JOURNAL OF VIROLOGY, July 1994, p. 4196–4203
0022-538X/94/$04.00+0

Vol. 68, No. 7

High Rates of Frameshift Mutations within Homo-Oligomeric Runs during a Single Cycle of Retroviral Replication

DAWN P. W. BURNS* AND HOWARD M. TEMIN†

McArdle Laboratory for Cancer Research, University of Wisconsin Medical School, Madison, Wisconsin 53706

Homo-oligomeric runs were inserted into a spleen necrosis virus-based retrovirus vector to determine the nature and rate of mutations within runs of 10 to 12 identical nucleotides during a single replication cycle. Clones of helper cells containing integrated copies of retroviral vectors were used to produce virus for infection of target (nonhelper) cells. Proviral sequences from target cell clones were compared with proviral sequences from helper cell clones to study mutations that occurred during a single cycle of replication. In addition to the internal region spanning the homo-oligomeric inserts, a naturally occurring run of 10 T's in the long terminal repeat (LTR) also was sequenced. Rates of mutation ranged from <0.01 to 0.38 frameshift mutations per run per cycle for different nucleotide runs. Frameshift mutations ranged from deletions of 2 bases to additions of 5 bases; the most common mutations were +1 and −1. Frameshift mutation rates did not increase as the run length increased from 10 to 12 bases. Rates of frameshift mutation for runs of T's and A's were significantly higher than rates for runs of C's and G's, and rates for runs of pyrimidines were significantly higher than those for runs of purines. Interestingly, the vast majority of frameshift mutations in the internal region (95%) were positive, suggesting that the primer strand tends to slip backward on the template in this region. LTR runs had a significantly lower number of positive frameshift mutations than the internal runs. By analyzing the types of frameshift mutations within runs and by comparing the patterns of frameshift mutations in the 5′ and 3′ LTRs of individual proviruses, we conclude that the majority of mutations observed in our system occurred during minus-strand DNA synthesis of reverse transcription.

Spleen necrosis virus (SNV) is a member of the murine leukemia virus-related genus of the *Retroviridae* family. The genome of SNV consists of a noncovalently linked dimer of identical, positive-strand RNA (21). SNV uses a virally encoded enzyme called reverse transcriptase (RT) (20, 23) to replicate its RNA genome through a proviral DNA intermediate (33). SNV replication is highly error prone (6, 7) as a result of lack of proofreading and mismatch repair mechanisms (1, 24, 29).

An SNV-based shuttle vector has been used previously to characterize mutations occurring in a single cycle of replication. Base substitutions, frameshifts, hypermutations, deletions, and deletions with insertions were observed (25, 26). Frameshift mutations occurred exclusively at runs of 2 to 5 identical nucleotides within the *lacZ*α target gene of the SNV vector. Within the *lacZ*α target gene of a similar bovine leukemia virus-based shuttle vector, runs of 3 to 5 identical nucleotides also served as hot spots for frameshift mutations (22).

Although these previous in vivo studies on SNV and bovine leukemia virus reverse transcription identified runs of identical nucleotides as hot spots for mutation, the number of mutants characterized was small. Within the *lacZ*α target genes of the shuttle vectors, only five frameshift mutations were identified in the SNV study (26) and only four frameshift mutations were identified in the bovine leukemia virus study (22). Therefore, it is difficult to conclude from these two studies whether frameshift mutation rates differ significantly for runs of different lengths and different base compositions.

Naturally occurring runs of 9 A's and 10 T's, located in different regions outside the *lacZ*α gene, were also found to be hot spots for frameshift mutation during SNV reverse transcription (26). Runs of 9 A's were located within a direct repeat of 110 nucleotides in a region of the shuttle vector which was not under selective pressure. Analysis of frameshift mutation rates within these runs of 9 A's was complicated by the fact that one copy of the direct repeat was undergoing deletion at a rate of approximately 41% per single cycle of replication; some of the frameshift mutations and deletions of one copy of the direct repeat may have been causally linked (26). Runs of 10 T's were located in the R region of the long terminal repeat (LTR) between the poly(A) signal and the poly(A) site, a location under potential biological selection. For these reasons, it is again difficult to conclude whether frameshift mutation rates differ significantly for runs of different lengths and different base compositions.

An in vitro study using purified human immunodeficiency virus type 1 (HIV-1) RT has also shown that runs of 2 to 5 identical bases are hot spots for frameshift mutation (2). In this study, a gapped M13mp2 molecule was used as the substrate for DNA synthesis, and the wild-type *lacZ*α gene served as the mutational target. The results from this study revealed a ninefold preference for −1 rather than +1 frameshift errors. Frameshift error rates also differed for identical runs located in different regions of the *lacZ*α gene. Thus, sequence context affected the frameshift mutation rate, and runs of different lengths and different base compositions could not be compared directly.

A subsequent in vitro study using purified HIV-1 RT analyzed the fidelity of both RNA- and DNA-dependent DNA synthesis (4). Fidelity with an RNA template was >10-fold higher than fidelity with a DNA template for −1 frameshift errors at five different homo-oligomeric runs. These results suggest that frameshift mutation errors may occur more fre-

* Corresponding author. Mailing address: McArdle Laboratory for Cancer Research, University of Wisconsin Medical School, 1400 University Ave., Madison, WI 53706. Phone: (608) 262-2024. Fax: (608) 262-2824. Electronic mail address: burns@oncology.wisc.edu.
† Deceased.

quently during plus-strand rather than minus-strand synthesis of reverse transcription.

The observations made from cell-free studies may not be an accurate representation of reverse transcription in vivo. In a recent study comparing in vitro and in vivo base substitution mutation rates of murine leukemia virus on the same template sequence, it was found that the mutation rate of purified murine leukemia virus RT in vitro was about 30 times higher than the in vivo mutation rate (34). Accessory viral and cellular factors may have contributed to the increase in RT fidelity during DNA synthesis in vivo. In addition, cellular DNA mismatch repair mechanisms may have corrected mistakes made during reverse transcription (34).

In this report, we describe a comprehensive in vivo study of SNV reverse transcription within tracts of 10 to 12 identical nucleotides. SNV-based retroviral vectors were designed to contain runs of all four nucleotides on both the plus and minus strands of the proviral DNA. The synthetic runs were all inserted into the same internal region of the retroviral vector so that frameshift mutation rates for runs of different lengths and different base compositions could be directly compared. We found that rates of frameshift mutation varied depending on the nucleotide composition of the run. Frameshift mutation rates did not increase as run length increased from 10 to 12 bases. Unlike cell-free studies with HIV-1 RT in which negative frameshift mutations predominated (2), frameshift mutations which occurred during SNV reverse transcription in vivo were almost exclusively positive. Cell-free studies with RNA and DNA templates in the HIV-1 system suggested that frameshift mutation errors may occur more frequently during plus-strand DNA synthesis of reverse transcription (4). Interestingly, the majority of frameshift mutations observed after SNV reverse transcription in vivo appear to have been generated during minus-strand DNA synthesis.

MATERIALS AND METHODS

Definitions. Although frameshift mutation traditionally refers to additions or deletions of nucleotides within an open reading frame (except multiples of three), frameshift mutation in this paper refers to a small change in sequence length (≤5 bp) at any sequence location. Positive frameshift mutations add bases, while negative frameshift mutations delete bases.

Cells. D-17 and DSN cell lines were maintained in Temin-modified Eagle's medium (32) supplemented with 6% calf serum (EC_6) at 37°C and 5% CO_2. The D-17 cell line (ATCC CCL 183; American Type Culture Collection, Rockville, Md.) was derived from a dog osteosarcoma and is permissive for SNV infection (28). The DSN helper cell line (ATCC CRL-9939) expresses the *gag-pol* and *env* genes of SNV and was derived from D-17 cells (8). Selection with hygromycin B (Calbiochem-Novabiochem Corp., La Jolla, Calif.) was performed at 80 U/ml. DSN cell clones harboring replication competent proviruses were maintained in the constant presence of neutralizing anti-SNV sera to prevent superinfection as described previously (6).

Vectors. Long synthetic oligonucleotides containing runs of 10 T's and 10 C's or 10 G's and 10 A's were annealed and inserted in both orientations into the unique *Cla*I site located between the hygromycin B phosphotransferase gene (*hyg*) and the polypurine tract (ppt) of the retroviral vector pJD214Hy (Fig. 1) (6). Synthetic oligonucleotides containing runs of 12 T's and 12 C's or 12 G's and 12 A's were also annealed and inserted in both orientations into pJD214Hy; the sequences of these oligonucleotides were the same as those shown in Fig. 1 except for the lengths of the runs and are correspondingly designated TC12 and GA12. Prior to annealing complementary oligonucleotides, 5′ phosphate groups were added by using T4 polynucleotide kinase. Kinase activity was removed by heating reaction mixtures to 85°C, and the oligonucleotides were annealed by incubation at 65°C, 37°C, room temperature, and 0°C for 15 min each. The pJD214Hy vector was cleaved with *Cla*I, and 5′ phosphate groups were removed by using calf intestinal phosphatase. Approximately 0.2 ng of the annealed insert was mixed with 20 ng of *Cla*I-digested pJD214Hy, and the ligation mixture was used to transform either strain NM522 or strain XL1-Blue of *Escherichia coli*. Individual bacterial colonies were screened for the presence and orientation of the insert. The final clones were selected by sequencing the internal region of the vector spanning the insert.

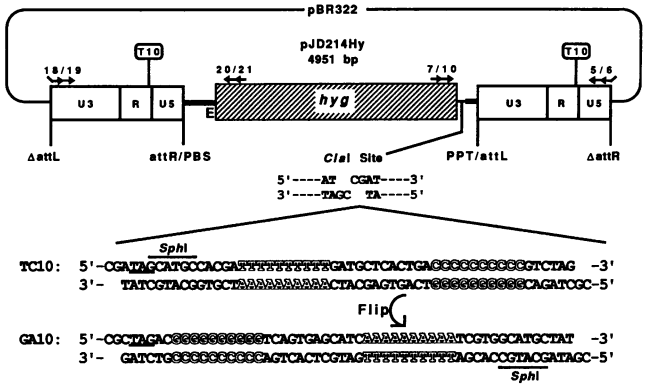

FIG. 1. Construction of vectors. Shown at the top is the plasmid form of the SNV-based vector pJD214Hy (6). U3, unique 3′ sequences; R, repeat sequences; U5, unique 5′ sequences; E, encapsidation sequence; attR, right attachment site; attL, left attachment site; PBS, primer-binding site; PPT, ppt; Δ, deletion; →, oligonucleotide primer location for PCR; T10, naturally occurring run of 10 T's. Thin horizontal lines represent pBR322 sequences. Thick horizontal lines represent SNV sequences. Shown at the bottom of the figure are the sequences of the synthetic oligonucleotide inserts, TC10 and GA10, which were cloned into the unique *Cla*I site between *hyg* and ppt. The locations of *Sph*I sites and stop codons (TAG) for *hyg* are indicated by overlining and underlining within the inserts. Similar insert sequences containing runs of 12 nucleotides, TC12 and GA12, also were synthesized and cloned into the *Cla*I site. These insert sequences (data not shown) are identical to TC10 and GA10 except for the lengths of the homo-oligomeric runs.

Transfection. Transfections were performed by the dimethyl sulfoxide-polybrene method as described previously (14).

Virus. Supernatant media containing neutralizing anti-SNV sera were removed from subconfluent 100-mm-diameter dishes of DSN helper cell clones, and the cells were washed with TD buffer (137 mM NaCl, 5 mM KCl, 0.7 mM $Na_2HPO_4 \cdot H_2O$, 25 mM Tris [pH 7.5]). Fresh EC_6 medium was added, and the cells were incubated for 1 h in a CO_2 incubator. Again, the supernatant medium was removed, and the cells were washed with TD buffer. Ten milliliters of fresh EC_6 medium containing 80 U of hygromycin B per ml was added, and the cells were incubated in a CO_2 incubator. After 24 h, cell supernatants were centrifuged at 1,600 × *g* for 12 min at 21°C to remove cellular debris. The top 6 ml of supernatant was removed and stored as virus stock in aliquots of 1.0 to 2.0 ml at −70°C.

For infection, virus was thawed quickly at 37°C, and serial 10-fold dilutions were made in EC_6 medium. A mixture of 0.2 ml of virus and 0.2 ml of polybrene (100 μg/ml in TD) was used to infect 2×10^5 cells per 60-mm-diameter plate. Virus was incubated with cells for 45 min in a CO_2 incubator, during

which time the plates were rocked by hand every 15 min. After 45 min, the virus was removed by aspiration, and 5 ml of EC_6 medium was added to the cells. The next day, cells were placed under selection by changing the medium to EC_6 containing 80 U of hygromycin B per ml. Fresh medium was added to the cells every 3 to 4 days for a total of approximately 3 to 4 weeks, at which time individual colonies of cells were large enough to be picked and expanded.

Cellular lysates. Cells from a confluent 100-mm-diameter tissue culture plate were trypsinized and washed twice in TD buffer. Cell pellets were resuspended in either 0.1 ml (for DSN cells) or 0.15 ml (for D-17 cells) of lysis buffer (100 mM KCl, 200 mM Tris-HCl [pH 8.8], 100 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$, 1% Triton X-100, 200 μg proteinase K per ml); this represented approximately 3×10^7 cells per ml for each cell line. The lysates were then diluted by a factor of 10 with sterile double-distilled H_2O containing 200 μg of proteinase K per ml. The lysates were incubated at 50 to 55°C overnight. The next day, the lysates were boiled for 10 min and spun in a microcentrifuge at full speed for 1 min. Lysates were stored at −20°C.

PCR. Genomic DNAs from cell lysates were used as templates in PCR to amplify specifically the proviral DNA. In order to obtain sufficient quantities of DNA for direct sequencing, two rounds of 30 cycles were performed. The first round of PCR was performed with an outer set of oligonucleotide primers, and the second round was performed with a second set of primers complementary to the amplified product of the first round. The internal region and 3′ LTR were amplified as one 740-bp fragment, with primers 7 and 6 in the first round and primers 10 and 5 in the second round (Fig. 1). The 5′ LTR was amplified as a 945-bp fragment, with primers 18 and 21 in the first round and primers 19 and 20 in the second round (Fig. 1) (the sequences of the primers are available upon request). For the first round of PCR, 50 μl of cellular lysate was amplified by using $Vent_R$ DNA polymerase in a 100-μl reaction mixture according to the manufacturer's suggestions (New England Biolabs, Beverly, Mass.). Five microliters was transferred from the first reaction mixture into a second 100-μl reaction mixture. The reaction mixture for round one was supplemented with 4.0 mM $MgSO_4$. Samples were amplified on a Perkin-Elmer DNA thermal cycler. Detailed cycle profile conditions are available upon request.

DNA sequencing. For direct DNA sequencing, three 100-μl reaction mixtures from the second round of PCR were combined and concentrated to a volume of 40 to 75 μl. Primers from PCR were removed by purification on CHROMA SPIN-400 columns (CLONTECH Laboratories, Inc., Palo Alto, Calif.). Sequencing was performed by using the CircumVent Thermal Cycle Dideoxy DNA Sequencing Kit (New England Biolabs) or the Sequenase DNA Sequencing Kit (United States Biochemical Corp., Cleveland, Ohio).

RESULTS

Design of vectors. Long synthetic oligonucleotides containing runs of 10 or 12 of each of the four nucleotides were inserted between *hyg* and the ppt of the SNV-based vector pJD214Hy (Fig. 1) (6). Purines flanked runs of pyrimidines, and pyrimidines flanked runs of purines. An in-frame amber stop codon was engineered into the synthetic insert to provide a translational stop for *hyg*. Sites for *Sph*I cleavage were also engineered into the inserts to distinguish between the two orientations. Insert sequences were further designed to remove the 3′ *Cla*I sites of inserts TC10 and TC12 and the 5′ *Cla*I sites of GA10 and GA12 in order to avoid the formation of a 6-bp direct repeat; short direct repeats have been previously shown to be hot spots for deletion during SNV reverse transcription (25). The remaining insert sequences were generated randomly. A region of 251 and 655 bp spanning the insert was analyzed by using the FoldRNA program of the University of Wisconsin Genetics Computer Group software package to ensure that the synthetic insert did not create unusually stable stem-loop structures in the viral RNA.

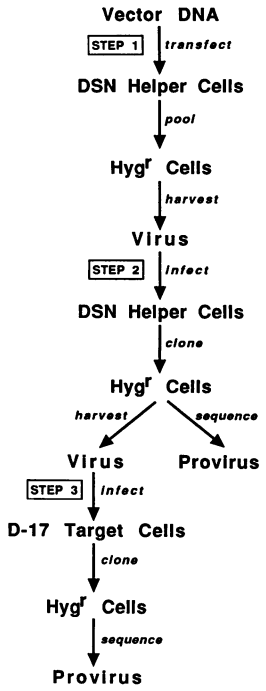

FIG. 2. Flow chart of experimental procedure.

In addition to the homo-oligomeric runs inserted between *hyg* and the ppt of pJD214Hy, a naturally occurring run of 10 T's was located in the R region of the LTR (Fig. 1). Runs of nucleotides in the synthetic insert are referred to as internal runs, while this naturally occurring run is referred to as either the 5′ or 3′ LTR run.

Single round of viral replication. Vector DNA was transfected into DSN helper cells, which express the *gag-pol* and *env* genes of SNV (Fig. 2, step 1). Hygromycin B was added to the medium of transfected cells to select for cells harboring integrated vector sequences. To remove mutations that may have occurred during transfection, virus was harvested from these cells and was used to infect fresh DSN helper cells (Fig. 2, step 2). Infected DSN helper cells were grown in the constant presence of neutralizing anti-SNV sera to prevent multiple rounds of infection. Hygromycin B-resistant (Hyg^r) DSN cells were cloned, the proviral sequences were amplified by PCR, and the amplified DNA was sequenced directly. Hyg^r DSN cell clones are referred to as parental cell clones. Virus was harvested from parental cell clones containing desired proviral sequences and was used to infect D-17 target cells (Fig. 2, step 3). D-17 cells do not express structural proteins of SNV, and virus cannot be produced from these cells. The infection at step 3 therefore represents one cycle of viral replication (from provirus in step-2 cells to provirus in step-3 cells). Hyg^r D-17 cells were cloned, the proviral sequences were amplified by PCR, and the amplified DNA was se-

TABLE 1. Frequencies of frameshift mutation for internal runs[a]

Nucleotide (plus strand)[b]	Type of frameshift	No. of frameshift mutations for parental run length of:			Totals for all lengths
		10	11	12	
T	−1	0	1	1	2
	+1	3	13	4	20
	+2	2	1	3	6
	+3	2	0	1	3
	+5	2	0	1	3
	Total (%)[c]	9/30 (30)	15/30 (50)	10/30 (33)	34/90 (38)
A	+1	6	4	4	14
	+2	1	0	3	4
	Total (%)	7/30 (23)	4/30 (13)	7/29 (24)	18/89 (20)
C	−1	1		0	1
	+1	4	NT[d]	0	4
	Total (%)	5/60 (8)		0/30 (<3)	5/90 (6)
G	Total (%)	0/60 (<2)	NT	0/29 (<3)	0/89 (<1)

[a] Statistical analysis is discussed in the text.
[b] Since proviral DNA was sequenced, deoxyribonucleotides are listed; plus-strand DNA corresponds to the plus-strand RNA genome of the virus.
[c] Total, no. of frameshift mutations/total no. of runs sequenced.
[d] NT, not tested.

quenced directly. Hyg^r D-17 cell clones are referred to as target cell clones.

Sequences of parental cell clones. Although the original vector DNAs contained runs of 10 or 12 nucleotides, some of the proviruses in parental cell clones had already undergone mutation and had runs of 11 nucleotides. Mutations may have occurred during transfection at step 1 or during infection at step 2 (Fig. 2). Apparently, some of the other proviruses in parental cell clones had inactivating mutations since they did not produce detectable levels of virus as measured by virus assay.

Ten different parental cell clones were chosen as producers of virus for step-3 infections of D-17 target cells. These clones had internal runs ranging in length from 10 to 12 nucleotides and were distributed as follows: two clones had 10 T's and 10 C's, one clone had 11 T's and 10 C's, two clones had 12 T's and 12 C's, one clone had 10 G's and 10 A's, one clone had 10 G's and 11 A's, and three clones had 12 G's and 12 A's in the plus strand. There were no base substitution mutations in the 36 bp surrounding the runs in any of the parental cell clones, as determined by nucleotide sequencing (data not shown). In addition, all parental cell clones had proviruses with 10 T's in the 5′ and 3′ LTR runs.

Frameshift mutations within internal runs. Proviral sequences from 179 randomly picked target cell clones were analyzed. In total, 57 of 358 internal runs had frameshift mutations (Table 1). Fifty-four of the 57 frameshift mutations (95%) were positive, adding bases to the length of the run. The +1 frameshift mutations were the most frequent type of positive frameshift observed for all nucleotides; 38 of 54 positive frameshifts (70%) were of the +1 type. No base substitution mutations were found at the ends of runs or in the 36 bp surrounding the runs in any of the target cell clones.

Runs of T's exhibited the greatest variety of frameshift mutations; frameshifts ranged from −1 to +5 (Table 1). For runs of A's, only +1 and +2 frameshift mutations were observed, and for runs of C's, only −1 and +1 frameshifts were observed. No frameshift mutations were observed for runs of 12 C's, 10 G's, or 12 G's. Double mutants were rarely seen on the same molecule. Of 54 mutant proviruses, only 3 had mutations in both internal runs.

Since mutation rates were not significantly different for run lengths of 10 to 12 bp, frameshift mutations were totaled for each nucleotide, and mutation rates were calculated as percentages per cycle of replication (Table 1). Frequencies from Table 1 were used in Fisher's exact test to determine if rate differences were statistically significant. The mutation rate for runs of T's was significantly higher than the rates for runs of A's ($P = 0.008$), C's ($P < 0.001$), and G's ($P < 0.001$). The mutation rate for runs of A's was significantly higher than the rate for runs of C's ($P = 0.003$) and G's ($P < 0.001$). Finally, the mutation rate for runs of C's was significantly higher than for runs of G's ($P = 0.03$).

Frameshift mutations within LTR runs. The 3′ LTR and internal runs were amplified as one fragment in PCR. Therefore, proviruses from the same 179 randomly picked target cell clones were studied. One 3′ LTR run could not be sequenced as a result of a deletion, so only 178 3′ LTR runs were analyzed. Thirty-one of the 178 LTR runs had frameshift mutations ranging from −2 to +2 (Table 2). Unlike the internal runs, positive frameshift mutations did not predominate in the LTR run. Instead, positive and negative frameshift mutations numbered approximately half and half; 14 of 31 (45%) frameshift mutations were positive, and 17 of 31 (55%) were negative. Differences in the frequencies of positive and negative frameshift mutations between runs of 10 T's in the internal region and in the 3′ LTR were statistically significant ($P = 0.003$). Frameshift mutations in LTR runs also differed from mutations in internal runs with regard to the lengths of frameshifts; long, positive frameshift mutations (+3 and +5)

TABLE 2. Frequencies of frameshift mutation for LTR runs

Type of frameshift	No. of frameshift mutations in:		
	3′ LTR, random[a]	5′ LTR, +3′ mutation[b]	5′ LTR, −3′ mutation[c]
−2	1	1	0
−1	16	15	3
+1	8	8	1
+2	6	6	0
Total[d]	31/178	30/31	4/24

[a] Cell clones were randomly chosen for sequencing of the 3′ LTR.
[b] Cell clones that had a frameshift mutation in the 3′ LTR were chosen for sequencing of the 5′ LTR.
[c] Cell clones that did not have a frameshift mutation in the 3′ LTR were chosen for sequencing of the 5′ LTR.
[d] Total, no. of frameshift mutations/total no. of runs sequenced.

were observed in internal runs but not in LTR runs (Tables 1 and 2). The frameshift mutation rate for runs of 10 T's in the 3′ LTR was 17% per cycle (31 of 178) and was lower than the rate of 30% per cycle (9 of 30) for runs of 10 T's in the internal region; however, the difference was not statistically significant. Like the internal runs, no base substitution mutations were found at the ends of 3′ LTR runs or in the surrounding 24 bp.

The 5′ LTR was amplified and sequenced for all 31 cell clones that had mutations in the 3′ LTR of the provirus. Thirty of 31 proviruses from these selected cell clones had identical mutations in the 5′ and 3′ LTR runs (Table 2). One provirus had discordant LTRs; the 5′ LTR had 10 T's, and the 3′ LTR had 9 T's.

The 5′ LTR was amplified and sequenced for an additional 24 cell clones that did not have a mutation in the 3′ LTR of the provirus. The rate and nature of mutations in the 5′ LTR runs of proviruses from these cell clones were similar to those in the 3′ LTR runs (Table 2). Four frameshift mutations were found among 24 runs, yielding a rate of 17% per cycle. Four proviruses with 5′ LTR mutations had discordant LTRs; one provirus had 11 T's in the 5′ LTR and 10 T's in the 3′ LTR, and three proviruses had 9 T's in the 5′ LTR and 10 T's in the 3′ LTR.

Of 55 proviruses that had mutations in either the 5′ or 3′ LTR run or both, only 5 had frameshift mutations in the internal regions; these 5 proviruses did not correspond to the 3 proviruses described above that had frameshift mutations in both internal runs. Thus, there was no correlation between frameshift mutations in the LTRs and internal regions.

Control experiments. Since proviral sequences had undergone 80 cycles of PCR amplification (two rounds of 30 cycles each plus an additional 20 cycles for sequencing), there is a possibility that the frameshift mutations observed in this system were a result of the polymerase used in PCR. This possibility is unlikely because the PCR-amplified DNA was directly sequenced (not cloned). Although mutations do occur during PCR, direct sequencing of PCR-amplified DNA yields a predominant DNA sequence.

In order to rule out the possibility that the observed frameshift mutations were caused by $Vent_R$ DNA polymerase during PCR amplification, a control experiment was performed. Lysates from cell clones containing proviruses with runs of either 12 T's and 12 C's or 12 G's and 12 A's in the plus strand were subjected to 21 independent PCR amplifications. DNA from these 42 reactions was directly sequenced. No frameshift mutations were found in any of the 84 internal runs, and no base substitutions were found in the sequences surrounding the runs. The frameshift mutations observed in our system, therefore, are not the result of the polymerase used in PCR.

DISCUSSION

In the studies described in this report, we determined the nature and rate of mutations within homo-oligomeric runs of 10 to 12 bases during a single cycle of SNV replication. Synthetic runs of all four nucleotides, located in the same internal region of an SNV-based retroviral vector, were analyzed along with a naturally occurring run of 10 T's in the LTR. By analyzing the types of frameshift mutations within the runs and by comparing the patterns of frameshift mutations in the 5′ and 3′ LTR runs of individual proviruses, we deduced mechanisms for the frameshift mutations observed in our system.

Mechanisms of frameshift mutation. Positive frameshift mutations add bases to the length of a run, while negative frameshift mutations delete bases. In the absence of recombination and nontemplated additions of nucleotides, a positive frameshift mutation could occur only if the primer strand slips backward on the template strand, one or more nucleotides loop out of the primer strand (9), and one or more bases of the run are recopied. On the other hand, a negative frameshift mutation could occur only if the primer strand slips forward on the template, one or more nucleotides loop out of the template strand, and one or more bases of the run are bypassed. Fifty-four of 57 frameshift mutations in the internal region and 14 of 31 frameshift mutations in the 3′ LTR run were positive and were most likely generated by the first mechanism described above in which the primer strand loops out. The three negative frameshift mutations in the internal region and 17 negative frameshift mutations in the 3′ LTR were most likely generated by the second mechanism in which the template strand loops out. The predominance of positive frameshift mutations in the internal region contrasts with cell-free studies of other polymerases such as HIV-1 RT, eukaryotic polymerases α and β, Klenow fragment (exo^-) (17), yeast DNA polymerase I (19), and phage T4 polymerase (31) in which negative frameshift mutations occur much more frequently than positive frameshift mutations.

During reverse transcription, positive-strand RNA is copied into double-stranded DNA (Fig. 3). Positive and negative frameshift mutations could be generated theoretically during both minus- and plus-strand DNA syntheses. If a positive frameshift mutation occurs during minus-strand DNA synthesis, the primer strand loops out (Fig. 4a). The template RNA is degraded, and the looped-out primer strand is straightened. When the straightened primer strand is copied during subsequent steps of reverse transcription, the added base(s) is copied and maintained in the provirus. If a negative frameshift mutation occurs during minus-strand DNA synthesis, the template strand loops out (Fig. 4b). The looped-out template strand is degraded, and the deletion is maintained in the provirus. Positive and negative frameshift mutations that occurred during plus-strand DNA synthesis have looped-out nucleotides on the primer and template strands, respectively (Fig. 4c and d); in either case, the completed double-stranded viral DNA has looped-out nucleotides in one of two strands.

Since reverse transcription takes place in the cytoplasm in a subviral particle in the cell, mismatch repair processes are not available at this stage of viral replication. However, when double-stranded viral DNA is finally transported into the nucleus and is integrated into the host chromosome, host cell repair mechanisms are operative. Proviral DNA with looped-

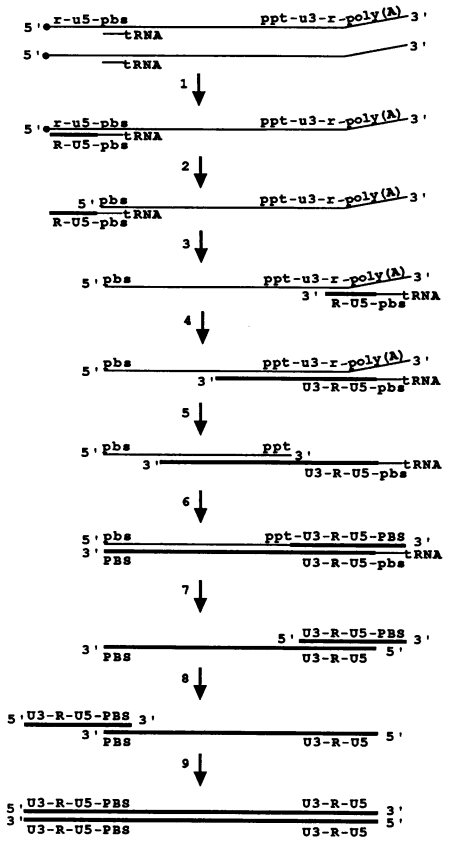

FIG. 3. Reverse transcription process. Thin lines and lowercase letters represent RNA. Thick lines and capital letters represent DNA. R, U5, U3, PBS, and PPT are defined in the legend to Fig. 1. ●, 5′ cap. DNA synthesis is performed by the polymerase function, and RNA degradation is performed by the RNase H function of reverse transcriptase. The starting products (top) are two identical strands of 5′-capped, poly(A) RNA bound to tRNA at the primer-binding site. Since one molecule of retroviral RNA is sufficient for viral DNA synthesis in the absence of recombination (13), products from only one RNA molecule are shown. The steps of reverse transcription are (i) minus-strand DNA primer synthesis, (ii) removal of RNA bound to minus-strand DNA primer, (iii) minus-strand DNA primer transfer (intrastrand), (iv) minus-strand DNA synthesis, (v) nicking of RNA at the ppt and removal of 3′ RNA sequences, (vi) plus-strand DNA primer synthesis, (vii) removal of tRNA and viral RNA sequences, (viii) plus-strand DNA primer transfer (intrastrand), and (ix) completion of plus- and minus-strand DNA synthesis.

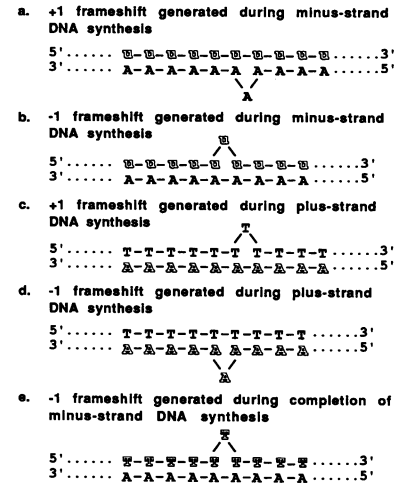

FIG. 4. Proposed structures of intermediates (a to e) for the generation of frameshift mutations. Lowercase letters represent RNA. Uppercase letters represent DNA. Outlined letters represent the template strand, and boldface letters represent the primer strand. Only one type of run is shown for intermediates a to d; runs of other nucleotides would have similar structures. For simplicity, only +1 and −1 frameshift mutations are shown; frameshift mutations of different lengths would have similar structures except for the sizes of the looped-out nucleotides. The position of the loop is arbitrary; the exact position of the frameshift mutation was not experimentally determined.

out nucleotides may or may not be repaired by host cell proteins. If proviral DNA is not repaired, then two cell populations exist after division, and the proviral DNA sequence is mixed (27); this was not observed in our experimental system. Since the homo-oligomeric runs are located immediately downstream of the actively transcribed *hyg* gene, mistakes in this region may have been preferentially repaired (11).

Most likely, proviral DNA with a looped-out nucleotide is repaired by a mechanism similar to one used for removal of thymidine dimers or bulky chemical adducts; a region of DNA is excised from the damaged (or looped-out) strand, and the single-stranded gap is filled in by DNA polymerase (10, 12). Therefore, a positive frameshift mutation that occurred during plus-strand DNA synthesis is corrected and is not observed in the proviral sequence. On this basis, we conclude that the 54 positive frameshift mutations observed in the internal runs were generated during minus-strand DNA synthesis. Positive frameshift mutations could actually have occurred with more frequency than negative frameshift mutations during plus-strand DNA synthesis, but such mutations would have gone undetected in our system. The three negative frameshift mutations observed in the internal runs could have occurred during synthesis of either minus- or plus-strand DNA.

Thirty one of 178 proviruses had mutations in the 3′ LTR run. Thirty of these 31 proviruses had identical mutations in the 5′ LTR run. In the absence of two mutational events, identical mutations in the LTRs could only be the result of a frameshift mutation that occurred during minus-strand DNA primer synthesis (Fig. 3, step 1). A mutated minus-strand DNA primer is transferred to the 3′ end of the viral RNA and is used as the template for synthesis of a mutated plus-strand DNA primer. The mutated plus-strand DNA primer is then transferred to the 5′ end of the viral DNA, resulting in identical mutations in both the 5′ and 3′ LTRs (Fig. 3).

LTR runs. Although rates of frameshift mutation for internal and LTR runs were not significantly different, LTR runs differed from internal runs in terms of the nature of frameshift mutations. While positive frameshift mutations predominated in the internal region, positive and negative frameshift mutations numbered approximately half and half in the LTR; the difference in the number of positive frameshift mutations between these two regions was statistically significant. Long, positive frameshift mutations (+3 and +5) were also observed within internal runs but not within LTR runs. We believe that these differences in the nature of frameshift mutations between the internal and LTR regions may be the result of sequence location, the presence of the C run near the T run in the internal region, or biological selection. Biological selection may occur at the level of mRNA as a result of the location of

the LTR run in R between the poly(A) signal and the poly(A) addition site (Fig. 1).

The rates of frameshift mutation for runs of 10 T's in the LTR were significantly higher in a previous study (26) than in the current study ($P = 0.001$). In the previous study, the helper cell line was based on reticuloendotheliosis virus strain A (35), while in the current study the helper cell line was based on SNV (8). Although reticuloendotheliosis virus strain A and SNV are highly related viruses (different strains of the same virus species), the RT enzymes may have different fidelities. In the previous study, target cells underwent higher passage number, and proviral DNA was cloned and replicated in bacteria prior to sequencing (as opposed to direct sequencing in the current study). Either of these two experimental differences could explain the higher mutation rate observed in the previous study since runs of identical nucleotides have been reported to be hot spots for mutation in both eukaryotic (15, 16, 18) and bacterial (3, 5) systems.

Discordant LTR sequences. Of 31 proviruses that had mutations in the 3′ LTR run, one did not have a mutation in the 5′ LTR run; this provirus had 10 T's in the 5′ LTR and 9 T's in the 3′ LTR. On the basis of our models of strand slippage, nucleotide looping, and cell repair described above and on the basis of the assumption that only one mutational event occurred, we propose that this pattern of mutation is the result of a frameshift mutation that occurred during completion of plus-strand DNA synthesis at step 9 of reverse transcription (Fig. 3 and 4d).

In sequencing the 5′ LTR of 24 proviruses that did not have mutations in the 3′ LTR run, we found four other proviruses that had discordant LTRs. One of these proviruses had 11 T's in the 5′ LTR and 10 T's in the 3′ LTR. Assuming one mutational event, we propose that the pattern of mutation observed in this provirus is the result of a frameshift mutation that occurred during plus-strand DNA primer synthesis at step 6 of reverse transcription (Fig. 3 and 4c). The other three proviruses with discordant LTRs had 9 T's in the 5′ LTR and 10 T's in the 3′ LTR. We propose that the pattern of mutation observed in these proviruses was the result of a frameshift mutation that occurred either during plus-strand DNA primer synthesis at step 6 (Fig. 3 and 4d) or during completion of minus-strand DNA synthesis at step 9 (Fig. 3 and 4e).

Nucleotide composition of runs. In comparing mutation rates for runs of different nucleotides, we found that rates for runs of T's and A's are higher than rates for runs of C's and G's. We propose that hydrogen bonding may explain this difference. We also found that rates for runs of pyrimidines are higher than rates for runs of purines. This difference may be explained by base-stacking energies which are greater between adjacent purines than between adjacent pyrimidines (30). Fitting with our model of positive frameshift mutations generated during minus-strand DNA synthesis, mutations within template runs of pyrimidines occur when purine residues loop out of the primer strand. Higher stacking energies between adjacent purines in the primer strand might help to stabilize a misaligned intermediate, thereby increasing the chance that such an intermediate is extended during reverse transcription.

Concluding remarks. Working in an in vivo system that isolates a single cycle of retroviral replication, we found that homo-oligomeric runs of 10 to 12 bases are extremely unstable, at least for certain nucleotides. For template runs of T's, error frequencies approached 50% in just one cycle of replication. For template runs of G's, on the other hand, error frequencies were less than 1% per replication cycle. With regard to the nature of frameshift mutations, we have observed predominantly positive frameshift mutations (additions of bases). In terms of the mechanism of frameshifting, the majority of the observed frameshift mutations appear to have been generated by primer strand looping during minus-strand synthesis of reverse transcription.

ACKNOWLEDGMENTS

D.P.W.B. dedicates this paper to the memory of H.M.T.

We thank Xiao-juan Bi, Eric Hausmann, and Bradley Seufzer for technical assistance. We thank Kathleen Boris-Lawrie, Jeff Jones, Louis Mansky, Gary Pulsinelli, Shiaolan Yang, and Jiayou Zhang for helpful discussions and suggestions. We also thank Thomas Kunkel, Shiaolan Yang, Louis Mansky, and Kathleen Boris-Lawrie for critical reading of the manuscript.

This research was supported by Public Health Service grants CA22443 and CA07175 from the National Cancer Institute (NCI). D.P.W.B. was supported by postdoctoral training grant CA09075 from NCI. H.M.T. was an American Cancer Society research professor.

REFERENCES

1. **Battula, N., and L. A. Loeb.** 1976. On the fidelity of DNA replication: lack of exodeoxyribonuclease activity and error-correcting function in avian myeloblastosis virus DNA polymerase. J. Biol. Chem. **251:**982–986.
2. **Bebenek, K., J. Abbotts, J. D. Roberts, S. H. Wilson, and T. A. Kunkel.** 1989. Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase. J. Biol. Chem. **264:**16948–16956.
3. **Bebenek, K., C. M. Joyce, M. P. Fitzgerald, and T. A. Kunkel.** 1990. The fidelity of DNA synthesis catalyzed by derivatives of *Escherichia coli* DNA polymerase I. J. Biol. Chem. **265:**13878–13887.
4. **Boyer, J. C., K. Bebenek, and T. A. Kunkel.** 1992. Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. Proc. Natl. Acad. Sci. USA **89:**6919–6923.
5. **Cupples, C. G., M. Cabrera, C. Cruz, and J. H. Miller.** 1990. A set of *lacZ* mutations in *Escherichia coli* that allow rapid detection of specific frameshift mutations. Genetics **125:**275–280.
6. **Dougherty, J. P., and H. M. Temin.** 1986. High mutation rate of a spleen necrosis virus-based retrovirus vector. Mol. Cell. Biol. **168:**4387–4395.
7. **Dougherty, J. P., and H. M. Temin.** 1988. Determination of the rate of base-pair substitution and insertion mutations in retrovirus replication. J. Virol. **62:**2817–2822.
8. **Dougherty, J. P., R. Wisniewski, S. Yang, B. W. Rhode, and H. M. Temin.** 1989. New retrovirus helper cells with almost no nucleotide sequence homology to retrovirus vectors. J. Virol. **63:**3209–3212.
9. **Fresco, J. R., and B. M. Alberts.** 1960. The accommodation of noncomplementary bases in helical polyribonucleotides and deoxyribonucleic acids. Proc. Natl. Acad. Sci. USA **46:**311–321.
10. **Hansson, J., M. Munn, W. D. Rupp, R. Kahn, and R. D. Wood.** 1989. Localization of DNA repair synthesis by human cell extracts to a short region at the site of a lesion. J. Biol. Chem. **264:**21788–21792.
11. **Hoeijmakers, J. H. J.** 1993. Nucleotide excision repair II: from yeast to mammals. Trends Genet. **9:**211–217.
12. **Huang, J.-C., D. L. Svoboda, J. T. Reardon, and A. Sancar.** 1992. Human nucleotide excision nuclease removes thymine dimers from DNA by incising the 22nd phosphodiester bond 5′ and the 6th phosphodiester bond 3′ to the photodimer. Proc. Natl. Acad. Sci. USA **89:**3664–3668.
13. **Jones, J. S., R. W. Allan, and H. M. Temin.** 1994. One retroviral RNA is sufficient for synthesis of viral DNA. J. Virol. **68:**207–216.
14. **Kawai, S., and M. Nishizawa.** 1984. New procedure for DNA transfection with polycation and dimethyl sulfoxide. Mol. Cell. Biol. **4:**1172–1174.
15. **Kunkel, T. A.** 1985. The mutational specificity of DNA polymerase-β during *in vitro* DNA synthesis. J. Biol. Chem. **260:**5787–5796.
16. **Kunkel, T. A.** 1985. The mutational specificity of DNA polymerases-α and -γ during *in vitro* DNA synthesis. J. Biol. Chem. **260:**12866–12874.

17. **Kunkel, T. A.** 1990. Misalignment-mediated DNA synthesis errors. Biochemistry **29:**8003–8011.
18. **Kunkel, T. A., and P. S. Alexander.** 1986. The base substitution fidelity of eucaryotic DNA polymerases. J. Biol. Chem. **261:**160–166.
19. **Kunkel, T. A., R. K. Hamatake, J. Motto-Fox, M. P. Fitzgerald, and A. Sugino.** 1989. Fidelity of DNA polymerase I and the DNA polymerase I-DNA primase complex from *Saccharomyces cerevisiae*. Mol. Cell. Biol. **9:**4447–4458.
20. **Leis, J., D. Baltimore, J. M. Bishop, J. Coffin, E. Fleissner, S. P. Goff, S. Oroszlan, H. Robinson, A. M. Skalka, H. M. Temin, and V. Vogt.** 1988. Standardized and simplified nomenclature for proteins common to all retroviruses. J. Virol. **62:**1808–1809.
21. **Levy, J. A.** 1992. The Retroviridae. Plenum Press, New York.
22. **Mansky, L. M., and H. M. Temin.** 1994. Lower mutation rate of bovine leukemia virus relative to that of spleen necrosis virus. J. Virol. **68:**494–499.
23. **Mizutani, S., and H. M. Temin.** 1975. Purification and properties of spleen necrosis virus DNA polymerase. J. Virol. **16:**797–806.
24. **Morrison, A., J. B. Bell, T. A. Kunkel, and A. Sugino.** 1991. Eukaryotic DNA polymerase amino acid sequence required for 3′→5′ exonuclease activity. Proc. Natl. Acad. Sci. USA **88:**9473–9477.
25. **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.
26. **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.
27. **Pulsinelli, G. A.** 1994. Ph.D. thesis. University of Wisconsin, Madison.
28. **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.
29. **Roberts, J. D., K. Bebenek, and T. A. Kunkel.** 1988. The accuracy of reverse transcriptase from HIV-1. Science **242:**1171–1173.
30. **Saenger, W.** 1984. Principles of nucleic acid structure. Springer-Verlag, New York.
31. **Streisinger, G., and J. E. Owen.** 1985. Mechanisms of spontaneous and induced frameshift mutation in bacteriophage T4. Genetics **109:**633–659.
32. **Temin, H. M.** 1968. Studies on carcinogenesis by avian sarcoma viruses. VIII. Glycolysis and cell multiplication. Int. J. Cancer **3:**273–282.
33. **Temin, H. M.** 1976. The DNA provirus hypothesis. Science **192:**1075–1080.
34. **Varela-Echavarría, A., N. Garvey, B. D. Preston, and J. P. Dougherty.** 1992. Comparison of Moloney murine leukemia virus mutation rate with the fidelity of its reverse transcriptase *in vitro*. J. Biol. Chem. **267:**24681–24688.
35. **Watanabe, S., and H. M. Temin.** 1983. Construction of a helper cell line for avian reticuloendotheliosis virus cloning vectors. Mol. Cell. Biol. **3:**2241–2249.