–1201.
 9. Emerman, M., and H. M. Temin. 1986. Comparison of promoter suppression in avian and murine retrovirus vectors. *Nucleic Acids Res.* **14**:9381–9396.
 10. Gelinis, C., and H. M. Temin. 1986. Nondefective spleen necrosis virus-derived vectors define the upper size limit for packaging reticuloendotheliosis viruses. *Proc. Natl. Acad. Sci. USA* **83**:9211–9215.
 11. Gilboa, E., S. W. Mitra, S. Goff, and D. Baltimore. 1979. A detailed model of reverse transcription and tests of crucial aspects. *Cell* **18**:93–100.
 12. 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**:179–188.
 13. Halvas, E. K., E. S. Svarovskaia, and V. K. Pathak. 2000. Development of an in vivo assay to identify structural determinants in murine leukemia virus reverse transcriptase important for fidelity. *J. Virol.* **74**:312–319.
 14. Hauber, J., and B. R. Cullen. 1988. Mutational analysis of the *trans*-activation-responsive region of the human immunodeficiency virus type 1 long terminal repeat. *J. Virol.* **62**:673–679.
 15. Herman, S. A., and J. M. Coffin. 1987. Efficient packaging of readthrough RNA in ALV: implications for oncogene transduction. *Science* **236**:845–848.
 16. 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.
 17. Inoue, J., M. Yoshida, and M. Seiki. 1987. Transcriptional (p40x) and post-transcriptional (p27x-III) regulators are required for the expression and replication of human T-cell leukemia virus type I genes. *Proc. Natl. Acad. Sci. USA* **84**:3653–3657.
 18. Iwasaki, K., and H. M. Temin. 1990. The efficiency of RNA 3'-end formation is determined by the distance between the cap site and the poly(A) site in spleen necrosis virus. *Genes Dev.* **4**:2299–2307.
 19. Iwasaki, K., and H. M. Temin. 1992. Multiple sequence elements are involved in RNA 3' end formation in spleen necrosis virus. *Gene Expr.* **2**:7–18.
 20. Iwasaki, K., and H. M. Temin. 1990. The U3 region is not necessary for 3' end formation of spleen necrosis virus RNA. *J. Virol.* **64**:6329–6334.
 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. Julius, J. G., T. Kim, G. Arnold, and V. K. Pathak. 1997. The antiretrovirus drug 3'-azido-3'-deoxythymidine increases the retrovirus mutation rate. *J. Virol.* **71**:4254–4263.
 23. Kim, T., R. A. Mudry, Jr., C. A. Rexrode II, and V. K. Pathak. 1996. Retroviral mutation rates and A-to-G hypermutations during different stages of retroviral replication. *J. Virol.* **70**:7594–7602.
 24. Klaver, B., and B. Berkhout. 1994. Premature strand transfer by the HIV-1 reverse transcriptase during strong-stop DNA synthesis. *Nucleic Acids Res.* **22**:137–144.
 25. Kulpa, D., R. Topping, and A. Telesnitsky. 1997. Determination of the site of first strand transfer during Moloney murine leukemia virus reverse transcription and identification of strand transfer-associated reverse transcriptase errors. *EMBO J.* **16**:856–865.
 26. Leis, J., A. Aiyar, and D. Corbrinik. 1993. Regulation of initiation of reverse transcription of retroviruses, p. 33–47. *In* A. M. Skalka and S. P. Goff (ed.), Reverse transcriptase. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
 27. Linial, M. L. 1999. Foamy viruses are unconventional retroviruses. *J. Virol.* **73**:1747–1755.
 28. Lobel, L. I., and S. P. Goff. 1985. Reverse transcription of retroviral genomes: mutations in the terminal repeat sequences. *J. Virol.* **53**:447–455.
 29. Luciw, P. A. 1996. Human immunodeficiency viruses and their replication, p. 1881–1952. *In* B. N. Fields, D. M. Knipe, P. M. Howley and (ed.), Fields virology, 3rd ed., vol. 2. Lippincott-Raven Publishers, Philadelphia, Pa.
 30. Martinez, I., and R. Dornburg. 1995. Improved retroviral packaging lines derived from spleen necrosis virus. *Virology* **208**:234–241.
 31. 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.
 32. Miller, A. D., and G. J. Rosman. 1989. Improved retroviral vectors for gene transfer and expression. *BioTechniques* **7**:980–990.
 33. Moore, R., M. Dixon, R. Smith, G. Peters, and C. Dickson. 1987. Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: two frameshift suppression events are required for translation of *gag* and *pol*. *J. Virol.* **61**:480–490.
 34. Parthasarathi, S., A. Varela-Echavarría, Y. Ron, B. D. Preston, and J. P. Dougherty. 1995. Genetic rearrangements occurring during a single cycle of murine leukemia virus vector replication: characterization and implications. *J. Virol.* **69**:7991–8000.
 35. 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.
 36. Rabson, A. B., and B. J. Graves. 1997. Synthesis and processing of viral RNA, p. 205–261. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (ed.), Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
 37. Ramsey, C. A., and A. T. Panganiban. 1993. Replication of the retroviral terminal repeat sequence during in vivo reverse transcription. *J. Virol.* **67**:4114–4121.
 38. Riggs, J. L., R. M. McAllister, and E. H. Lennette. 1974. Immunofluorescent studies of RD-114 virus replication in cell culture. *J. Gen. Virol.* **25**:21–29.
 39. Rosen, C. A., J. G. Sodroski, and W. A. Haseltine. 1985. The location of cis-acting regulatory sequences in the human T cell lymphotropic virus type III (HTLV-III/LAV) long terminal repeat. *Cell* **41**:813–823.
 40. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
 41. Seiki, M., J. Inoue, M. Hidaka, and M. Yoshida. 1988. Two cis-acting elements responsible for posttranscriptional trans-regulation of gene expression of human T-cell leukemia virus type I. *Proc. Natl. Acad. Sci. USA* **85**:7124–7128.
 42. Sodroski, J., M. Trus, D. Perkins, R. Patarca, F. Wong-Staal, E. Gelmann, R. Gallo, and W. A. Haseltine. 1984. Repetitive structure in the long-terminal-repeat element of a type II human T-cell leukemia virus. *Proc. Natl. Acad. Sci. USA* **81**:4617–4621.
 43. Svarovskaia, E. S., K. A. Delviks, C. K. Hwang, and V. K. Pathak. 2000. Structural determinants of murine leukemia virus reverse transcriptase that affect the frequency of template switching. *J. Virol.* **74**:7171–7178.
 44. Swain, A., and J. M. Coffin. 1992. Mechanism of transduction by retroviruses. *Science* **255**:841–845.
 45. Swain, A., and J. M. Coffin. 1989. Polyadenylation at correct sites in genomic RNA is not required for retrovirus replication or genome encapsidation. *J. Virol.* **63**:3301–3306.
 46. Swanstrom, R., H. E. Varmus, and J. M. Bishop. 1982. Nucleotide sequence of the 5' noncoding region and part of the *gag* gene of Rous sarcoma virus. *J. Virol.* **41**:535–541.
 47. Tanese, N., A. Telesnitsky, and S. P. Goff. 1991. Abortive reverse transcription by mutants of Moloney murine leukemia virus deficient in the reverse transcriptase-associated RNase H function. *J. Virol.* **65**:4387–4397.
 48. Telesnitsky, A., and S. P. Goff. 1997. Reverse transcriptase and the generation of retroviral DNA, p. 121–160. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (ed.), Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
 49. Temin, H. M. 1976. The DNA provirus hypothesis. *Science* **192**:1075–1080.
 50. Temin, H. M. 1981. Structure, variation and synthesis of retrovirus long terminal repeat. *Cell* **27**:1–3.
 51. Temin, H. M., and S. Mizutani. 1970. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**:1211–1213.
 52. Tisdale, M., T. Schulze, B. A. Larder, and K. Moelling. 1991. Mutations within the RNase H domain of human immunodeficiency virus type 1 reverse transcriptase abolish virus infectivity. *J. Gen. Virol.* **72**:59–66.
 53. Tomonaga, K., and J. M. Coffin. 1999. Structures of endogenous nonretroviral murine leukemia virus (MLV) long terminal repeats in wild mice: implication for evolution of MLVs. *J. Virol.* **73**:4327–4340.
 54. Trubetskoy, A. M., S. A. Okenquist, and J. Lenz. 1999. R region sequences in the long terminal repeat of a murine retrovirus specifically increase expression of unspliced RNAs. *J. Virol.* **73**:3477–3483.
 55. Yin, P. D., V. K. Pathak, A. E. Rowan, R. J. Teufel II, and W. S. Hu. 1997. Utilization of nonhomologous minus-strand DNA transfer to generate recombinant retroviruses. *J. Virol.* **71**:2487–2494.
 56. Zhang, J., and H. M. Temin. 1994. Retrovirus recombination depends on the length of sequence identity and is not error prone. *J. Virol.* **68**:2409–2414.