JOURNAL OF VIROLOGY, Dec. 1995, p. 7991–8000 Vol. 69, No. 12
0022-538X/95/$04.00+0

Genetic Rearrangements Occurring during a Single Cycle of Murine Leukemia Virus Vector Replication: Characterization and Implications

SHOBHA PARTHASARATHI,[1,2] ALFREDO VARELA-ECHAVARRÍA,[1]† YACOV RON,[1] BRADLEY D. PRESTON,[3] AND JOSEPH P. DOUGHERTY[1]*

Department of Molecular Genetics and Microbiology, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854-5635[1]; Graduate Program in Microbiology and Molecular Genetics, Rutgers University, New Brunswick, New Jersey 08903[2]; and Section of Experimental Oncology, Eccles Institute of Human Genetics, University of Utah, Salt Lake City, Utah 84112[3]

Received 20 July 1995/Accepted 21 September 1995

Retroviruses evolve at rapid rates, which is presumably advantageous for responding to selective pressures. Understanding the basic mutational processes involved during retroviral replication is important for comprehending the ability of retroviruses to escape immunosurveillance and antiviral drug treatment. Moreover, since retroviral vectors are important vehicles for somatic cell gene therapy, knowledge of the mechanism of retroviral variation is critical for anticipating untoward mutational events occurring during retrovirus-mediated gene transfer. The focus of this report is to examine the spectrum of genomic rearrangements arising during a single cycle of Moloney murine leukemia virus (MoMLV) vector virus replication. An MoMLV vector containing the herpes simplex virus thymidine kinase (*tk*) gene was constructed. MoMLV vector virus was produced in packaging lines, and target cells were infected. From a total of 224 mutant proviruses analyzed, 114 had gross rearrangements readily detectable by Southern blotting. The remaining proviruses were of parental size. PCR and DNA sequence analysis of 73 of the grossly rearranged mutant proviruses indicated they resulted from deletions, deletions combined with insertions, duplications, and complex mutations that were a result of multiple genomic alterations in the same provirus. Complex hypermutations distinct from those previously described for spleen necrosis virus and human immunodeficiency virus were detected. There was a correlation between the mutation breakpoints and single-stranded regions in the predicted viral RNA secondary structure. The results also confirmed that the *tk* gene is inactivated at an average rate of about 8.8% per cycle of retroviral replication, which corresponds to a rate of mutation of 3%/kbp.

To study the mutational processes occurring during retroviral replication, protocols have been developed to determine retroviral mutation rates, that is, the mutation frequency in a single cycle of retroviral replication, and to examine the nature of the mutations arising after one cycle of replication. A few mutation rate studies have been performed with replication-competent retroviruses (17, 22), but they have essentially been limited to base substitution mutation rate determinations and did not allow an extensive examination of the spectrum of mutations that might arise during a single cycle of replication. An inherent difficulty with the use of replication-competent retroviruses for determining mutation rates and characterizing the mutations is posed by their ability to spread throughout a culture and by a lack of selection protocols to identify mutants phenotypically. However, vector systems that facilitate the determination of mutation rates in the absence of helper virus and allow the detailed analysis of resulting mutations have been devised (7, 18, 25, 26, 35, 36). To date, these studies have focused primarily upon spleen necrosis virus (SNV), an avian retrovirus, although there have been some recent reports concerning Moloney murine leukemia virus (MoMLV) (35, 36), bovine leukemia virus (18), and human immunodeficiency virus (19).

Previously, we reported the development of a forward mutation rate assay using an MoMLV vector containing the herpes simplex virus (HSV) thymidine kinase (*tk*) gene (36). The average rate of inactivation of the *tk* gene was 14.4% per cycle of replication. Southern analysis and DNA sequencing of a modest number of mutant proviruses suggested that genetic rearrangements occurred at a high rate during vector virus replication (36). About half of the mutants analyzed resulted from deletions involving a 30-bp direct repeat flanking the *tk* gene. Excluding the mutations involving the 30-bp repeats, the rate was adjusted to 7.7% per replication cycle.

In this report, we examine genetic rearrangements by Southern blotting and DNA sequencing of a large number of mutant proviruses in order to shed light upon the mechanisms of retroviral mutation and to develop improved strategies for gene transfer with retroviral vectors. Furthermore, this was done with an MoMLV vector that did not contain the 30-bp direct repeats to confirm that our previous adjustment of the rate of mutation was accurate. The mutation rate obtained with the modified vector was 8.8% per cycle of replication, and since the locus is 2.62 kbp, the mutation rate is 3.3%/kbp per replication cycle. Analysis of 244 mutant proviruses showed that 114 were the result of gross rearrangements, including some novel mutations not observed heretofore.

* Corresponding author. Mailing address: Department of Molecular Genetics and Microbiology, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, 675 Hoes Ln., Piscataway, NJ 08854-5635. Phone: (908) 235-4588. Fax: (908) 235-5223.

† Present address: Division of Anatomy and Cell Biology, Guy's Hospital, London Bridge, London SE1 9RT, England.

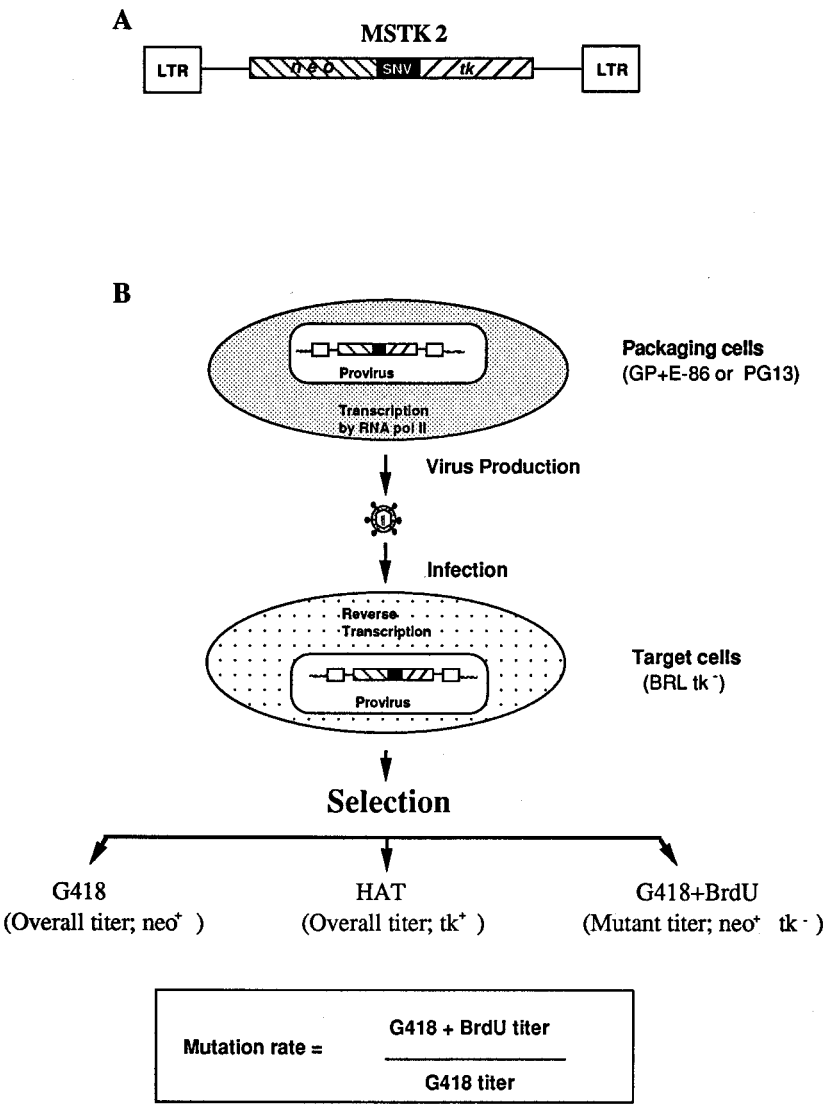

FIG. 1. Retroviral vector MSTK2 and the infection protocol used for forward mutation rate assay. (A) Schematic representation of the MoMLV-based vector MSTK2. LTR, MoMLV long terminal repeat; SNV, SNV U3 promoter; horizontal lines, MoMLV sequences containing *cis*-acting elements required for vector virus replication; *neo*, neomycin phosphotransferase gene; *tk*, HSV *tk* gene. (B) Protocol used to assay the forward mutation rate occurring in a single cycle of viral replication. GP+E-86 is an MoMLV-derived helper cell line; PG13 is a helper cell line which produces the MoMLV vector particles pseudotyped with GaLV Env glycoproteins. BRLtk$^-$ is a buffalo rat liver cell line permissive for infection by MoMLV-based vectors and by MoMLV-based vectors packaged with GaLV Env glycoproteins. The MSTK2 vector provirus inside the cells is shown in panel A. The jagged lines denote genomic DNA. G418, HAT, and G418+BrdU indicate the selections used to obtain the vector virus titers; neo$^+$, tk$^+$, and neo$^+$ tk$^-$ represent the phenotypes conferred by the vector proviruses.

MATERIALS AND METHODS

Plasmids. The pMSTK2 vector is derived from pMSTK, which has been described elsewhere (36). It contains the SNV U3 promoter ligated to the HSV *tk* gene. Briefly, the pMSTK2 vector was made by excising the SNV-*tk* cassette from pSNVtk (36) with *Eco*RI and *Ppu*MI and ligating the fragment to pN2 cut with *Xho*I (13). The resultant clone lacks an 86-bp sequence present at the 5′ end of the SNV promoter-*tk* cassette of pMSTK. The 86-bp sequence contained a 30-bp stretch which formed a direct repeat with vector sequences 3′ to *tk*; deletions between these direct repeats were responsible for at least half of the genetic rearrangements seen with the original pMSTK construct (36). Thus, pMSTK2 has only the 3′ copy of this 30-bp region and no longer contains direct repeats flanking the *tk* gene.

Cells. NIH 3T3 is a murine fibroblast cell line permissive for infection by MoMLV and MoMLV-based vectors. GP+E-86 is an NIH 3T3-derived packaging cell line which produces MoMLV proteins required for the propagation of replication-defective MoMLV-based vectors, without producing replication-competent virus (20). PG13 is another NIH 3T3-derived packaging cell line which expresses the MoMLV Gag and Pol proteins and gibbon ape leukemia virus (GaLV) Env proteins (21). This cell line can be used to propagate pseudotyped MoMLV-based vector virus with GaLV receptor specificity. The advantage of using this cell line is that it prevents any spread of vector virus in the packaging cells, since NIH 3T3 does not have the receptor for GaLV. The BRLtk$^-$ cell line is a bromodeoxyuridine (BrdU)-resistant buffalo rat liver cell line permissive for infection by ecotropic MoMLV vector virus and GaLV-pseudotyped MoMLV vector virus.

Transfections and infections. Vector plasmid DNA was transfected by the Polybrene-dimethyl sulfoxide method (12) into the GP+E-86 packaging cell line. Cells were infected by inoculating 60-mm-diameter dishes containing 2×10^5 cells with virus in the presence of 50 μg of Polybrene per ml in 0.4 ml of medium. Vector virus titers were determined by the infection of BRLtk$^-$ cells with 10-fold serial dilutions of virus stocks. Twenty-four hours after infection, the infected BRLtk$^-$ cells were placed in selective media containing G418 (300 μg/ml), HAT (0.1 mM hypoxanthine, 0.5 μM aminopterin, 30 μM thymidine), or both G418 and BrdU (G418+BrdU; 300 and 50 μg/ml, respectively). Titers were obtained by multiplying the number of resistant colonies by the dilution factor.

Generation of packaging cell clones harboring a single parental vector provirus. Packaging cell clones containing a single parental provirus were produced by (i) the transfection of GP+E-86 cells with vector plasmid pMSTK2 and selection in medium containing G418, (ii) harvesting vector virus from the G418-

TABLE 1. Titers and forward mutation rates obtained from GP+E-86 and PG13 cells containing MSTK2 proviruses

Clone[a]	Titer (CFU/ml of virus stock)			Mutation rate[b]
	G418	HAT	G418+BrdU	
1	1.7×10^6	1.5×10^6	0.9×10^5	0.053
2	2.5×10^6	2.2×10^6	1.1×10^5	0.044
3	2.5×10^6	2.7×10^6	1.9×10^5	0.076
4	1.8×10^6	1.7×10^6	1.3×10^5	0.072
5	2.3×10^6	1.9×10^6	1.5×10^5	0.065
6	0.2×10^6	0.2×10^6	0.3×10^5	0.150
7	0.7×10^6	1.3×10^6	7.5×10^4	0.107
8	0.7×10^6	1.3×10^6	0.4×10^5	0.057
9	0.3×10^5	0.2×10^5	0.4×10^4	0.133
10	2.0×10^5	1.9×10^5	2.5×10^4	0.125
Mean ± SD				0.088 ± 0.035

[a] Packaging cell clones 1 to 5 were derived from GP+E-86 cells, and clones 6 to 10 were derived from PG13 cells.

[b] The mutation rate is the ratio of the mutant virus titer (G418+BrdU) to the overall virus titers (G418).

TABLE 2. Size distribution of MSTK2 mutant proviruses arising in a single cycle of vector virus replication

Packaging cell line	Clone	No. of clones	No. of mutant proviruses		
			Same size as wild type	Smaller than wild type	Larger than wild type
GP+E-86	G2	82	45	27	10
	G3	81	42	31	8
PG13	P7	45	22	20	3
	P8	36	21	14	1
Total		244	130	92	22

resistant cells, and (iii) the infection of fresh GP+E-86 and PG13 packaging cells at a low multiplicity of infection to obtain G418-resistant cell clones harboring a single provirus (36). Those packaging cell clones that yielded G418-resistant colonies but no HAT-resistant cells upon the infection of BRLtk$^-$ cells were not used for mutation rate determinations because they probably represented clones with defective proviruses generated during either the transfection of GP+E-86 cells or the infection of GP+E-86 and PG13 cells.

Southern blot analysis. Genomic DNA was prepared and gel electrophoresis and Southern blotting were performed according to standard procedures (33). Genomic DNA was digested with *Xba*I, electrophoresed in a 1% agarose gel, and blotted onto nitrocellulose, after which it was hybridized with a *neo*-specific ^{32}P-labeled probe (an *Eco*RI-*Xho*I DNA fragment of plasmid N2) and autoradiographed.

DNA amplification and sequencing. To analyze the MSTK2 mutant proviruses, different combinations of primer pairs spanning the SNV promoter and *tk* locus were used to amplify the genomic DNA from the infected clones by PCR with *Taq* DNA polymerase (Perkin-Elmer Cetus). The primers used are listed below from 5′ to 3′ with their orientations and their locations on MSTK2: 1, GCCTTCTTGACGAGTTCTTC (plus, 2303 to 2325); 2, CTGCAAGCAAT TCGTTCTGT (minus, 2715 to 2695); 3, ATGGCCGCTTTGGTCGA (plus, 2652 to 2668); 4, AGCCATACGCGCTTC (minus, 3315 to 3295); 5, CTAACT GAGCCAATTGGTTG (plus, 3134 to 3152); 6, ACATGAGGAGCCAGAA CGGCGTCGGTCA (minus, 3732 to 3705); 7, TACGTACCCGAGCCGAT GACTTACTGGCA (plus, 3542 to 3570); 8, ATTGGCAAGCAGCCCGTAAA (minus, 4037 to 4017); 9, CGCGCGATACCTTATGGGCAGCATGA (plus, 3826 to 3851); 10, GTTCCTTCCGGTATTGT (minus, 4466 to 4450); 11, TTGG ACGTCTTGGCCAAA (plus, 4239 to 4256); 12, CTTGCCAAACCTACA (minus, 5034 to 5016); 13, AGTAAGTCATCGGCTCG (minus, 3566 to 3550); 14, AAAACCACCACCACGCAA (plus, 3487 to 3505); and 15, CAGGCAAACAC GTTATA (minus, 4234 to 4227). The primers were used in standard reactions, and their products were visualized after gel electrophoresis. PCR fragments of abnormal sizes were then purified from 1% agarose gels with the Geneclean kit (Bio 101, Inc., La Jolla, Calif.) or Qiaquick columns (QIAGEN Inc., Chatsworth, Calif.). The purified fragments were sequenced with [γ-^{33}P]dATP (1,000 to 3,000 Ci/mmol; NEN Research Products, Boston, Mass.) and the double-stranded DNA cycle sequencing kit (BRL, Life Technologies, Inc., Gaithersburg, Md.).

Computer sequence analysis. The MoMLV sequence was obtained from GenBank (accession numbers J02255, J02256, and J02257). The programs FASTA and BLAST NCBI of the Genetics Computer Group sequence analysis package (6) were used to determine sequence homology. RNA secondary structure was predicted with the FOLD and MFOLD programs (40, 41) with the updated free energy parameters included in the Genetics Computer Group software package.

RESULTS

Forward mutation rate with the MSTK2 vector. This study was made to analyze the spectrum of mutations and the underlying mechanisms that may be responsible for the genetic rearrangements occurring during retrovirus replication as well as to confirm our previous determination of the adjusted mutation rate. The MoMLV-based vector MSTK2 was con-

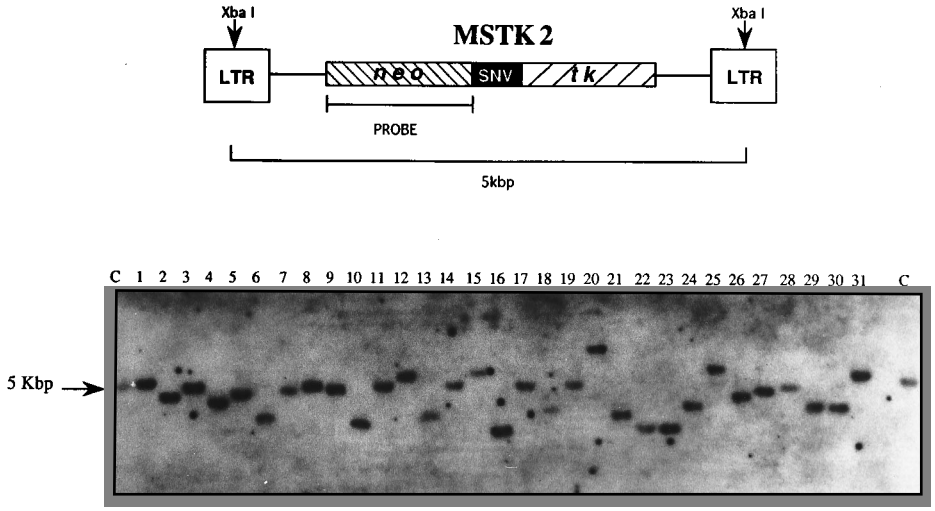

FIG. 2. Southern blot analysis of proviral DNA from G418+BrdU-resistant BRLtk$^-$ cell clones. The MSTK2 diagram represents the parental vector provirus. Genomic DNA was isolated from G418+BrdU-resistant BRLtk$^-$ cell clones infected with MSTK2 vector virus. Ten micrograms of genomic DNA from the different BRLtk$^-$ cell clones (lanes 1 to 31) was digested with *Xba*I and analyzed by Southern blotting with a *neo*-specific probe. As a control for full-length proviral DNA, 2 μg of genomic DNA isolated from a G418-resistant clone infected with MSTK2 vector virus and previously determined to be of wild-type size was also analyzed (lanes C). LTR, long terminal repeat.

structed to measure the mutation rate with the HSV *tk* gene as a reporter (Fig. 1A). Ten different packaging cell clones, each containing a single parental MSTK2 provirus, were established as previously described (36; see also Materials and Methods). Five were based upon GP+E-86, the MoMLV-based packaging cell line, and five were based upon PG13, a packaging cell line which produces MoMLV vector virus pseudotyped with GaLV envelope glycoproteins. These packaging cell clones served as the vector virus source that was used to infect the target BRLtk$^-$ cells. During this process, a single cycle of retroviral replication is completed: one round of proviral transcription by RNA polymerase II in the packaging cell clones to produce the vector virion RNA and one round of reverse transcription by the viral reverse transcriptase (RT) in the target BRLtk$^-$ cells (Fig. 1B). The *tk* mutation rate was measured by infecting BRLtk$^-$ cells and by separately obtaining the G418, HAT, and G418+BrdU titers (Fig. 1B and Table 1). G418 or HAT titers represent the overall virus titers, and the G418+BrdU titers reflect the mutant virus titer (Fig. 1B). Cells infected with the mutant virus titer represent clones that have a functional *neo* gene and an inactivated *tk* gene. The ratio of the mutant titer to the overall titer yields the mutation rate, that is, the mutation frequency in a single cycle of retroviral replication. The average rate of *tk* inactivation in the MSTK2 vector virus obtained from the 10 packaging cell clones was 0.088, or 8.8% (Table 1). As a control to rule out interference due to dual drug selection, experiments were performed with a helper cell clone harboring an MoMLV-based vector (N2) containing only the *neo* gene as a source of vector virus for the infection of BRLtk$^-$ cells. This vector did not yield any HAT-resistant colonies and its G418 and G418-BrdU titers were very similar (data not shown).

Southern blot analysis of MSTK2 mutant proviruses. Two packaging cell clones, each from GP+E-86 and PG13 cells harboring a single MSTK2 provirus, were used to generate a large number of mutant vector proviruses after one round of viral replication. Mutant proviruses were first analyzed by Southern blotting to identify those proviruses different in size from the parental MSTK2 provirus. Genomic DNA was isolated from individual G418+BrdU-resistant colonies, and Southern blotting was performed with a *neo*-specific probe. A total of 244 mutant proviruses were analyzed by Southern blotting, of which 22 were larger and 92 were smaller than the parental provirus and were presumably the results of genetic rearrangement. A representative Southern blot is depicted in Fig. 2. The remaining 130 were of parental size and likely the results of small mutations undetectable by Southern blotting. The gross rearrangements observed by Southern blotting are summarized in Table 2.

DNA sequence analysis of mutant proviruses. Different combinations of PCR primer pairs spanning the SNV promoter and *tk* locus were used to determine the mutation breakpoints in the mutant proviruses. PCR fragments of abnormal sizes defined by a particular primer pair were purified and directly sequenced. The sequences of 73 such mutant proviruses were obtained and are shown (Fig. 3, 4, 5, and 7). The mutants characterized were grouped according to the class of mutation they represented, and they are summarized in Table 3.

(i) Simple deletions. The most abundant class of mutation observed was the deletion of vector sequences (seen in 51 of 73 mutants) (Fig. 3). They involved both completely divergent sequences at the deletion junctions and regions of homology of up to 14 bp. Three of them were the result of independent but identical deletions involving a 7-bp repeat (Fig. 3, mutants 54.G3, 79.G3, and 46.G3). Several of the breakpoints were common to more than one mutant provirus. For example, coordinate 4366 is a breakpoint common to mutants 25.G2 and 37.P7, coordinate 3618 is a breakpoint common to 16.G2 and 77.G2, and coordinate 2490 is a breakpoint common to 5.P7 and 46.G2.

TABLE 3. Summary of genetic rearrangements and rates of mutation arising during a single cycle of MSTK2 replication

Class of mutation	No. of mutant proviruses sequenced	Mutation rate/10^5 bp/replication cycle	Mutation rate (%)/kbp/replication cycle
Simple deletions[a]	51	0.97	0.97
Deletions with insertions[a]	8	0.15	0.15
Duplications[b]	7	0.30	0.30
Complex hypermutations[a]	7	0.13	0.13
Overall mutation rate[c]	73	1.55	1.55

[a] The mutation rate for this class equals the overall mutation rate (see footnote *c*) $\times$ $n/73 \times 0.893$, where n is the number of mutant proviruses in this class and 0.893 is the ratio of the percentage of mutants that were smaller than the parental vector as determined by Southern blotting ($92/114 \times 100 = 80.7\%$) to the percentage of mutants that were smaller as determined by DNA sequencing ($66/73 \times 100 = 90.41\%$). The conversion value of 0.893 was used in the mutation rate calculation to prevent overrepresentation of these mutations.

[b] The mutation rate for this class equals the overall mutation rate $\times$ $n/73 \times 2.01$, where n refers to the number of mutant proviruses in this class identified by DNA sequencing and 2.01 is the ratio of the percentage of mutants that were larger than the parental vector as ascertained by Southern blotting ($22/114 \times 100 = 19.3\%$) to the percentage of mutants that were larger as determined by DNA sequencing ($7/73 \times 100 = 9.6\%$). Multiplication by the conversion value 2.01 was necessary to avoid underrepresentation of this class of mutation since only 7 of the 22 mutants larger than the parental mutant proviruses were sequenced. This assumes that the mutations occurring in larger proviruses are due to the duplication of vector sequences.

[c] The overall mutation rate equals $0.088/2{,}620 \times 114/244$, where 0.088 is the rate of *tk* inactivation, 2,620 is the size of the target gene, and 114/244 is the number of proviruses with rearrangements over the total number of proviruses analyzed.

(ii) Deletions combined with insertions. Ten of seventy-three mutants had deletions ranging in size from 155 to 2,004 bp with short insertions at the deletion junctions (Fig. 4). Most of the insertions were short sequences ranging from 1 to 8 nucleotides in length. Two of the mutants in this class were distinct from the other mutants with deletions and insertions, as they involved more complex alterations. Both mutant 56.G2 and mutant 65.G2 appear to have undergone a combination of a small deletion of 129 and 5 bp, respectively, and a deletion combined with the insertion of a sequence for which RNase H cleavage products were apparently incorporated to serve as the templates. The site of insertion of these fragments involved both divergent and homologous regions.

(iii) Duplications. Southern blotting indicated that 22 of 244 mutant proviruses were larger than the parental MSTK2 provirus. On the basis of the sequence analysis of 73 mutant proviruses, 7 were seen to have formed as a result of duplications ranging from 40 to 800 bp in length (Fig. 5 and 6). About half of these mutants (four of seven) contained short direct repeats (2 to 6 bp) at the breakpoints (Fig. 5). All seven mutant proviruses had the duplicated sequence within the *tk* gene, and none involved the SNV promoter.

(iv) Complex hypermutations. Seven of seventy-three mutant proviruses displayed a combination of mutations near the deletion junctions. These involved large deletions combined with base substitutions, small deletions, and/or insertions of a few nucleotides near the deletion junctions (Fig. 7). Two of the mutant proviruses also had a junctional insertion of 8 bp each (mutants 56.G3 and 45.G2 [Fig. 4 and 7]). The deletions ranged from 0.4 to about 1.6 kbp. The mutant 72.P8 exhibited

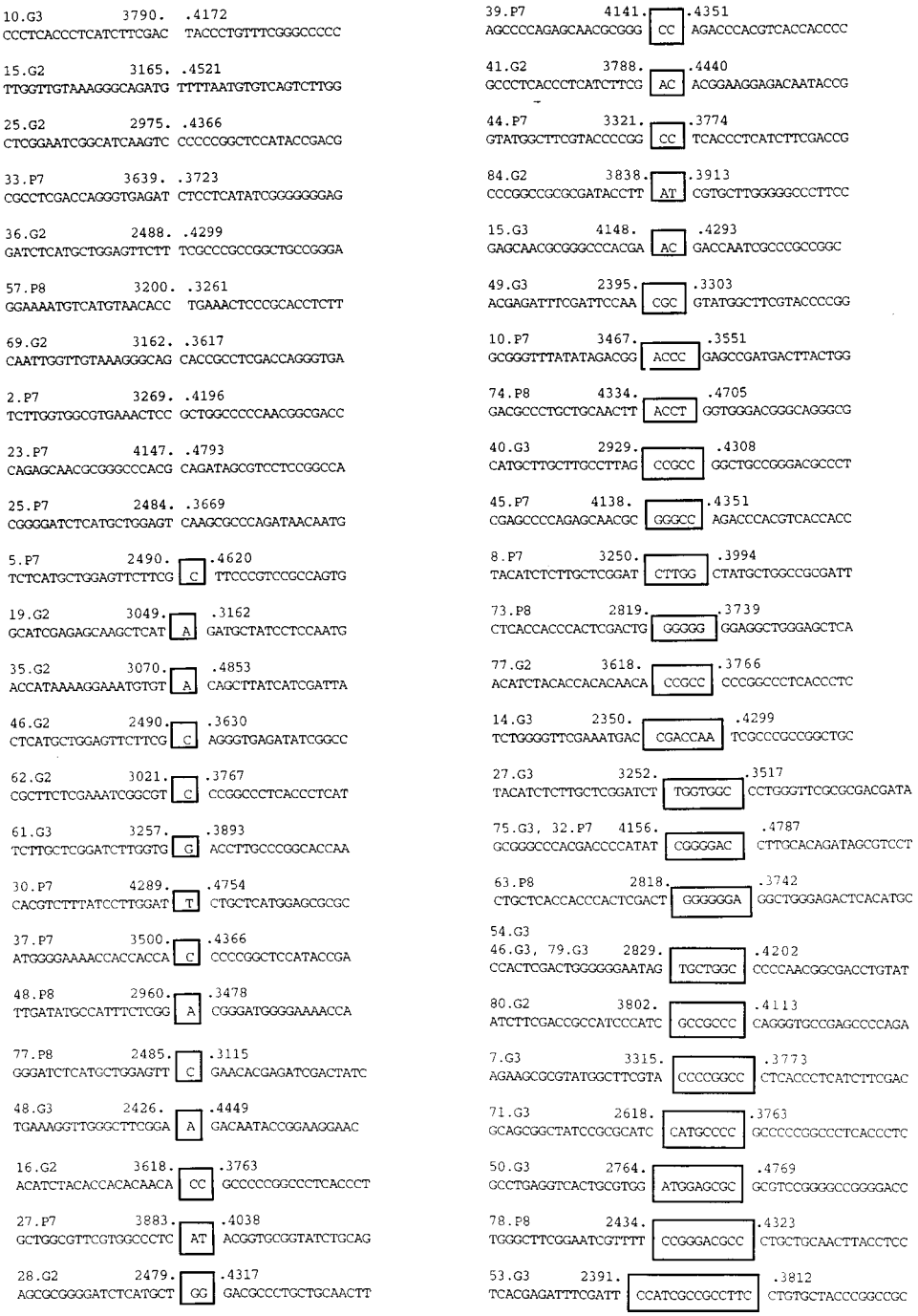

FIG. 3. Sequences of mutant proviruses arising as a result of simple deletions. The first half in the nomenclature of the mutant proviruses represents the clone number, and the second half represents the producer cell clone that was used to infect the target cell. G2 and G3 correspond to GP+E-86 clones 2 and 3 from Table 2, and P7 and P8 are PG13 clones 7 and 8 from Table 2, respectively. Sequences shown are those with simple deletion of vector sequences. The coordinates of the deletion breakpoints are given above the sequences. Sequences within boxes represent sequences common to both recombination points.

all of the combined classes of mutations within a span of 25 bp from the deletion junction.

RNA secondary structure analysis. Mapping all the 5′ and 3′ junctions along the target sequence revealed some clusters of mutation breakpoints. To ascertain if the RNA secondary structure played a role, 150 breakpoints from 73 mutants, schematically depicted in Fig. 8, were mapped to the predicted secondary structure of the MSTK2 vector virus sequence. With the programs FOLD and MFOLD, several overlapping fragments of various window sizes were analyzed with squiggles, circles, mountains, and domes as output for the display of the breakpoints. A squiggle output that corresponds to the coordinates 3720 to 4075 is shown in Fig. 9. Despite 34% of the secondary structure being single stranded, 80% (120 of 150) of all breakpoints were located in single-stranded regions (data not shown). The distribution of the breakpoints was analyzed

FIG. 4. Mutant provirus sequences in which vector sequences were deleted and short sequences were inserted at the deletion junctions. The insertions are shown within brackets with a plus sign. The underlined sequences represent sites of base substitution mutations; the wild-type base is shown above each mutated base. The boxes enclose overlapping sequences that are present at both breakpoints in the wild-type sequence; only one copy remains in the mutant provirus. The coordinates on either side of the boxes represent recombination events. The coordinate on the left marks the 5′ and that on the right marks the 3′ recombination junction. The sequences within the boxes indicate homology at the breakpoints. The ellipsis in 56.G2 is to indicate a segment of the sequence that is not shown.

for statistical significance. The probability of 80% of all breakpoints occurring randomly in single-stranded regions is <0.00001.

DISCUSSION

In the absence of any obvious hot spots such as direct repeats, the rate of *tk* inactivation measured in one round of retroviral replication was 8.8% per replication cycle with the MSTK2 vector. This rate concurred with our previous studies (36). The target size was 2.62 kbp and consisted of the SNV U3 promoter, the *tk* gene, and flanking sequences that are not required for replication or for Neo protein production; thus, the rate of inactivation per kilobase was 3.3%. Since 114 of the 244 mutant proviruses analyzed by Southern blotting were the results of gross rearrangements, the rate of genetic rearrangement per base pair was 1.55×10^{-5}/bp per replication cycle (Table 3). The specific types of mutations observed were simple deletions involving either short regions of homology or regions of no homology at the breakpoints, deletions with insertions, duplications, and "complex hypermutations." Mutation rates are given for each class of mutation (Table 3). However, it should be noted that the rates shown in Table 3 represent minimum mutation rates, since it is possible that mutations arose which did not inactivate the *tk* gene, so they would not have been scored.

At what step do the mutations occur? Theoretically, mutations can arise during transcription by RNA polymerase II, during proviral replication by cellular DNA polymerases, or during reverse transcription by the viral RT. It is not likely that mutations occurring during cellular DNA replication contribute significantly to the measured mutation rates because rates of mutation at this step are orders of magnitude lower than the observed rates (15). Although it cannot be formally excluded that the genetic rearrangements observed occur during RNA transcription, it is more likely they are introduced by a relatively nonprocessive enzyme. Therefore, it is likely they occur during reverse transcription, since RT is not particularly processive owing to its need to switch templates at least twice during normal replication. Moreover, this low degree of processivity of RT is supported by numerous experiments in vitro using purified RT (1, 2, 9, 14, 28, 29, 31, 32).

RNA secondary structure. An inspection of the local primary sequence at the sites of breakpoints did not reveal any underlying common theme, although sequence context is likely to play a role in determining pausing and/or dissociation. However, there was significant correlation between the breakpoints and single-stranded regions in the predicted RNA secondary structure with the Genetics Computer Group programs FOLD and MFOLD. Although these programs have limitations and it is uncertain that the predicted RNA secondary structure accurately reflects that of an RNA undergoing reverse transcription in a viral nucleoprotein core, there was a correlation between the location of mutations and the predicted RNA structure. The majority of the mutations mapped in regions corresponding to predicted hairpin loops, internal loops, and bulges (Fig. 9), which correspond to bases that are unpaired.

Analysis of the distribution of the breakpoints showed that the mutations were clustered at the loops and bulges and at the boundaries between stems and bulges (border bases). The clustering was statistically significant. One hundred twenty of one hundred fifty breakpoints fell within single-stranded regions, while single-stranded bases constituted only 34% of the RNA structure of the vector. Assuming no relationship between the

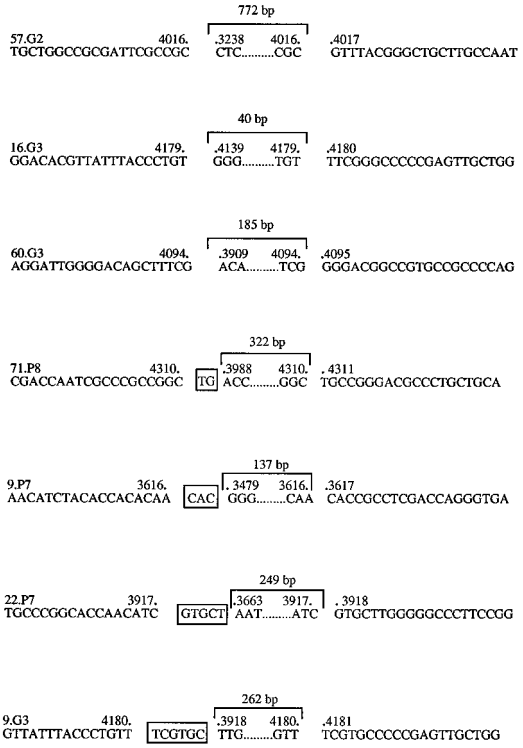

FIG. 5. Sequences of mutant proviruses that were larger than parental vector MSTK2 because of a duplication of vector sequences. The coordinates within the brackets represent the duplicated sequences. The length of the duplicated sequence for each mutant provirus is noted above the bracket. The sequences within the boxes represent regions of homology.

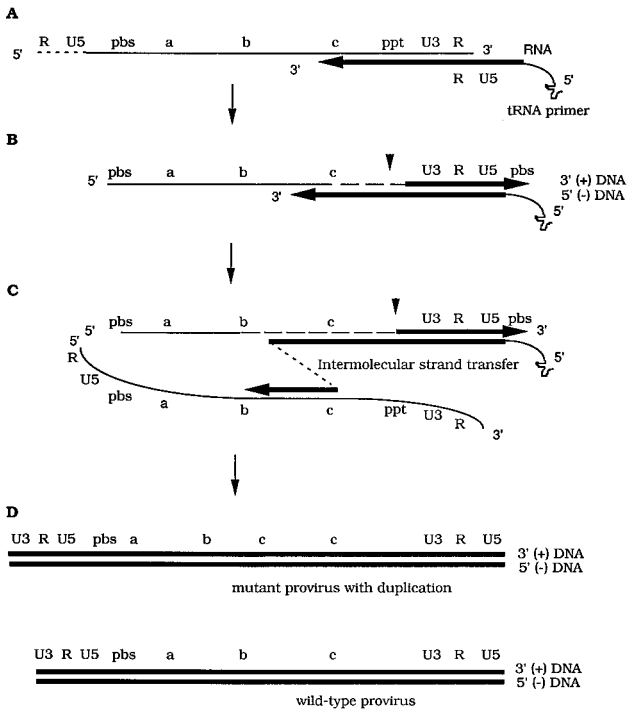

FIG. 6. Proposed model for the generation of mutant proviruses with duplications. Thin lines represent RNA, thick lines represent DNA. The arrowheads mark RNase H cleavage sites for primer initiation of plus-strand DNA synthesis. a, b, and c, different loci 5′ to 3′; pbs, primer binding site; ppt, polypurine tract; R, repeat. U3 and U5 represent the U3 and U5 regions of MLV. (A) Synthesis of minus-strand DNA subsequent to the first strand jump. (B) Pausing or dissociation of RT on the nascent DNA minus strand. (C) Switching of the nascent strand to a second template at a site downstream of its original location, which most likely is an intermolecular jump, as the original strand would have been degraded by RNase H activity. (D) Completion of the synthesis of plus and minus DNA strands. The resultant mutant is larger than the parental provirus because of the duplication of vector viral sequences.

locations of the mutations and the RNA structure (the null hypothesis), the probability of obtaining the observed distribution (i.e., 120 of 150 breakpoints in single-stranded regions by random occurrence) is <0.00001. This analysis exhibited a highly significant clustering of mutations in unpaired regions, which is consistent with the work of Le et al. (16), who observed a correlation between hypervariable regions of the human immunodeficiency virus envelope protein and nonhelical regions of the RNA structure predicted for the envelope.

If RNA secondary structure plays an important role in genetic rearrangements, this would suggest that a majority of mutations occur during minus-strand synthesis. One could envisage that the single-stranded regions are more susceptible to modifications such as cleavage, which would favor genetic rearrangements via the forced copy-choice mechanism (5, 10). Furthermore, if the genetic rearrangements occurred during plus-strand synthesis, the provirus would have a bulging single-strand loop corresponding to the size of the deletion obtained, which could be greater than 2 kbp in some cases. This would seem to be a difficult structure to resolve, and to our knowledge, an activity which could repair such a structure has not been described. However, it cannot be ruled out that many of the mutations occur during plus-strand synthesis (3, 8).

Simple deletions. The 31 simple deletions with homology at the breakpoints are likely to have arisen by misalignment of the growing DNA endpoint with the homologous downstream sequence after the copying of the first repeat by RT (26, 30, 34, 36). This results in the deletion of intervening sequences together with one of the repeats. These mutations indicate that the retroviral RT is capable of strand transfer between templates with little or no sequence similarity. Deletions of sequences flanked by divergent sequences with only one base in common have also been observed in other retroviruses (23, 30, 38, 39). However, the rate of 1.09×10^{-5} for this class of mutation is high compared with what has been observed by others for bovine leukemia virus and SNV, namely, 4×10^{-7} and 2×10^{-6}/bp per replication cycle, respectively (18, 25, 26). This discrepancy may be due to properties of the different RTs involved, sequence effects, or a combination of both.

Deletions with insertions. The generation of all the short insertions can be accounted for by either the nontemplated addition of nucleotides (24), particularly those from 1 to 3 nucleotides in length, or the use of a short template from a different portion of the viral genome, such as a short RNase H digestion product, during minus-strand synthesis, or a combination of both events. Deletions with insertions of RNase H cleavage products have also been described previously for SNV-based vectors (26, 34). Alternatively, it is possible that the recombination occurred during plus-strand synthesis, but then it is unclear if a large looped-out single-stranded structure equivalent in length to that of the large deletions is efficiently repaired on the minus-strand DNA.

Of the nine mutant proviruses with insertions, none had transduced cellular sequences. These results were unexpected because in our previous study with an almost identical vector (36), a sequence examination of seven mutant proviruses demonstrated that three had transduced sequences from the producer cell line ranging from 30 to 854 bp. The disparity between the two vectors is currently being further investigated to

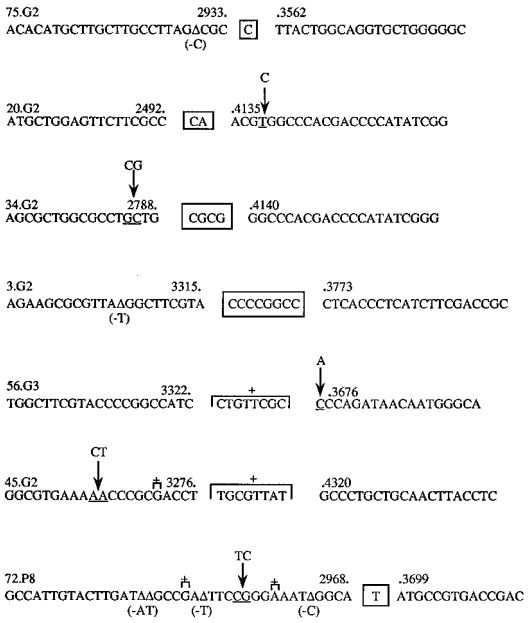

FIG. 7. Mutant proviruses with complex hypermutations. The triangles represent sites of frameshifts created by the deletions of the bases shown in parentheses below the mutant sequence. The underlined bases are sites of base substitution mutations; the base(s) of the parental sequence is shown above an arrow pointing to the mutated base. Frameshifts created by the insertion of nucleotides are shown within brackets surmounted by a plus sign. Sequences within boxes represent regions of homology. Coordinates denote the 5′ and 3′ deletion junctions. Mutants 45.G2 and 56.G3 are shown in Fig. 5 as well since they represent both classes of mutations.

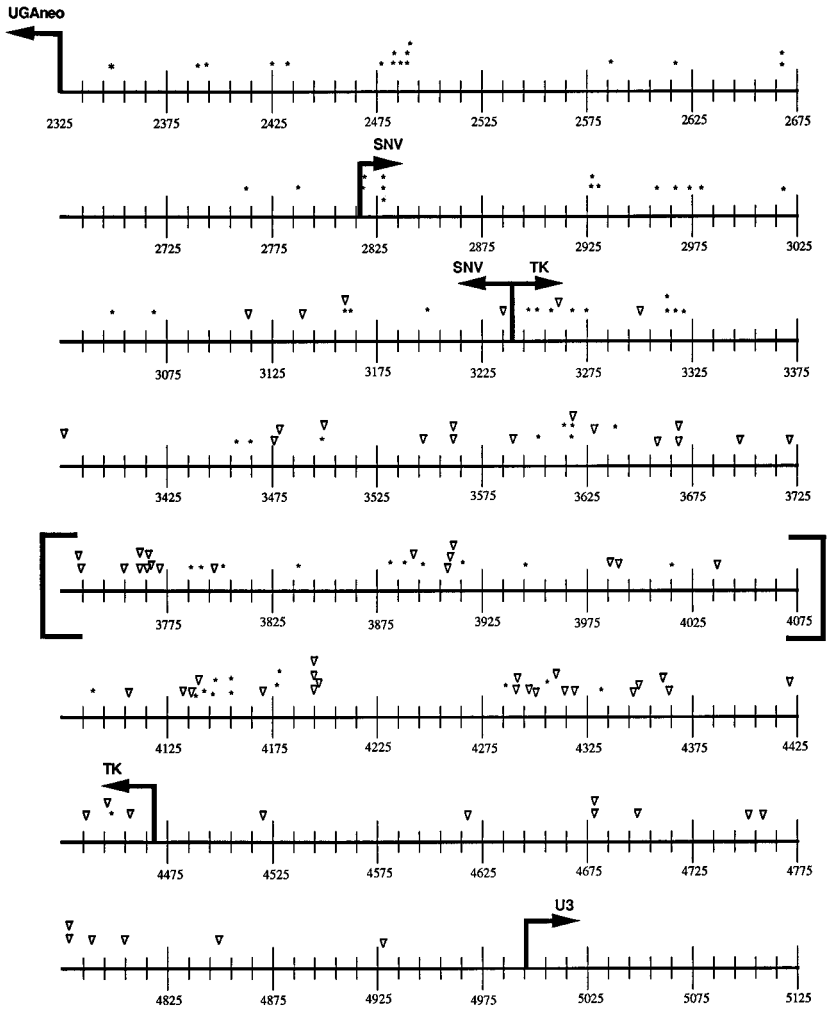

FIG. 8. Schematic representation of the distribution of the 5′ and 3′ recombination points on the parental MSTK2 provirus. The asterisks and downward-pointing open triangles represent 5′ and 3′ junctions, respectively. Bent arrows delimit various regions. UGAneo, the translational stop codon of the neomycin phosphotransferase gene; SNV, the SNV promoter; TK, the HSV *tk* gene; U3MLV, the U3 region of the MLV 3′ long terminal repeat. The coordinates correspond to those of the MSTK2 vector. The brackets enclose the region depicted in Fig. 9 for the squiggle output.

help shed light upon the mechanism of retrovirus-mediated transduction of cellular sequences and the evolution of acutely transforming retroviruses.

Duplications. Duplications involving a single recombination event were also detected. The duplicated sequences ranged from 40 to 772 bp (Fig. 5). If they occurred during minus-strand synthesis, they would most likely result from an intermolecular strand transfer to the second vector RNA molecule present in the virus particle, because the corresponding RNA template on the initial template would have already been digested by the RT-associated RNase H activity (Fig. 6) (4, 27).

Complex hypermutations. Previously, hypermutations have been described for both SNV (26) and human immunodeficiency virus type 1 (37). In both cases, there were a number of G to A transitions over a short stretch of sequence. With MSTK2, mutations distinct from the G to A hypermutations, termed complex hypermutations, were observed. Two potential mechanisms that have been proposed for the G to A transitions may also account for these complex mutations: (i) a mutator RT, the product of a translational error, may be engaged for a period of unfaithful polymerization before being replaced by a normal RT and after succeeding a normal RT that dissociated from the template primer complex, possibly at a pause site, or (ii) normal RT may enter a highly error-prone metastable state that is maintained over the mutated stretch of DNA (34). It has been proposed that this could be the underlying cause of mutational events during retroviral replication in general. According to the hypothesis, RT enters a metastable state as a consequence of a pause during nonprocessive polymerization. The outcome of such a state can be continued polymerization at the next template nucleotide, in which case no mutation is produced, or strand transfer and polymerization at a different location, which would generate mutations of different classes (34). If such a phenomenon is at work in the complex mutations observed in our study, the implication is that the metastable state can be maintained under certain circumstances for a period long enough to give rise to clustered mutations of different classes.

Implications for use of MoMLV-based vectors. What are the consequences of these results with respect to the utilization of MLV-based vectors for gene transfer? First, one should consider whether these results can be generalized to other exogenous genes. Formally, they cannot; however, we suggest that they might be a useful predictor, at least within a few fold, of

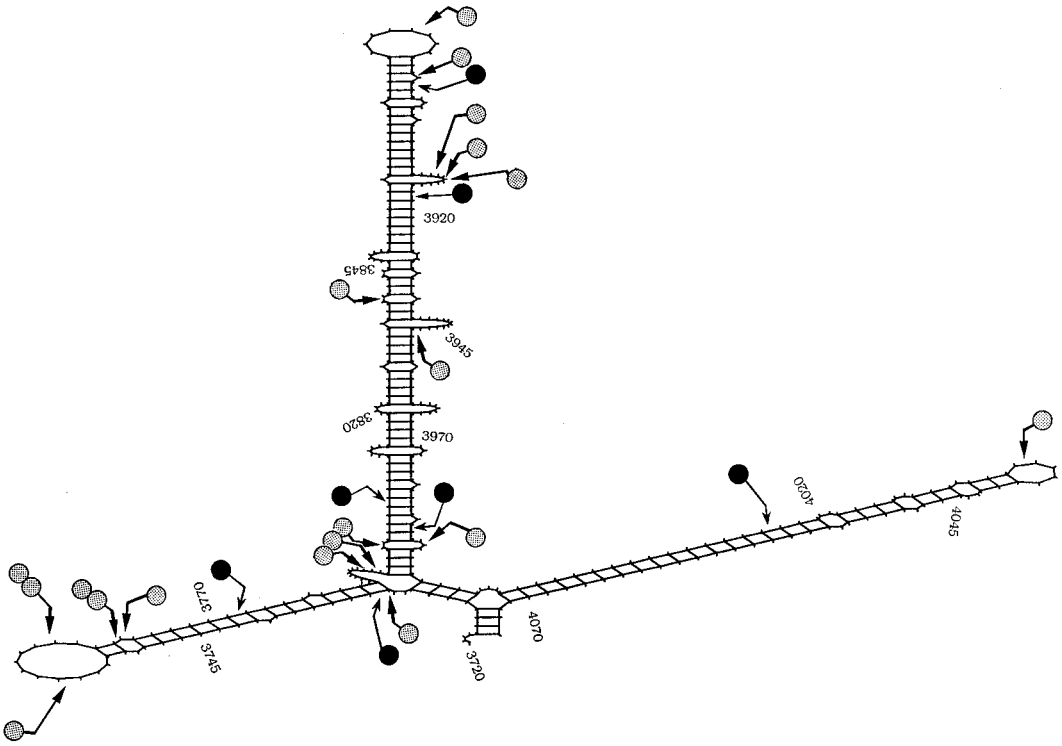

FIG. 9. Squiggle output from FOLD program of the Genetics Computer Group package between coordinates 3720 and 4075 of the wild-type MSTK2 provirus sequence. Each circle with an extended arrow indicates a recombination event (5′ or 3′), and the number of circles represents the number of mutant proviruses at each breakpoint. The stippled and filled circles represent mutants with breakpoints in single-stranded and double-stranded regions, respectively.

the rates of gene inactivation that can be expected for other genes, especially considering the mutation rate was obtained in the absence of sequences that would contribute to an abnormally high mutation rate such as long direct repeats, cryptic splice sites, and long homopolymeric stretches of nucleotides.

The consequences of the observed rates and spectra of mutations most likely depend on specific applications. For example, in the case of gene replacement therapies, one would anticipate that the majority of transferred genes would be intact. Thus, as long as the levels of efficiency of gene transfer and of gene expression are appropriate, the transferred gene is likely to have an important therapeutic benefit. This seems to be true for most protocols. On the other hand, we suggest that considerable prudence be exercised when utilizing genes involved in signal transduction or cell cycle control, as activating mutations with oncogenic potential might be introduced into such genes even after a single cycle of replication. Of particular concern are the complex hypermutations with the potential to introduce multiple mutations into a single gene in one cycle of replication.

ACKNOWLEDGMENTS

S. Parthasarathi and A. Varela-Echavarría contributed equally to this study.

We thank Arnold B. Rabson, Stuart W. Peltz, Jie Gu, Hong Yu, and Malvika Kaul for discussion and helpful comments on the manuscript. We also thank David Harrison and Seshu Pedapudi for helping with the RNA secondary structure analysis. Experimental assistance provided by Michael Correa is appreciated as well.

This work was supported by grants RO1 CA50777 and 5RO1 AI 3483 from the National Institutes of Health and grant 3923 from the Council for Tobacco Research to J.P.D.

REFERENCES

1. **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.
2. **Bebenek, K., and T. A. Kunkel.** 1993. The fidelity of retroviral reverse transcriptases, p. 85–102. *In* A. M. Skalka and S. P. Goff (ed.), Reverse transcriptase. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
3. **Boone, L. R., and A. M. Skalka.** 1993. Strand displacement synthesis by reverse transcriptase, p. 119–113. *In* A. M. Skalka and S. P. Goff (ed.), Reverse transcriptase. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
4. **Champoux, J. J.** 1993. Roles of ribonuclease H in reverse transcription, p. 103–117. *In* A. M. Skalka and S. P. Goff (ed.), Reverse transcriptase. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
5. **Coffin, J. M.** 1990. Genetic variation in retroviruses, p. 11–33. *In* E. Kurstak, R. G. Marusyk, F. A. Murphy, and M. H. V. Van Regenmortel (ed.), Applied virology research, virus variability, epidemiology and control. Plenum Press, New York.
6. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
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. **Hu, W.-S., and H. M. Temin.** 1990. Retroviral recombination and reverse transcription. Science **250:**1227–1233.
9. **Ji, J. P., and L. A. Loeb.** 1992. Fidelity of HIV-1 reverse transcriptase copying RNA in vitro. Biochemistry **31:**954–958.
10. **Katz, R. A., and A. M. Skalka.** 1990. Generation of diversity in retroviruses. Annu. Rev. Genet. **24:**409–445.
11. **Katz, R. A., and A. M. Skalka.** 1994. The retroviral enzymes. Annu. Rev. Biochem. **63:**133–173.
12. **Kawai, S., and M. Nishizawa.** 1984. New procedure for DNA transfection with polycation and dimethyl sulfoxide. Mol. Cell. Biol. **4:**1172–1174.
13. **Keller, G., C. Paige, E. Gilboa, and E. F. Wagner.** 1985. Expression of a foreign gene in myeloid and lymphoid cells derived from multipotent haematopoietic precursors. Nature (London) **318:**149–154.
14. **Klarmann, G. J., C. A. Schauber, and B. D. Preston.** 1993. Template-directed pausing of DNA synthesis by HIV-1 reverse transcriptase during polymerization of HIV-1 sequences in vitro. J. Biol. Chem. **268:**9793–9802.

15. **Kunkel, T. A.** 1992. DNA replication fidelity. J. Biol. Chem. **267:**18251–18254.
16. **Le, S.-Y., J.-H. Chen, D. Chatterjee, and J. V. Maizel.** 1989. Sequence divergence and open regions of RNA secondary structures in the envelope regions of the 17 human immunodeficiency isolates. Nucleic Acids Res. **17:**3275–3288.
17. **Leider, J. M., P. Palese, and F. I. Smith.** 1988. Determination of the mutation rate of a retrovirus. J. Virol. **62:**3084–3091.
18. **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.
19. **Mansky, L. M., and H. M. Temin.** 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J. Virol. **69:**5087–5094.
20. **Markowitz, D., S. Goff, and A. Bank.** 1988. A safe packaging line for gene transfer: separating viral genes on two different plasmids. J. Virol. **62:**1120–1124.
21. **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.
22. **Monk, R. J., F. G. Malik, D. Stokesberry, and L. H. Evans.** 1992. Direct determination of the point mutation rate of a murine retrovirus. J. Virol. **66:**3683–3689.
23. **Omer, C. A., K. Pogue-Geile, R. Guntaka, K. A. Staskus, and A. J. Faras.** 1983. Involvement of directly repeated sequences in the generation of deletions of the avian sarcoma virus *src* gene. J. Virol. **47:**380–382.
24. **Patel, P. H., and B. D. Preston.** 1994. Marked infidelity of human immunodeficiency virus type 1 reverse transcriptase at RNA and DNA template ends. Proc. Natl. Acad. Sci. USA **91:**549–553.
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. **Peliska, J. A., and S. J. Benkovic.** 1992. Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. Science **258:**1112–1118.
28. **Preston, B. D., and N. Garvey.** 1992. Retroviral mutation and reverse transcriptase fidelity. Pharm. Technol. **16:**34–52.
29. **Preston, B. D., B. J. Poiesz, and L. A. Loeb.** 1988. Fidelity of HIV-1 reverse transcriptase. Science **242:**1168–1171.
30. **Pulsinelli, G. A., and H. M. Temin.** 1991. Characterization of large deletions occurring during a single round of retrovirus vector replication: novel deletion mechanism involving errors in strand transfer. J. Virol. **65:**4786–4797.
31. **Roberts, J. D., K. Bebenek, and T. A. Kunkel.** 1988. The accuracy of reverse transcriptase from HIV-1. Science **242:**1171–1173.
32. **Roberts, J. D., B. D. Preston, L. A. Johnston, A. Soni, L. A. Loeb, and T. A. Kunkel.** 1989. Fidelity of two retroviral reverse transcriptases during DNA-dependent DNA synthesis in vitro. Mol. Cell. Biol. **9:**469–476.
33. **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.
34. **Temin, H. M.** 1993. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. Proc. Natl. Acad. Sci. USA **90:**6900–6903.
35. **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. J. Biol. Chem. **34:**24681–24688.
36. **Varela-Echavarría, A., C. M. Prorock, Y. Ron, and J. P. Dougherty.** 1993. High rate of genetic rearrangement during replication of a Moloney murine leukemia virus-based vector. J. Virol. **67:**6357–6364.
37. **Vartanian, J.-P., A. Meyerhans, B. Åsjö, and S. Wain-Hobson.** 1991. Selection, recombination, and G→A hypermutation of human immunodeficiency virus type 1 genomes. J. Virol. **65:**1779–1788.
38. **Voynow, S. L., and J. M. Coffin.** 1985. Truncated *gag*-related proteins are produced by large deletion mutants of Rous sarcoma virus and form virus particles. J. Virol. **55:**79–85.
39. **Voynow, S. L., and J. M. Coffin.** 1985. Evolutionary variants of Rous sarcoma virus: large deletion mutants do not result from homologous recombination. J. Virol. **55:**67–78.
40. **Zuker, M., J. A. Jaeger, and D. H. Turner.** 1991. A comparison of optimal and suboptimal RNA secondary structures predicted by free energy minimization with structures determined by phylogenetic comparison. Nucleic Acids Res. **19:**2707–2714.
41. **Zuker, M., and P. Steigler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxillary information. Nucleic Acids Res. **9:**133–148.