JOURNAL OF VIROLOGY, Nov. 1993, p. 6357–6364
0022-538X/93/116357-08$02.00/0

Vol. 67, No. 11

High Rate of Genetic Rearrangement during Replication of a Moloney Murine Leukemia Virus-Based Vector

ALFREDO VARELA-ECHAVARRÍA, CHRISTINE M. PROROCK,† YACOV RON, AND JOSEPH P. DOUGHERTY*

Department of Molecular Genetics and Microbiology, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, 675 Hoes Lane, Piscataway, New Jersey 08854-5635

Received 5 May 1993/Accepted 26 July 1993

A protocol was designed to measure the forward mutation rate over an entire gene replicated as part of a Moloney murine leukemia virus-based vector. For these studies, the herpes simplex virus thymidine kinase (*tk*) gene under the control of the spleen necrosis virus U3 promoter was used as target sequence since it allows selection for either the functional or the inactivated gene. Our results indicate that after one round of retroviral replication, the *tk* gene is inactivated at an average rate of 0.08 per cycle of replication. Southern blotting revealed that the majority of the mutant proviruses resulted from gross rearrangements and that deletions of spleen necrosis virus and *tk* sequences were the most frequent cause of the gene inactivation. Sequence analysis of the mutant proviruses suggested that homologous as well as nonhomologous recombination was involved in the observed rearrangements. Some mutations consisted of simple deletions, and others consisted of deletions combined with insertions. The frequency at which these mutations occurred during one cycle of retroviral replication provides evidence indicating that Moloney murine leukemia virus-based vectors may undergo genetic rearrangement at high rates. The high rate of rearrangement and its relevance for retrovirus-mediated gene transfer are discussed.

Retroviruses are RNA viruses that evolve at rapid rates (3, 13). Several studies have demonstrated that retroviral base substitution mutation rates are several orders of magnitude higher than those of cellular sequences (6, 18, 22, 26, 40). In addition to generating diversity by base substitution mutations, retroviruses are subject to genetic rearrangements involving deletions (24, 41, 42) or duplications (1, 36) of viral elements and to recombination with cellular sequences (29, 39) or other retroviral genomes present in the same cell (2, 11, 12, 38). This characteristically imperfect replication process is responsible for the generation of genetic diversity which is presumably advantageous for responding to selective forces (3, 13).

As a result of mutational events, defective retroviral genomes have been found to arise spontaneously during retroviral replication. For instance, after replication in culture, replication-defective variants of Rous sarcoma virus (41, 42) and Moloney murine leukemia virus (MoMLV) (34) as well as nontransforming variants of avian sarcoma virus (24) have been observed. An indication that these defective retroviral genomes might be generated at a high frequency came from the finding that about one-third of the viral progeny produced by clonal stocks of MoMLV was replication defective (34). In that study, however, the precise nature of the mutations leading to the defects was not determined, and the possibility that defective genomes were already present in the virus producer cells could not be ruled out. The analysis of genomic RNA derived from some of the defective MoMLV proviral clones suggested that at least one in five was the result of deletion of viral sequences. Other defective genomes resulting from deletion of part of the viral genes were also detected during passage of Rous sarcoma virus in cultured cells (41, 42), but their rate of appearance was not determined.

An inherent difficulty of using replication-competent retroviruses for determining mutation rates, the mutation frequency per replication cycle, is posed by their ability to spread throughout a culture. However, vector systems that facilitate the determination of mutation rates in the absence of helper viruses and allow the detailed analysis of mutations that occur in a single cycle of retroviral replication have been developed. Spleen necrosis virus (SNV)- and MoMLV-based vector systems have been used to determine base substitution mutation rates (6, 26, 27, 40), but studies concerning other types of mutations, such as genetic rearrangements of different types, have been performed mainly with SNV-based vectors. In one such study, a forward mutation rate was determined by using the *lacZ*α gene (280 bp) as a reporter gene in an SNV-based vector, and it was found that deletions occur at a rate of 2×10^{-6}/bp per cycle (27). Deletions combined with insertions of up to 34 bp of sequences present in the parental vector were found to occur at a similar rate.

In this study, a protocol was designed to measure the forward mutation rate over an entire gene carried within an MoMLV-based vector. This is the first time that a forward mutation rate analysis has been performed with use of an MoMLV-based system. The herpes simplex virus thymidine kinase (*tk*) gene was used as the reporter gene because conditions are available to select for either the functional or the inactivated gene and because the *tk* gene is potentially relevant for some gene therapy procedures (20). Moreover, since the size of the potential target locus is 2.7 kbp, it is possible to efficiently score for gross rearrangements as well as for base substitution mutations, frameshifts, and small insertions and deletions. It should be noted that the 2.7-kbp locus includes the *tk* gene, the SNV U3 promoter directing its expression, and flanking sequences that can be involved in rearrangements without affecting viral replication or expression of the other exogenous gene in the vectors. Our results indicate that the *tk* gene is inactivated in about 8% of the MoMLV proviruses after a single cycle of replication. South-

* Corresponding author.
† Present address: Center for Advanced Biotechnology and Medicine, Piscataway, NJ 08854-5638.

ern blotting and DNA sequence analyses revealed that the majority of the inactivating mutations involved deletion of *tk* coding sequences and the promoter driving its expression. Moreover, the deletions arose from both homologous and nonhomologous recombination events. In some mutant proviruses, the rearrangements resulted in deletion of vector sequences combined with the insertion of sequences of unknown origin. These results suggest that at least some MoMLV vectors can undergo genetic rearrangement at a high rate.

In addition to constituting a step toward the understanding of the principles underlying rates and mechanisms of mutation in retroviruses, this study provides information pertaining to the stability of retroviral vectors, an important consideration for their use as vehicles for retrovirus-mediated gene transfer.

MATERIALS AND METHODS

DNA constructions. Plasmid pN2 has been described elsewhere (15). The vector pMSTK was constructed in two steps. First, the *Bgl*II-*Bam*HI fragment containing the herpes simplex virus *tk* gene from pTKΔterR (35) was blunt-end ligated to the *Ava*I fragment of pME111 (7), which contains the SNV promoter (U3 region of the long terminal repeat [LTR]), yielding pSNV*tk*. The *Eco*RI-*Ppu*MI fragment from pSNV*tk* which contains the SNV promoter-*tk* cassette was cloned into the *Xho*I site of pN2 by blunt-end ligation to yield the pMSTK vector.

Cells. GP+E-86 is an NIH 3T3-derived helper cell line which expresses the MoMLV proteins required for the propagation of replication-defective MoMLV-based vectors without producing replication-competent virus (19). PG13 is another NIH 3T3-derived helper cell line which expresses the MoMLV Gag and Pol proteins as well as the Gibbon ape leukemia virus (GaLV) *env* gene products (21). This cell line can be used to propagate MoMLV-based vector virus with the GaLV receptor specificity (21). The BRLtk$^-$ cell line is a bromodeoxyuridine (BrdU)-resistant buffalo rat liver cell line permissive for infection by MoMLV vector virus and MoMLV vector virus pseudotyped with the GaLV Env glycoproteins.

Transfections and infections. Transfections were performed by the Polybrene-dimethyl sulfoxide method (14). Infections were done 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 infection of BRLtk$^-$ cells with 10-fold serial dilutions of virus stocks. At 24 h after infection, the cells were placed in medium containing G418 (300 μg/ml), HAT (0.1 mM hypoxanthine, 0.5 μM aminopterin, 30 μM thymidine), or both G418 and BrdU (300 and 30 μg/ml, respectively). Titers were obtained by multiplying the number of resistant colonies by the dilution factor.

Establishment of helper cell clones containing a single parental provirus. To establish helper cell clones containing a single parental provirus, the following steps were performed. (i) GP+E-86 helper cells were transfected with vector plasmid DNA (pMSTK) and placed in selection medium containing G418 (300 μg/ml). (ii) Vector virus was harvested from the resulting G418-resistant cells and used to infect either fresh GP+E-86 or PG13 helper cells at low multiplicity of infection as previously described (40) to obtain G418-resistant cell clones harboring a single provirus. (iii) To determine which helper cells harbored parental MSTK proviruses, HAT and G418 titers were obtained by infection of BRLtk$^-$ cells. The helper cell clones that yielded no HAT-resistant cells were not used for mutation rate determinations because they probably represent clones with defective proviruses generated during transfection of GP+E-86 cells or infection of either GP+E-86 or PG13 cells. Thus, the procedure described above ensures that only helper cell clones harboring a single parental vector provirus are used in the mutation rate analysis.

Southern blot analysis. Southern blotting was performed by standard protocols (32). Briefly, 10 μg of genomic DNA from the infected BRLtk$^-$ clones of interest was digested with *Xba*I, electrophoresed in 1% agarose gels, and blotted onto nitrocellulose. The blots were then hybridized with a neomycin phosphotransferase (*neo*) gene-specific ^{32}P-labeled probe (*Eco*RI-*Xho*I fragment from pN2).

PCR and sequencing. Proviral DNA was amplified from genomic DNA of infected cell clones by the polymerase chain reaction (PCR) (33), using *Taq* polymerase (Perkin-Elmer-Cetus, Norwalk, Conn.) and two pairs of oligonucleotide primers (see Fig. 3). The first pair was 5′-GCCTTCTTGAC GAGTTCTTC-3′ (primer I), located upstream and adjacent to the stop codon of the *neo* gene, and 5′-CTTGCCAAACCTA CAGGT-3′ (primer II), located 14 bp downstream of the start of the MoMLV LTR U3 sequence. The second pair was 5′-AGCGGATCCCCTTCTATCGCCTTC-3′ (primer III), located 15 bp upstream of the *neo* stop codon, and 5′-AG CAAGCTTGCCAAACCTACA-3′ (primer IV), located 17 bp downstream of the start of the MoMLV LTR U3 sequence. The primers of the second set had base substitutions which upon amplification generated the restriction sites *Bam*HI (from primer III) and *Hin*dIII (from primer IV). The fragments amplified by using the primer pair I-II were cloned by blunt-end ligation into the *Sma*I site of M13mp19, and those amplified with the pair III-IV were digested with *Bam*HI and *Hin*dIII and cloned into the corresponding sites of M13mp19. Single-stranded phage DNA from these recombinants was then prepared (32) and used in sequencing reactions carried out with the Sequenase 2.0 kit (United States Biochemicals, Cleveland, Ohio). It should be noted that gross rearrangements detected after PCR amplification and DNA sequencing were also readily detectable during Southern blotting, so they are not likely to be due to artifacts occurring during amplification.

Sequences and computer analysis. The MoMLV sequence coordinates given in Fig. 3D and E were obtained from the MoMLV complete genome sequence (GenBank accession numbers J02255, J02256, and J02257). Searches for sequence homology were performed by using the FASTA program of the Genetics Computer Group sequence analysis package (5).

RESULTS

Forward mutation rate measured by using an MoMLV vector containing the herpes simplex virus *tk* gene. To measure the mutation rate over an entire gene in an MoMLV-based vector system, the retroviral vector MSTK was constructed (Fig. 1A). The *neo* gene is expressed from the 5′ LTR promoter, and the herpes simplex virus *tk* gene is expressed from the internal SNV promoter (U3 region of the SNV LTR). Expression of *neo* confers resistance to G418, whereas expression of *tk* confers resistance to aminopterin present in HAT medium. The *tk* gene was chosen as target for mutations because an inactivating mutation in *tk* is readily detectable in cells lacking the endogenous *tk* gene by selection with the nucleotide analog BrdU. Therefore, BrdU selection was used to measure the rate of inactivation of *tk* in an MoMLV-based vector during a single cycle of retroviral replication as a measure of the forward mutation rate (see below).

For the mutation rate determination, GP+E-86 helper cells harboring one copy of the MSTK provirus were established as

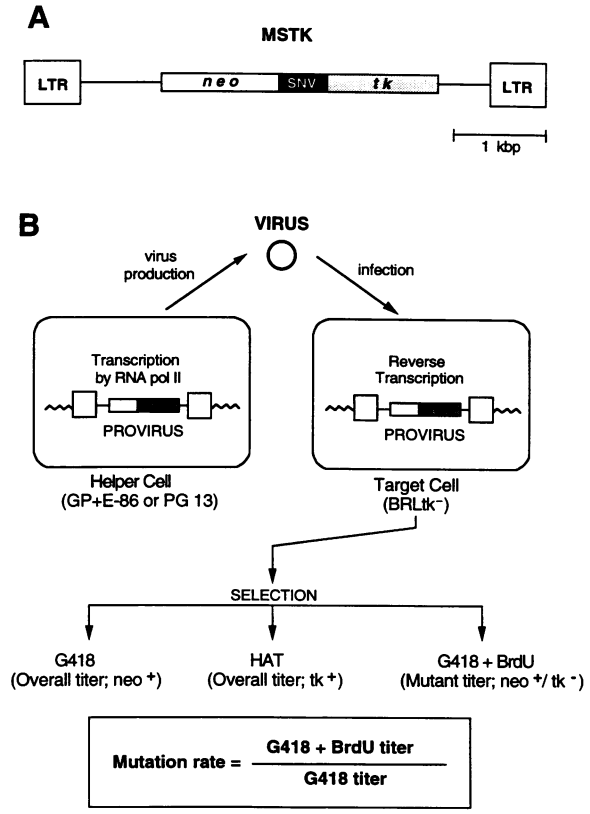

FIG. 1. Retroviral vector MSTK and protocol used for forward mutation rate determination. (A) Diagram of the MoMLV-based vector MSTK. LTR, MoMLV LTR; SNV, SNV promoter; horizontal lines, MoMLV sequences containing *cis*-acting elements required for vector virus replication. (B) Schematic representation of the protocol used to determine the forward mutation rate involving a single cycle of viral replication. Inside the cells, a vector provirus is depicted. The two open squares represent LTRs; the two rectangles represent exogenous *neo* and *tk* genes; the jagged lines represent genomic DNA; 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 by using the GaLV Env glycoproteins; G418, HAT, and G418+BrdU denote drugs used for selection to obtain the vector virus titers; neo$^+$, tk$^+$, and neo$^+$/tk$^-$ represent the phenotypes conferred by the vector proviruses present in cells obtained under each selection condition. pol, polymerase.

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

Clone	Titer (CFU/ml of virus stock)[a]			Mutation rate[b]
	G418	HAT	G418+BrdU	
1	5.0×10^3	5.0×10^3	5.0×10^2	0.100
2	4.5×10^3	1.1×10^4	1.0×10^3	0.222
3	4.0×10^4	3.5×10^4	4.5×10^3	0.112
4	1.5×10^4	1.5×10^4	1.5×10^3	0.100
5	1.0×10^4	2.0×10^4	5.0×10^2	0.050
6	1.0×10^4	1.0×10^4	7.5×10^2	0.075
7	5.0×10^5	6.0×10^5	4.0×10^4	0.080
8	3.5×10^5	4.0×10^5	1.5×10^4	0.043
9	3.5×10^5	5.0×10^5	1.0×10^5	0.286
10	3.0×10^5	2.5×10^5	4.5×10^4	0.150
11	3.0×10^5	6.5×10^5	3.5×10^4	0.117
12	1.0×10^4	1.2×10^4	7.5×10^2	0.075
13	5.0×10^5	6.5×10^5	2.5×10^4	0.050
14	4.0×10^5	6.5×10^5	4.5×10^4	0.112
15	1.2×10^5	1.5×10^5	2.0×10^3	0.017
16	6.5×10^4	5.5×10^4	5.0×10^3	0.077
17	4.0×10^4	4.0×10^4	8.0×10^3	0.200
18	6.5×10^5	6.0×10^5	1.0×10^5	0.154
				0.112 (avg)[c]

[a] G418, HAT, and G418+BrdU represent the selection drugs used to obtain the titers shown.

[b] Ratio of the G418+BrdU titer to the G418 titer. The variation in the rates obtained from the different helper cell clones is probably a consequence of normal experimental variation in viral titers. When one is dealing with viral titers in the range of those shown, a twofold variation is not unexpected. If we assume a 2-fold error, the titers can vary from 0.5- to 2-fold of the actual titer, as indicated by the following expressions, in which x represents the G418 titer and y represents the G418+BrdU titer: $x/2 < x < 2x$ and $y/2 < y < 2y$. Since the mutation rate is the ratio of the dual G418+BrdU titer to the G418 titer (y/x), then the rates from individual helper cell clones can be expected to vary as follows: $y/4x < y/x < 4y/x$. Thus, the mutation rate can be 4-fold lower to 4-fold higher than the average rate, accounting for at least a 16-fold difference between the highest and the lowest possible rates.

[c] Standard deviation = 0.069.

previously described (6, 40) and were initially used as a source of vector virus which was in turn used to infect BRLtk$^-$ cells (Fig. 1B). During this process, a single cycle of retroviral replication is completed; the cycle includes one round of proviral transcription in the helper cell by RNA polymerase II to produce the vector virion RNA and one round of reverse transcription by the viral reverse transcriptase (RT) to produce a new provirus in the target cell (Fig. 1B). It should be noted that the spread of vector virus is greatly reduced in the helper cells as a result of superinfection interference (37) (superinfection of GP+E-86 is reduced by a factor of 10^3 [17]) and in the target cells is totally blocked as a result of the absence of viral proteins required for replication. The infected BRLtk$^-$ cells were then selected in medium containing G418, HAT, or both G418 and BrdU (G418+BrdU; Fig. 1B). The vector virus titers obtained with G418 or HAT yield the overall virus titer, while the G418+BrdU titer yields the mutant titer (Fig. 1B). Cells resistant to the dual drug selection represent infected cells harboring a provirus with a functional *neo* gene and an inactivated *tk* gene (Fig. 1B). These mutant proviruses could arise as a consequence of different types of mutations, such as base substitutions, frameshifts, deletions, and insertions, involving the *tk* gene or the SNV promoter which directs its expression. The target sequence (2.7 kbp) includes untranslated regions of *neo* and *tk* and some viral sequences which are not required for *tk* expression but could be involved in rearrangements that cause its inactivation. The ratio of the mutant titer (G418+BrdU) to the overall titer (G418) represents the mutation frequency. Since the mutation frequency is measured in a single cycle of retroviral replication, it constitutes a measure of the mutation rate, that is, the mutation frequency in a single cycle of viral replication (Fig. 1B).

A total of 18 GP+E-86 helper cell clones harboring the MSTK vector were used as source of vector virus for infection of BRLtk$^-$ cells. After one cycle of retroviral replication, these helper cells yielded Neo$^+$ Tk$^-$ proviruses at an average rate of 0.11 per cycle (Table 1), indicating that *tk* was inactivated in about 11% of the G418-resistant clones. It should be noted that the apparent low vector virus titers obtained with the helper cells used in this study are due to a reduced susceptibility of BRLtk$^-$ cells for infection by the vector systems used (unpublished observations).

Control experiments were performed to rule out possible

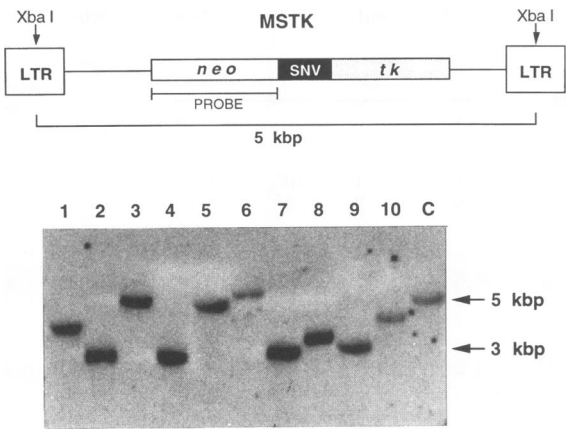

FIG. 2. Southern blot analysis of proviral DNA from G418+BrdU-resistant BRLtk$^-$ cell clones. The diagram depicts the structure of the parental MSTK vector provirus. Genomic DNA was isolated from G418+BrdU-resistant BRLtk$^-$ cell clones infected with MSTK vector virus. Ten micrograms of genomic DNA from the different BRLtk$^-$ cell clones (lanes 1 to 10) was digested with *Xba*I and analyzed by Southern blotting using a *neo*-specific probe. As a control for full-length proviral DNA (5 kbp after digestion with *Xba*I), 10 μg of genomic DNA isolated from a G418-resistant cell clone infected with MSTK vector virus was also analyzed (lane C). The mutant clones shown (in the order in which they appear) were derived from virus from the following helper cell clones shown in Table 1: 1, 17, 10, 10, 13, 14, 18, 9, 16, and 11.

interference or synergistic effects of dual selection with G418 and BrdU which might affect the G418 titers. A helper cell clone harboring an MoMLV-based vector containing only the *neo* gene (N2 [15]) was used as a source of vector virus for infection of BRLtk$^-$ cells. As expected, this vector did not yield HAT-resistant cells, and its G418 and G418+BrdU titers were very similar (2×10^6 CFU/ml was the average titer for both selections in seven experiments; data not shown). These results indicate that BrdU does not affect the G418 titer.

Southern blot analysis of mutant proviruses. To examine the structures of the proviruses in dual-resistant cells (Neo$^+$ Tk$^-$), Southern blot analysis was performed with genomic DNA isolated from G418+BrdU-resistant cells, using a *neo*-specific probe. The restriction enzyme *Xba*I was used to cleave the genomic DNA because it cuts only once within each LTR. This analysis revealed that the majority of the mutant proviruses had readily detectable gross rearrangements and that most proviruses were smaller than the parental vector (Fig. 2 and data not shown). Of 31 clones analyzed, 2 were larger and 30 were smaller than the parental vector provirus, and 2 showed no detectable size difference. Approximately half of the mutants were a consequence of an apparent 2.1-kbp deletion. It is noteworthy that these mutants were derived from viral stocks obtained from different helper cell clones, so the mutants yielding similar deletions are not arising from the viral stocks of only one helper cell clone.

An additional experiment was performed to determine whether the system used yields an accurate depiction of the mutation rate. If the observed high rate of *tk* inactivation by genetic rearrangement is not an artifact of the selections used, a similarly high rate of rearrangements should be expected in infected cells selected for a functional *neo* gene in the absence of selection for *tk* mutants. To test this prediction, 30 BRLtk$^-$ clones were obtained by infection with vector virus harvested from clone 7 (Table 1) and selected with G418 only. Southern blot analysis of these clones revealed that two of them harbored rearranged proviruses; one was smaller and the other was larger than the parental MSTK vector (data not shown). This finding indicates that the observed high rate of rearrangement is not an artifact of the selection system used.

Sequence analysis of deletion mutants. To further analyze the mutant proviruses, DNA fragments containing the deletion junctions were PCR amplified from genomic DNA and sequenced. The amplification was performed with primers flanking the SNV and *tk* sequences, and the resulting fragments were cloned into M13 to assist in determining their DNA sequences. The results of that analysis are shown in Fig. 3. About half of the deletions (8 of 14 clones analyzed) resulted from excision of the entire SNV-*tk* cassette involving 30-bp direct repeats flanking the deleted sequence which consist of plasmid sequences introduced during construction of the vector (Fig. 3A). The mutants with this deletion belong to the group which yielded a 2.1-kbp deletion in the Southern blot analysis and probably resulted from misalignment of the template during proviral DNA synthesis (26, 29). Another deletion also involved direct repeats at the ends of the deleted sequence (Fig. 3B). It was a deletion of 1,954 bp flanked by 10-bp direct repeats present in the 3′ noncoding region of *neo* and at the 3′ end of the coding region of *tk*. A deletion of 1,768 bp in a different mutant clone did not involve direct repeats except for a single-base-pair overlap at the ends of the deleted fragment (Fig. 3C). Another type of mutation involved recombination of vector viral sequences with MoMLV *env* sequences (Fig. 3D and E). The recombination resulted in the loss of most of the SNV and *tk* sequences and the insertion of *env* sequences which at their 3′ ends were indistinguishable from the MoMLV provirus. One of the mutants had an additional insert of 30 bp between the MSTK vector and MoMLV *env* sequences (Fig. 3D). The central 26-bp stretch of this insert is identical to an internal segment of a centromeric satellite DNA species found in murine chromosomes (10, 30). In this mutant, as well as in the other mutant which acquired MoMLV *env* sequences, no sequence similarity was found at the 5′ recombination points between the vector and the transduced sequences (Fig. 3D and E). It is noteworthy that the virus stocks used to obtain these two mutant proviruses originated from different helper cell clones (see implications in Discussion). Two other mutant proviruses combined deletions and insertions. One of them had a 1,133-bp deletion combined with an insertion of three C nucleotide residues (sequence corresponding to the plus strand of the provirus; Fig. 3F). The other mutant had a 2,388-bp deletion and an insertion of 261 bp (Fig. 3G and H). The inserted sequence in this last mutant does not correspond to any sequence found in the parental MSTK vector or to any sequence represented in the GenBank data base (as of April 1993). It probably represents transduced cellular sequences.

Although proviral transcription by RNA polymerase II remains a possible source of mutations, it is likely that reverse transcription is the major contributor to the high rate of rearrangement (see Discussion).

Forward mutation rate analysis using PG13 helper cells. Although the spread of the vector virus in the GP+E-86 helper cells is effectively reduced as a result of superinfection interference, it is not blocked completely. To determine whether limited cycling which might occur during propagation of the GP+E-86 helper cell clones significantly contributed to the observed high mutation rate, we performed similar mutation rate analyses using PG13 cells (21). PG13 is an NIH 3T3-derived helper cell line which produces the MoMLV Gag and Pol proteins and the GaLV Env glycoproteins. Therefore, the

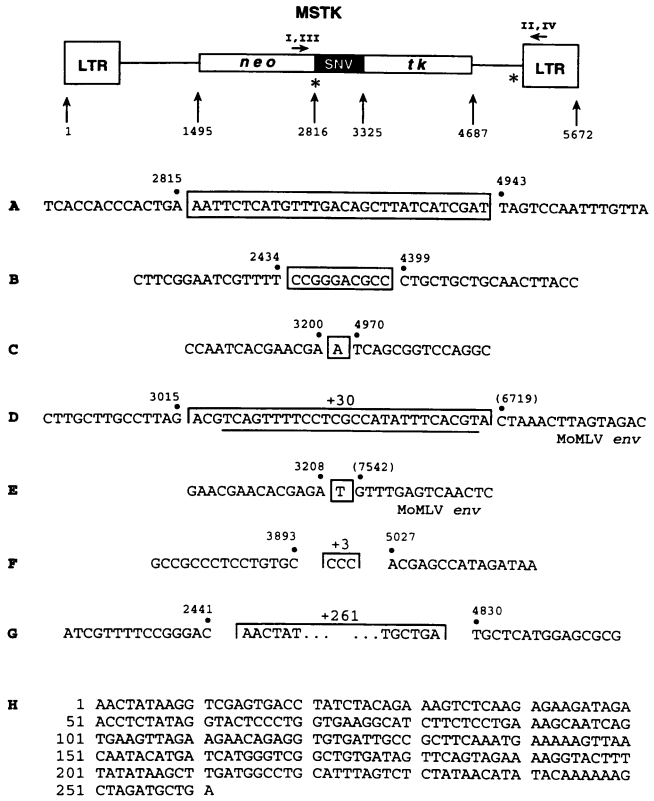

FIG. 3. Sequences of mutant proviruses obtained under G418+BrdU selection. This figure depicts the sequence of mutant proviruses analyzed by DNA sequencing after PCR amplification and molecular cloning of proviral DNA from BRLtk$^-$ cell clones that grew in the presence of G418+BrdU after infection by MSTK vector virus. The parental provirus is 5,672 bp long, and relevant coordinates are given below the vector. The approximate locations of the PCR primers are shown by the horizontal arrows above the vector (not drawn to scale). A, B, and C represent simple deletions for which the coordinates of the breakpoints are given above the sequences. For these three mutants, the deletion breakpoint was flanked by repeated sequences which are designated within the rectangles. The approximate locations of the 30-bp repeats shown in A are indicated by the asterisks below the vector. D and E represent examples of recombination which resulted in insertion of endogenous MoMLV *env* sequences. The sequence within the bracket in D represents an insertion of 30 bp between the MSTK and MoMLV *env* sequences. The underlined sequence in the insert was found to be identical to an internal segment of a murine centromeric satellite DNA. The breakpoints within the vector in D and E are again represented by numbers without parentheses, whereas the *env* sequence coordinates corresponding to their locations in the MoMLV genome are indicated by numbers with parentheses. The *env* inserts extend from the coordinates shown to nucleotide 7673 of the MoMLV genome. F and G represent examples of deletion with accompanying insertions of sequence of unknown origin. The insertions are indicated by brackets, and the sizes of the insertions are given above the brackets. The sequence of the entire 261-bp insertion found in G is given in H. The sequences of mutants displayed were derived from virus from the helper cell clones in Table 1 as follows: A from 1, B from 16, C from 1, D from 11, E from 9, F from 1, and G from 17.

virus produced by this cell line is a GaLV-pseudotyped vector virus which uses the RT of MoMLV but has the GaLV receptor specificity for infection. This pseudotyped vector virus is unable to infect NIH 3T3 cells because GaLV Env glycoproteins cannot mediate infection of these cells, at least under the conditions used in these experiments (21). Therefore, the vector virus cannot cycle in the PG13 helper cells even at low levels. This vector virus, however, can infect BRLtk$^-$ cells. PG13 cell clones containing MSTK vector proviruses were established by infection with virus produced in GP+E-86 helper cells at low multiplicity of infection and used in the forward mutation rate assay. The titers obtained from these helper cell clones revealed an average rate of *tk* inactivation of 0.18 per cycle (Table 2), which is not significantly different from that obtained from GP+E-86 helper cells. Southern blot analysis of the G418+BrdU-resistant cell clones again demonstrated that many of the mutant proviruses (7 of 10 analyzed) resulted from gross rearrangements similar to those observed in the assays using GP+E-86 helper cell clones (Fig. 4). These results indicate that the high forward mutation rate displayed by these vectors is not due to cycling of vector virus during the propagation of the helper cell clones.

TABLE 2. Titers and forward mutation rates obtained from PG13 helper cells containing MSTK proviruses

Clone	Titer (CFU/ml of virus stock)[a]			Mutation rate[b]
	G418	HAT	G418+BrdU	
1	5.0×10^4	4.0×10^4	1.6×10^4	0.320
2	7.0×10^4	8.5×10^4	1.0×10^4	0.143
3	4.5×10^4	3.5×10^4	1.2×10^4	0.267
4	9.0×10^4	1.2×10^5	1.0×10^4	0.111
5	5.5×10^4	5.0×10^4	3.5×10^3	0.064
6	5.0×10^4	9.0×10^4	9.0×10^3	0.180
7	5.5×10^4	5.5×10^4	1.5×10^4	0.273
8	6.5×10^4	6.0×10^4	7.5×10^3	0.115
9	7.5×10^4	1.2×10^5	1.8×10^4	0.240
10	1.5×10^3	5.5×10^3	5.0×10^1	0.033
				0.175 (avg)[c]

[a] G418, HAT, and G418+BrdU represent the selection drugs used to obtain the titers shown.
[b] Ratio of the G418+BrdU titer to the G418 titer.
[c] Standard deviation = 0.097.

DISCUSSION

High rate of genetic rearrangement during retrovirus vector replication. Our results show that an MoMLV-based vector carrying the *tk* gene is subject to a high forward mutation rate upon replication. In a single cycle of retroviral replication, the *tk* gene was inactivated in the MSTK vector at an average rate of 14.4% per replication cycle (this value represents an average

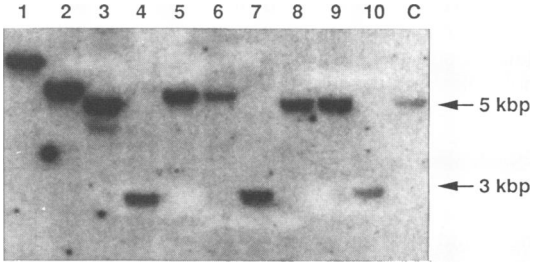

FIG. 4. Southern blot analysis of proviral DNA from mutant proviruses obtained by infection with the MSTK and MSTK2 vectors produced by PG13 helper cells. Genomic DNA from infected BRLtk$^-$ cells was analyzed as described in the legend to Fig. 2. Lane C, full-length MSTK provirus control clone (5 kbp) described in the legend to Fig. 2.

of the rates shown in Tables 1 and 2). The majority of the mutations (36 of 41) were the result of gross rearrangements, and about half of them were the result of an identical deletion involving 30-bp direct repeats (Fig. 2 to 4 and data not shown). Since such direct repeats would normally not be found between the primer binding site and the polypurine tract of a retrovirus genome, the 14.4% mutation rate is probably an overestimation of the actual rate. Excluding mutations involving the 30-bp repeats (19 of 41), the rate is 7.7%, which is probably a more accurate reflection of the mutation rate.

In the MSTK vector, the potential target size is 2.7 kbp, which includes the SNV U3 promoter, the *tk* gene, and 1.0 kbp of flanking sequence not required in *cis* for replication and not required for Neo protein production. Thus, the rate of genetic rearrangement is 2.2×10^{-5}/bp per replication cycle (i.e., 0.077/2,700 bp multiplied by 17/22, which excludes rearrangements involving 30-bp repeats). If the measured mutation rate is extrapolated to replication-competent MoMLV (8.3 kbp), it would correspond approximately to the production of one mutant provirus in every five produced. Although caution should be exercised in making this extrapolation, this estimation is similar to that observed during MoMLV replication in culture in which replication-defective mutants were produced at a frequency of 33% after a single cycle of replication (34), which lends support to the significance of our findings. The rearrangement rate observed in our MoMLV-based vectors is about five times higher than that obtained with an SNV-based vector by using a target sequence of 280 bp (27). Further experimentation is needed to determine whether this discrepancy reflects different properties of the RTs involved or it is just the result of differences in the experimental protocols used in the two studies.

Genetic rearrangements are likely to occur during reverse transcription. The mutations could have occurred during proviral replication by cellular DNA polymerases, during proviral transcription by RNA polymerase II, or during reverse transcription by the viral RT. Because of its high fidelity, cellular DNA replication is not likely to contribute significantly to the high rate of mutation observed. The high rate of gene rearrangement suggests that the polymerase responsible for the rearrangements exhibits low processivity. In vitro measurements have shown that RT polymerizes DNA with lower processivity than do other replicative polymerases (reviewed in reference 16), probably because of its need to switch templates twice during normal retroviral replication (26, 27). Although an important role for RNA polymerase II in the generation of the observed mutations cannot be excluded, it would seem that gross rearrangements are more likely to occur during reverse transcription because there is no obvious mechanism for keeping an incomplete transcript in close proximity to RNA polymerase II if it dissociates from the template. On the other hand, if an incomplete reverse transcript dissociates from its template, it still remains in close association with numerous RT molecules in a core particle.

It should be noted that another mechanism independent of polymerase errors could result in generation of gross mutations. Such mutations could arise as a result of aberrant vector viral RNA processing. For example, if cryptic splice sites are present in the exogenous sequences inserted into the vector, splicing would produce defective vector viral RNA by removing vector sequences. If such an RNA was packaged, it could then serve as a template for reverse transcription generating a provirus with a deletion. However, in this study, this does not seem to be the case since analysis of the junctions in the deletion mutants did not reveal any concordance with consensus splice junctions (Fig. 3).

Homologous as well as nonhomologous recombination is involved in vector rearrangement. Sequence analysis of a number of deletion mutants suggested that homologous as well as nonhomologous recombination was involved in the generation of rearrangements. Some of the deletions observed most likely resulted from misalignment between direct repeats present in the target sequences (Fig. 3A and B). Similar types of rearrangements involving direct repeats have been observed for deletions obtained with SNV-based vectors (27, 29) and avian sarcoma virus (24). As previously proposed, deletions involving two direct repeats presumably occur while the RT copies the first of the direct repeats and the growing DNA strand misaligns with the second direct repeat found downstream in the template (27, 29, 43).

We observed a mutation in which nonhomologous sequences were involved in the deletion of about 1.8 kbp (Fig. 3C). In this mutant, only an A residue was common to both endpoints. Deletions of sequences flanked by divergent sequences with only one base in common have been observed in other retroviruses (29, 41). These mutations suggest that the retroviral RT is capable of strand transfer between templates with little or no sequence similarity.

Transduction of MoMLV *env* sequences. Two similar but not identical MSTK mutant proviruses lost most of the SNV promoter and *tk* gene and acquired sequences corresponding to the 3′ end of the MoMLV *env* gene (Fig. 3D and E). One of them had an additional insert between the vector and MoMLV *env* sequences (Fig. 3D) which was found to be homologous to an internal region of a mouse satellite DNA species (10). The *env* sequence transduced by MSTK vector virus is probably derived from p*env*, which is an expression vector encoding the MoMLV envelope glycoproteins present in GP+E-86 cells (19). Since these mutant proviruses originated from different helper cell clones, it is unlikely that they both integrated near p*env* DNA in the helper cell clones, thereby generating vector-p*env* recombinants. Therefore, it is possible that the mutations are the result of a double crossover after copackaging of the vector and p*env* RNA. This suggests an alternative mechanism by which retroviruses can transduce cellular sequences without the need of proviral integration adjacent to the sequence to be transduced. It should be noted that although the viral encapsidation signal and most of the MoMLV 3′ LTR have been deleted from p*env*, it shares a 170-bp identity with sequences present at the 3′ end of vector MSTK starting 3′ to the *tk* gene and ending 30 bp within the 3′ LTR. Furthermore, the 5′ LTR is still present in p*env*. The proposed mechanism of recombination is shown in Fig. 5. This model suggests that both homologous (3′ junction) and nonhomologous (5′ junction) strand transfers occurred during minus-strand DNA synthesis.

In the case of the mutant provirus depicted in Fig. 3D, an additional strand transfer event involving transcripts derived from the centromeric satellite DNA would have been needed to produce the 30-bp insertion at the 5′ junction. The helper cells used in this study have not been analyzed for the presence of such transcripts; however, it has been reported that satellite DNA is transcribed in a variety of rodent cells. For example, centromeric satellite DNA transcripts have been detected in mouse senescent cardiac muscle (8), in mouse L cells (4, 9), and in cultured rat hepatoma cells (33). Thus, it seems possible that these sequences were transduced during reverse transcription using copackaged satellite RNA as a template.

The finding that two independently generated mutants acquired MoMLV *env* sequences suggests that their appearance is not a rare event. This high frequency could be the result of more efficient copackaging of the MoMLV *env* mRNA than of other cellular RNAs, which would result in a concomitant

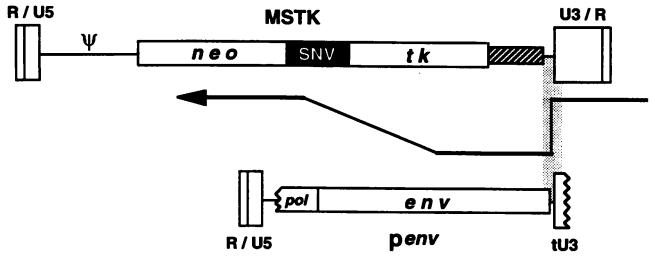

FIG. 5. Model for the generation of mutants with deletions combined with insertions of MoMLV *env* sequences. Structures of the mRNAs derived from the retroviral vector MSTK and the expression vector p*env* are shown. Ψ represents the encapsidation signal present only in the MSTK vector RNA. The hatched box represents untranslated sequences. The jagged lines in the vector p*env* represent truncation of *pol* and LTR sequences produced during its construction. The R, U5, and U3 elements of the LTRs are shown. tU3 represents the truncated U3 present at the 3′ end of the p*env* RNA. The shaded area corresponds to region of homology between the two vectors at their 3′ ends, and the arrow represents the minus-strand DNA being synthesized, showing the two template switches between the two mRNAs. The recombination point within the region of homology cannot be determined with precision.

increase in the probability of recombination with the MSTK vector sequences. This seems possible since elements within the U5 region have been shown to act as auxiliary sequences for packaging of MoMLV genomic RNA in addition to the already mapped encapsidation signal (23). High levels of p*env* expression and its homology with the MSTK vector could also increase the probability of copackaging and recombination, respectively. Further analysis is required to determine the contribution of these possible mechanisms, which are not mutually exclusive, to the transduction of MoMLV *env* sequences.

Deletions combined with insertions. Two other mutant proviruses resulted from deletions combined with insertions. In one of them, the deletion was combined with an insertion of three C residues of unknown origin (Fig. 3F). Small insertions of one or more nucleotides at deletion or recombination junctions have also been observed for other retroviruses (27, 29, 39, 44). This type of insertion could be the result of a strand transfer from one deletion endpoint to a template containing the inserted nucleotide residues and then another transfer to the template containing the second deletion endpoint. A second possibility is that the insertions are consequences of non-template-directed elongation of the nascent DNA chain by the RT combined with forced copy choice (3). To generate the insertion in our mutant, three G residues would have been incorporated during minus-strand DNA synthesis at the growing end of the molecule upon reaching a nick in the vector genomic RNA (3′ deletion endpoint), followed by the transfer to the template containing the 5′ deletion endpoint. Although the template-directed insertion remains a possibility in the absence of more experimental evidence, the second alternative is indirectly supported by biochemical analysis performed with RT purified from different retroviruses. In those studies, it was shown that human immunodeficiency virus, avian myeloblastosis virus, and MoMLV RTs are capable of elongating the end of a growing DNA strand past the end of the RNA template molecule by one or several nucleotides (25, 28).

In the second mutant of this class, the insertion was a sequence of 261 nucleotides which was probably transduced from the helper cells (Fig. 3G and H). Since the genomic locations and flanking sequences of the insert and the provirus in the helper cells are unknown, it is not possible at this point to conjecture about the mechanism of transduction involved. This and some other mutant proviruses which have not been fully characterized yet but which probably contain insertions of sequences of up to 1.5 kbp (Fig. 4 and data not shown) present an opportunity to study retroviral transduction.

Conclusions and perspectives. Our results suggest that at least some MoMLV-based retroviral vectors can undergo a high rate of genetic rearrangement. This inference is consistent with views of retroviral replication as a process that results in the generation of genetic diversity as well as defective proviruses (3, 13). These results have significance for the utilization of MoMLV-based vectors for retrovirus-mediated gene transfer, including somatic cell gene therapy. Retrovirus-mediated gene transfer usually involves a single cycle of replication since producer cell clones harboring the vector provirus are used as source of vector virus to infect target cells which do not express the viral proteins (20). Thus, there could be similar rates of gene inactivation for genetic units of the size described here. It should be pointed out, however, that the mutation rate is at least partly sequence dependent and that mutation rates could be lower with use of other exogenous sequences. Nevertheless, even a rate of gene inactivation of 7.7% is likely to be acceptable for many gene transfer protocols.

In our experiments, only one marker gene in a single vector was analyzed. The molecular analysis of a greater number of mutants obtained with this and other vector systems is required to gain a better insight into the sequence elements affecting the generation of mutations during retroviral replication. The effects of the relative positions of the target genes within the vector as well as the role of cellular factors upon the mutation rate are issues that deserve careful analysis. This effort will result in a better understanding of retroviral mutagenesis and will lead to a more rational design of vectors for applications in which high genetic stability of the vector is required.

ACKNOWLEDGMENTS

We thank Abram Gabriel, Celine Gelinas, Stuart W. Peltz, Bradley D. Preston, Han-Mo Koo, and Hong Yu for discussion and helpful comments on the manuscript.

This work was supported by Public Health Service grant 5 R29 CA50777 from the National Institutes of Health, grant 13-92 from the Foundation of the University of Medicine and Dentistry of New Jersey, grant 92-39-CCR-00 from the New Jersey Commission for Cancer Research, and basic research grant 1-FY92-0699 from the March of Dimes. J.P.D. is scholar of the Pew Foundation. A.V.-E. received partial support from the International Center for Molecular and Cellular Biology, Monterrey, Mexico.

REFERENCES

1. **Alizon, M., S. Wain-Hobson, L. Montagnier, and P. Sonigo.** 1986. Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. Cell **46**:63–74.
2. **Clavel, F., M. D. Hoggan, R. L. Willey, K. Strebel, M. A. Martin, and R. Repaske.** 1989. Genetic recombination of human immunodeficiency virus. J. Virol. **63**:1455–1459.
3. **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.
4. **Cohen, A. K., T. Y. Huh, and C. W. Helleiner.** 1973. Transcription of satellite DNA in mouse L-cells. Can. J. Biochem. **51**:529–532.
5. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
6. **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.

7. **Emerman, M., and H. M. Temin.** 1984. Genes with promoters in retrovirus vectors can be independently suppressed by an epigenetic mechanism. Cell **39:**459–467.
8. **Gaubatz, J. W., and R. G. Cutler.** 1990. Mouse satellite DNA is transcribed in senescent cardiac muscle. J. Biol. Chem. **265:**17753–17758.
9. **Harel, J., N. Hanania, H. Tapiero, and L. Harel.** 1968. RNA replication by nuclear satellite DNA in different mouse cells. Biochem. Biophys. Res. Commun. **33:**696–701.
10. **Horz, W., and W. Altenburger.** 1981. Nucleotide sequence of mouse satellite DNA. Nucleic Acids Res. **9:**683–696.
11. **Hu, W.-S., and H. M. Temin.** 1990. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. Proc. Natl. Acad. Sci. USA **87:**1556–1560.
12. **Hu, W.-S., and H. M. Temin.** 1990. Retroviral recombination and reverse transcription. Science **250:**1227–1233.
13. **Katz, R. A., and A. M. Skalka.** 1990. Generation of diversity in retroviruses. Annu. Rev. Genet. **24:**409–445.
14. **Kawai, S., and M. Nishizawa.** 1984. New procedure for DNA transfection with polycation and dimethyl sulfoxide. Mol. Cell. Biol. **4:**1172–1174.
15. **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.
16. **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.
17. **Kuo, M.-L. (Rutgers University).** Personal communication.
18. **Leider, J. M., P. Palese, and F. I. Smith.** 1988. Determination of the mutation rate of a retrovirus. J. Virol. **62:**3084–3091.
19. **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.
20. **Miller, A. D.** 1992. Human gene therapy comes of age. Nature (London) **357:**455–460.
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. **Murphy, J. E., and S. P. Goff.** 1989. Construction and analysis of deletion mutations in the U5 region of Moloney murine leukemia virus: effects on RNA packaging and reverse transcription. J. Virol. **63:**319–327.
24. **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.
25. **Patel, P. H., and B. D. Preston.** Infidelity of human immunodeficiency virus type 1 reverse transcriptase at RNA and DNA template ends. Proc. Natl. Acad. Sci. USA, in press.
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. **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.
28. **Peliska, J. A., and S. J. Benkovic.** 1992. Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. Science **258:**1112–1118.
29. **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.
30. **Radic, M. K., K. Lundgren, and B. A. Hamkalo.** 1987. Curvature of mouse satellite DNA and condensation of chromatin. Cell **50:**1101–1108.
31. **Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim.** 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230:**1350–1354.
32. **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.
33. **Sealy, L., J. Hartley, J. Donelson, R. Chalkley, N. Hutchinson, and B. Hamkalo.** 1981. Characterization of a highly repetitive sequence DNA family in rat. J. Mol. Biol. **145:**291–318.
34. **Shields, A., O. N. Witte, E. Rothenberg, and D. Baltimore.** 1978. High frequency of aberrant expression of Moloney murine leukemia virus in clonal infections. Cell **14:**601–609.
35. **Shimotohno, K., and H. M. Temin.** 1981. Formation of infectious progeny virus after insertion of herpes simplex thymidine kinase gene into DNA of an avian retrovirus. Cell **26:**67–77.
36. **Shimotohno, K., and H. M. Temin.** 1982. Spontaneous variation and synthesis in the U3 region of the long terminal repeat of an avian retrovirus. J. Virol. **41:**163–171.
37. **Steck, F. T., and H. Rubin.** 1966. The mechanism of interference between an avian leukosis virus and Rous sarcoma virus. I. Establishment of interference. Virology **29:**628–641.
38. **Stuhlman, H., and P. Berg.** 1992. Homologous recombination of copackaged retrovirus RNAs during reverse transcription. J. Virol. **66:**2378–2388.
39. **Swain, A., and J. M. Coffin.** 1992. Mechanism of transduction by retroviruses. Science **255:**841–845.
40. **Varela-Echavarria, 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.
41. **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.
42. **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.
43. **Xu, H., and J. D. Boeke.** 1987. High-frequency deletion between homologous sequences during retrotransposition of Ty elements in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA **84:**8553–8557.
44. **Zhang, J., and H. M. Temin.** 1993. 3′ junction of oncogene-virus sequences and the mechanisms for formation of highly oncogenic retroviruses. J. Virol. **67:**1747–1751.